

Why Second-Order Sufficient Conditions are, in a Way, Easy – or – Revisiting the Calculus for Second Subderivatives

Matúš Benko

*Applied Mathematics and Optimization, University of Vienna, Austria
matus.benko@univie.ac.at*

Patrick Mehrlitz

*Institute of Mathematics, BTU Cottbus-Senftenberg, Cottbus, Germany
mehrlitz@b-tu.de*

Dedicated to Roger J-B Wets on the occasion of his 85th birthday.

Received: June 8, 2022

Accepted: February 2, 2023

We readdress the classical topic of second-order sufficient optimality conditions for optimization problems with nonsmooth structure. Based on the so-called second subderivative of the objective function and of the indicator function associated with the feasible set, one easily obtains second-order sufficient optimality conditions of abstract form. In order to exploit further structure of the problem, e.g., composite terms in the objective function or feasible sets given as (images of) pre-images of closed sets under smooth transformations, to make these conditions fully explicit, we study calculus rules for the second subderivative under mild conditions. To be precise, we investigate a chain rule and a marginal function rule, which then also give a pre-image and image rule, respectively. As it turns out, the chain rule and the pre-image rule yield lower estimates, desirable in order to obtain sufficient optimality conditions, for free. Similar estimates for the marginal function and the image rule are valid under a comparatively mild inner calmness* assumption. Our findings are illustrated by several examples including problems from composite, disjunctive, and nonlinear second-order cone programming.

Keywords: Composite optimization, optimization with geometric constraints, second-order sufficient optimality conditions, second-order variational calculus, second subderivative.

2010 Mathematics Subject Classification: 49J52, 49J53, 90C26, 90C46.

1. Introduction

The derivation of second-order necessary and sufficient optimality conditions has been a key topic in mathematical optimization throughout the last decades. Clearly, using second-order information associated with the appearing data provides sharper conditions than those ones obtained from pure first-order information. Moreover, there is a huge interest in second-order sufficient optimality conditions since these do not only guarantee stability of the underlying local minimizer in certain sense, but can be applied in numerical optimization in order to show local fast convergence of diverse solution methods.

Second-order optimality conditions for smooth standard nonlinear optimization problems (NLPs) date back, e.g., to [3, 28] and essentially postulate positive (semi-) definiteness of the Hessian of the associated Lagrangian function over a suitable critical cone. We are interested in a much broader setting where the constraint region is given by so-called geometric constraints, which are given by feasible sets of the form $F^{-1}(C) := \{x \in \mathbb{R}^n \mid F(x) \in C\}$ where $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a twice continuously differentiable mapping and $C \subset \mathbb{R}^m$ is a closed set. Second-order optimality conditions typically combine the second-order derivative of the objective function and the curvature of the feasible set $F^{-1}(C)$. In case of standard NLPs, where the set C is convex and *polyhedral*, the curvature of the feasible set comes solely from the constraint mapping F and is incorporated in the Hessian of the Lagrangian. As shown e.g. in [19, 29], the situation is analogous for so-called disjunctive programs, where C is still polyhedral, i.e., the union of finitely many convex polyhedral sets. In general, however, the curvature of the feasible set comes from both, mapping F and set C . Exemplary, in nonlinear second-order cone programming, C corresponds to the well-known second-order cone, i.e., a curved, closed, convex set, and second-order optimality conditions additionally comprise a so-called curvature term associated with C , see [11]. Moreover, it has been suggested in [7] that various rather challenging problem classes like bilevel optimization problems, see e.g. [15], or optimization problems with (quasi-) variational inequality constraints, see e.g. [16, 27, 31], possess certain common structure where C itself is quite complicated and so the computation of its curvature is far from trivial.

Various variational tools have been used to capture the curvature of the feasible set. Among the standard ones are the support function, applied to a suitable (approximation of some) second-order tangent set to the feasible set, see [10, 12], and the second subderivative, applied to the indicator function associated with the feasible set, see the classical references [35, 38], [30, Section 7] and [40] for some recent developments, and [13, 41] for generalizations of this approach to optimization in abstract spaces. We believe that these two tools are more suitable for *sufficient* conditions since the related second-order sufficient conditions in terms of F and C can be derived without any additional requirements, such as constraint qualifications. This fact was shown in [7, Section 3], and we further clarify it in this paper from the perspective of calculus rules. On the other hand, to derive the respective necessary conditions, one typically needs rather severe assumptions, see e.g. [21, Theorem 4]. A new construction, the so-called lower generalized support function, was recently introduced in [21] as another tool to measure the curvature of the feasible set. The resulting second-order necessary conditions in terms of F and C are valid under very mild assumptions (so-called directional metric subregularity), see [21, Theorem 2]. Again, this approach seems to be less suitable for sufficient conditions. We point out that the replacement of the standard support function (particularly useful in the convex setting) by the lower generalized support function to some extent resembles the move from the convex subdifferential to the limiting one, which is mainly suitable for (first-order) *necessary* optimality conditions. All these three tools and approaches are currently being closely investigated in a larger context of second-order variational analysis in [7] and a forthcoming continuation.

Dealing with second-order conditions, it is a common aim to derive so-called “no-gap” conditions, meaning that the only difference between necessary and sufficient optimality conditions appears in the involved relation sign. The above considerations suggest, however, that this might be only possible in a restrictive setting, regardless of what tool is used to handle the curvature of the feasible set, and that it may not be of main interest. Perhaps, it is better to focus on the purpose of necessary and sufficient conditions separately, potentially even using different variational tools. This seems to be in line with the comments of Poliquin and Rockafellar from their paper [33] introducing the concept of tilt stability:

“The role of optimality conditions is seen rather in the justification of numerical algorithms, in particular their stopping criteria, convergence properties and robustness. From that angle, the goal of theory could be different. Instead of focusing on the threshold between necessity and sufficiency, one might more profitably try to characterize the stronger manifestations of optimality that support computational work.”

Let us also mention two related recent works of Rockafellar which use a conceptually very interesting approach and deal with sufficient conditions with focus on stability properties of minimizers as well as related aspects of numerical optimization, see [36, 37].

In this paper, we focus solely on sufficient optimality conditions via second subderivatives, and we proceed by exploring basic calculus rules that enable us to estimate the curvature of the feasible set by computable expressions. Let us point out that for sufficient conditions, only lower estimates are relevant, and so we focus on those ones without trying to get the best possible upper estimates which would be needed to infer necessary conditions. The most important calculus principle is the pre-image rule which provides an estimate for the curvature of the feasible set $F^{-1}(C)$ in terms of second derivatives of F and the curvature term associated with C . For NLPs and disjunctive programs, C has no curvature and so the sufficient conditions can be derived just from the pre-image rule. For more challenging programs, one may need to further apply calculus rules to estimate the curvature of C until eventually ending up with a set possessing no curvature, e.g. a polyhedral set, and fully explicit optimality conditions. This is precisely the case for the aforementioned setting from [7] which covers bilevel problems and problems with (quasi-) variational inequality constraints. Therein, the set C is the image of the pre-image of the graph of a normal cone mapping associated with a convex polyhedral set, and the latter is a polyhedral set as well.

The main message of the paper is that calculus for second subderivatives is very easy and so are the resulting sufficient second-order optimality conditions, which can be readily applied to a large variety of optimization problems, including very challenging ones. More precisely, the lower estimates relevant for sufficient conditions are valid under very mild assumptions. Indeed, the pre-image rule yields a general lower estimate in the absence of constraint qualifications. This is in fact a very simple observation, yet it seems that, with the exception of [7, Section 3], both standard works [35, 38] as well as the more recent contributions [30, Section 7] and [40] postulated a superfluous constraint qualification for that purpose. Let us

mention that the sufficient conditions from [30, 40] can be derived by just applying the pre-image rule to estimate the second subderivative of the indicator function associated with the feasible set $F^{-1}(C)$, while [7, Theorem 3.3] provides a stronger result: it needs no constraint qualification and a milder second-order condition, yet it yields a more stable minimizer. Interestingly, the calculus rules are so versatile, that we are able to fully recover this result by estimating the second subderivative of the function $f(x) := \max\{f_0(x) - f_0(\bar{x}), \text{dist}(F(x), C)\}$, $x \in \mathbb{R}^n$, where f_0 is the objective function, $\bar{x}$ is a fixed reference point, and $\text{dist}(\cdot, C)$ stands for the distance function associated with C . On the other hand, the image rule yields only an upper estimate for free, while the lower estimate is valid in the presence of a suitable qualification condition. Here, we exploit the recently introduced inner calmness* property from [4] for this purpose. The latter is not very restrictive and can be efficiently verified. Particularly, an easy consequence of the famous Walkup-Wets result on Lipschitzness of convex polyhedral mappings, see [42], is that polyhedral mappings enjoy inner calmness*, see [4, Theorem 3.4]. General sufficient conditions for inner calmness* can be found in [9]. Moreover, in applications to the challenging problem classes where the image rule is needed, the inner calmness* assumption has to hold for an associated so-called multiplier mapping which was shown to be inner calm* under reasonable conditions in [4, Theorem 3.9]. Finally, in Proposition 2.2, we prove that the projection mapping associated with any closed set is inner calm* at every point of the set.

The remainder of this paper is organized as follows. In Section 2, we provide the necessary preliminaries from variational analysis and generalized differentiation which are used in this paper. Particularly, we recall some essential theoretical foundations of second subderivatives in Section 2.3 and, specifically, their role in the context of second-order optimality conditions in Section 2.4. In Section 3, we first address the calculus of second subderivatives associated with indicator functions in Section 3.1, and as already pointed out in [7], this naturally leads to the introduction of a so-called directional proximal normal cone. For later use, we also introduce a directional proximal subdifferential in order to capture the finiteness of second subderivatives in Section 3.2. The essential Section 4 is dedicated to the derivation of calculus rules for second subderivatives. We derive a quite general composition rule as well as a marginal function rule which can be used to infer a pre-image and an image rule for indicator functions, respectively. Based on these findings, we are in position to easily derive second-order sufficient optimality conditions in constrained optimization over geometric constraints in Section 5 and composite optimization Section 6. Particular applications of these results in disjunctive optimization, second-order cone programming, and optimization with structured geometric constraints are presented in Section 5.1 to 5.3, respectively. The paper closes with some concluding remarks in Section 7.

2. Preliminaries

2.1. Notation

In this paper, we mainly use standard notation as utilized in [38]. For brevity of notation, we do not properly distinguish between a sequence and its terms, writing simply z_k instead of, say, $\{z_k\}$ or $\{z_k\}_{k \in \mathbb{N}}$.

Throughout the paper, we equip $\mathbb{R}^n$ with the Euclidean norm $\|\cdot\|$ and the Euclidean inner product $\langle \cdot, \cdot \rangle$. The open and closed ε -ball around some point $\bar{z} \in \mathbb{R}^n$ are denoted by $\mathbb{U}_\varepsilon(\bar{z})$ and $\mathbb{B}_\varepsilon(\bar{z})$, respectively. For given $w \in \mathbb{R}^n$,

$$\{w\}^\perp := \{z^* \in \mathbb{R}^n \mid \langle z^*, w \rangle = 0\}$$

is the *annihilator* of w . By $\text{dist}(\bar{z}, \Omega) := \inf_{z \in \Omega} \|z - \bar{z}\|$ we denote the distance of $\bar{z} \in \mathbb{R}^n$ to a nonempty set $\Omega \subset \mathbb{R}^n$. Additionally, for each nonempty index set $I \subset \{1, \dots, n\}$, the vector $\bar{z}_I$ results from $\bar{z}$ by deleting all components whose indices do not belong to I . We use $\mathbf{e}_1, \dots, \mathbf{e}_n \in \mathbb{R}^n$ to denote the canonical unit vectors of $\mathbb{R}^n$.

For a twice continuously differentiable function $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$ and some point $\bar{z} \in \mathbb{R}^n$, $\nabla f_0(\bar{z})$ and $\nabla^2 f_0(\bar{z})$ denote the *gradient* and the *Hessian* of f_0 at $\bar{z}$. For each $w \in \mathbb{R}^n$, we exploit $\nabla^2 f_0(\bar{z})(w, w) := w^\top \nabla^2 f_0(\bar{z}) w$. For a mapping $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ and a vector $y^* \in \mathbb{R}^m$, $\langle y^*, F \rangle: \mathbb{R}^n \rightarrow \mathbb{R}$ given by $\langle y^*, F \rangle(z) := \langle y^*, F(z) \rangle$, $z \in \mathbb{R}^n$, is the associated scalarization mapping. In case where F is twice continuously differentiable and $\bar{z}$ is fixed, $\nabla F(\bar{z})$ is the *Jacobian* of F at $\bar{z}$. Furthermore, we use

$$\langle y^*, \nabla^2 F(\bar{z})(w, w) \rangle := \sum_{i=1}^m y_i^* \nabla^2 F_i(\bar{z})(w, w)$$

for each $w \in \mathbb{R}^n$ for brevity of notation. A single-valued mapping $G: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is referred to as *calm* at $\bar{z} \in \mathbb{R}^n$ in direction $w \in \mathbb{R}^n$ if there is a constant $\kappa > 0$ such that, for each sequences $t_k \downarrow 0$ and $w_k \rightarrow w$, $\|G(\bar{z} + t_k w_k) - G(\bar{z})\| \leq \kappa t_k \|w_k\|$ holds for sufficiently large $k \in \mathbb{N}$. For $w := 0$, this notion recovers the classical property of G to be calm at $\bar{z}$, and this is weaker than local Lipschitz continuity of G at $\bar{z}$.

Fix a closed set $\Omega \subset \mathbb{R}^n$ and some point $\bar{z} \in \Omega$. The sets

$$\widehat{N}_\Omega(\bar{z}) := \{z^* \in \mathbb{R}^n \mid \langle z^*, z - \bar{z} \rangle \leq o(\|z - \bar{z}\|) \forall z \in \Omega\},$$

$$\widehat{N}_\Omega^p(\bar{z}) := \{z^* \in \mathbb{R}^n \mid \langle z^*, z - \bar{z} \rangle \leq O(\|z - \bar{z}\|^2) \forall z \in \Omega\}$$

are referred to as the *regular* and *proximal normal cone* to Ω at $\bar{z}$, respectively, and these are closed, convex cones by definition. In case $\bar{z} \notin \Omega$, we set $\widehat{N}_\Omega(\bar{z}) := \widehat{N}_\Omega^p(\bar{z}) := \emptyset$. Based on the regular normal cone, we can define the so-called *limiting normal cone* to Ω at $\bar{z}$ by means of

$$N_\Omega(\bar{z}) := \{z^* \in \mathbb{R}^n \mid \exists z_k \rightarrow \bar{z}, \exists z_k^* \rightarrow z^*, \forall k \in \mathbb{N}: z_k^* \in \widehat{N}_\Omega(z_k)\}.$$

The latter is a closed cone which does not need to be convex. We set $N_\Omega(\bar{z}) := \emptyset$ for each $\bar{z} \notin \Omega$. Clearly, we have $\widehat{N}_\Omega^p(\bar{z}) \subset \widehat{N}_\Omega(\bar{z}) \subset N_\Omega(\bar{z})$, and all these cones coincide with the standard normal cone of convex analysis whenever Ω is convex. For some direction $w \in \mathbb{R}^n$, let us recall that

$$N_\Omega(\bar{z}; w) := \{z^* \in \mathbb{R}^n \mid \exists t_k \downarrow 0, \exists w_k \rightarrow w, \exists z_k^* \rightarrow z^*, \forall k \in \mathbb{N}: z_k^* \in \widehat{N}_\Omega(\bar{z} + t_k w_k)\}$$

is referred to as the *directional limiting normal cone* to Ω at $\bar{z}$ in direction w . This notion has been introduced in [18, 22] in the Banach space setting, and a simplified finite-dimensional counterpart of the definition appears in [19]. The calculus of the

directional normal cone has been explored in the paper [6]. Whenever $w \in T_\Omega(\bar{z})$ holds, where

$$T_\Omega(\bar{z}) := \{w \in \mathbb{R}^n \mid \exists t_k \downarrow 0, \exists w_k \rightarrow w, \forall k \in \mathbb{N}: \bar{z} + t_k w_k \in \Omega\}$$

denotes the *tangent cone* to Ω at $\bar{z}$, then $N_\Omega(\bar{z}; w)$ is a closed cone which satisfies $N_\Omega(\bar{z}; w) \subset N_\Omega(\bar{z})$. For arbitrary $w \notin T_\Omega(\bar{z})$, $N_\Omega(\bar{z}; w)$ is empty. We also set $N_\Omega(\bar{z}; w) := \emptyset$ for each $\bar{z} \notin \Omega$.

Let $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ denote the extended real line. Recall that a function $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is called *proper* whenever $h(z) > -\infty$ hold for all $z \in \mathbb{R}^n$ while there is at least one $\bar{z} \in \mathbb{R}^n$ such that $h(\bar{z}) < \infty$. Let us note that whenever h is lower semicontinuous at $\bar{z} \in \mathbb{R}^n$ such that $|h(\bar{z})| < \infty$, then $h(z) > -\infty$ holds for all $z \in \mathbb{R}^n$ in a neighborhood of $\bar{z}$, i.e., h is locally proper at $\bar{z}$. The sets $\text{dom } h := \{z \in \mathbb{R}^n \mid h(z) < \infty\}$ and $\text{epi } h := \{(z, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid h(z) \leq \alpha\}$ are called *domain* and *epigraph* of h , respectively. Observe that whenever h is proper, we have $\text{dom } h = \{z \in \mathbb{R}^n \mid |h(z)| < \infty\}$. Furthermore, whenever h is lower semicontinuous, then $\text{epi } h$ is closed. For a lower semicontinuous function $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ and some point $\bar{z} \in \mathbb{R}^n$ where $|h(\bar{z})| < \infty$, the mapping $dh(\bar{z}): \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ defined by

$$dh(\bar{z})(w) := \liminf_{t \downarrow 0, w' \rightarrow w} \frac{h(\bar{z} + tw') - h(\bar{z})}{t} \quad \forall w \in \mathbb{R}^n$$

is called the *subderivative* of h at $\bar{z}$. Clearly, $dh(\bar{z})$ is a lower semicontinuous mapping. Observe that we have $\text{epi } dh(\bar{z}) = T_{\text{epi } h}(\bar{z}, h(\bar{z}))$. Whenever h is continuously differentiable at $\bar{z}$, we find $dh(\bar{z})(w) = \langle \nabla h(\bar{z}), w \rangle$ for all $w \in \mathbb{R}^n$. Let us mention that $dh(\bar{z})(w)$ is also referred to as lower Hadamard directional derivative of h at $\bar{z}$ in direction w in the literature. Note that the definition of the subderivative is also possible for arbitrary functions which are not necessarily lower semicontinuous. However, keeping in mind that our main purpose behind the consideration of variational objects is related to applications in mathematical optimization, it is reasonable to focus on lower semicontinuous functions. Some of the results in this paper can, via some nearby adjustments, be extended to functions which are not lower semicontinuous. Anyhow, for simplicity and brevity of presentation, we will use lower semicontinuity as a standing assumption in the remainder of the paper.

Recall that the sets

$$\begin{aligned} \widehat{\partial}h(\bar{z}) &:= \{z^* \in \mathbb{R}^n \mid (z^*, -1) \in \widehat{N}_{\text{epi } h}(\bar{z}, h(\bar{z}))\}, \\ \widehat{\partial}^p h(\bar{z}) &:= \{z^* \in \mathbb{R}^n \mid (z^*, -1) \in \widehat{N}_{\text{epi } h}^p(\bar{z}, h(\bar{z}))\}, \\ \partial h(\bar{z}) &:= \{z^* \in \mathbb{R}^n \mid (z^*, -1) \in N_{\text{epi } h}(\bar{z}, h(\bar{z}))\} \end{aligned}$$

are called *regular*, *proximal*, and *limiting subdifferential* of h at $\bar{z}$, respectively. Clearly, we have $\widehat{\partial}^p h(\bar{z}) \subset \widehat{\partial}h(\bar{z}) \subset \partial h(\bar{z})$, and whenever h is convex, all these subdifferentials coincide with the subdifferential in the sense of convex analysis.

2.2. Inner semicompactness and inner calmness* of set-valued mappings

For a set-valued mapping $\Gamma: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, the sets $\text{dom } \Gamma := \{x \in \mathbb{R}^n \mid \Gamma(x) \neq \emptyset\}$ and $\text{gph } \Gamma := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid y \in \Gamma(x)\}$ are referred to as *domain* and *graph* of Γ ,

respectively. Furthermore, $\Gamma^{-1}: \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ given by $\Gamma^{-1}(y) := \{x \in \mathbb{R}^n \mid y \in \Gamma(x)\}$, $y \in \mathbb{R}^m$, is the *inverse* of Γ .

Let the point $(\bar{x}, \bar{y}) \in \text{gph } \Gamma$ be fixed. We refer to $D\Gamma(\bar{x}, \bar{y}): \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given via $\text{gph } D\Gamma(\bar{x}, \bar{y}) := T_{\text{gph } \Gamma}(\bar{x}, \bar{y})$ as the *graphical derivative* of Γ at $(\bar{x}, \bar{y})$. If Γ is single-valued at $\bar{x}$, we exploit the notation $D\Gamma(\bar{x}) := D\Gamma(\bar{x}, \Gamma(\bar{x}))$ for brevity. Note that whenever $G: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a single-valued and continuously differentiable mapping, then we have $DG(\bar{x})(u) = \{\nabla G(\bar{x})u\}$ for each $u \in \mathbb{R}^n$.

Fix $\bar{x} \in \text{dom } \Gamma$. For a set $\Omega \subset \mathbb{R}^n$, we say that Γ is *inner semicompact* at $\bar{x}$ w.r.t. Ω whenever for each sequence $x_k \rightarrow \bar{x}$ such that $x_k \in \Omega$ for all $k \in \mathbb{N}$, there exist $y \in \mathbb{R}^m$ and a sequence $y_\ell \rightarrow y$ such that $y_\ell \in \Gamma(x_{k_\ell})$ holds for each $\ell \in \mathbb{N}$ where x_{k_ℓ} is a subsequence of x_k . If $\Omega := \mathbb{R}^n$ can be chosen, Γ is called *inner semicompact* at $\bar{x}$ for simplicity.

The next lemma shows that inner semicompactness at a single point already ensures inner semicompactness in a neighborhood of this point.

Lemma 2.1. *Let $\Gamma: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping and let $\bar{x} \in \text{dom } \Gamma$ be a point where Γ is inner semicompact w.r.t. $\text{dom } \Gamma$. Then there exists $\varepsilon > 0$ such that Γ is inner semicompact w.r.t. $\text{dom } \Gamma$ at all points from $\mathbb{B}_\varepsilon(\bar{x}) \cap \text{dom } \Gamma$. Moreover, $\mathbb{B}_\varepsilon(\bar{x}) \cap \text{dom } \Gamma$ is closed provided $\text{gph } \Gamma$ is closed.*

Proof. Suppose that the assertion is not true. Then we find a sequence $x_k \rightarrow \bar{x}$ such that $x_k \in \text{dom } \Gamma$ and Γ is not inner semicompact at x_k w.r.t. $\text{dom } \Gamma$ for each $k \in \mathbb{N}$. Thus, for each $k \in \mathbb{N}$, there is a sequence $x_k^\ell \rightarrow x_k$ such that $x_k^\ell \in \text{dom } \Gamma$ for each $\ell \in \mathbb{N}$ and $\text{dist}(0, \Gamma(x_k^\ell)) \rightarrow \infty$ as $\ell \rightarrow \infty$. For each $k \in \mathbb{N}$, fix $\ell(k) \in \mathbb{N}$ such that $\|x_k^{\ell(k)} - x_k\| \leq 1/k$ and $\text{dist}(0, \Gamma(x_k^{\ell(k)})) \geq k$. Due to $x_k^{\ell(k)} \rightarrow \bar{x}$ and $x_k^{\ell(k)} \in \text{dom } \Gamma$ for each $k \in \mathbb{N}$, Γ is not inner semicompact at $\bar{x}$ w.r.t. $\text{dom } \Gamma$ which is a contradiction.

The statement about closedness follows from [4, Lemma 2.1] as a consequence of the closedness of $\text{gph } \Gamma$. $\square$

Let us also recall the *inner calmness** property which has been coined in [4] and further studied in [9]. It can be interpreted as a quantitative version of the inner semicompactness from above. For $\bar{x} \in \text{dom } \Gamma$ and $\Omega \subset \mathbb{R}^n$, Γ is called *inner calm** at $\bar{x}$ w.r.t. Ω whenever there exists $\kappa > 0$ such that for each sequence $x_k \rightarrow \bar{x}$ satisfying $x_k \in \Omega$ for all $k \in \mathbb{N}$, there exist $y \in \mathbb{R}^m$ and a sequence $y_\ell \rightarrow y$ with $y_\ell \in \Gamma(x_{k_\ell})$ and $\|y_\ell - y\| \leq \kappa \|x_{k_\ell} - \bar{x}\|$ for each $\ell \in \mathbb{N}$ where x_{k_ℓ} is a subsequence of x_k . If $\Omega := \mathbb{R}^n$ can be chosen, Γ is called *inner calm** at $\bar{x}$ for brevity. If, for given $u \in \mathbb{R}^n$, the above property holds just for all sequences $\bar{x} + t_k u_k =: x_k \in \Omega$ where $t_k \downarrow 0$ and $u_k \rightarrow u$, then Γ is referred to as *inner calm** at $\bar{x}$ in direction u w.r.t. Ω . Similarly as above, *inner calmness** at $\bar{x}$ in direction u is defined.

Inner calmness* is not very restrictive. Let us mention some situations when it is satisfied, see also Proposition 2.2 below.

- (i) In [4, Theorem 3.4], it was shown that polyhedral set-valued mappings are inner calm* at every point of the domain w.r.t. the domain. This result provides a certain lower/inner counterpart to Robinson's result on upper/outer Lipschitzness of polyhedral mappings from [34], and it is an easy consequence of

the famous Walkup-Wets result on Lipschitzness w.r.t. the domain of convex polyhedral mappings, see [42].

- (ii) In [4, Theorem 3.9], inner semicompactness and inner calmness* of a certain multiplier mapping associated with standard geometric constraints was established under suitable constraint qualifications. Let us mention that this multiplier mapping is very relevant for the analysis of the normal cone mapping associated with this constraint system. Moreover, (essentially) the same mapping also appears in applications of our second-order conditions to the most challenging optimization problems, see Remark 5.8.
- (iii) Finally, due to [9, Lemma 4.3], whenever Γ is *isolatedly calm* at all points $(\bar{x}, y)$ such that $y \in \Gamma(\bar{x})$ and inner semicompact at $\bar{x}$ w.r.t. $\text{dom } \Gamma$, then Γ is inner calm* at $\bar{x}$ w.r.t. $\text{dom } \Gamma$. Let us recall that Γ is called *isolatedly calm* at some point $(\bar{x}, \bar{y}) \in \text{gph } \Gamma$ if there exist $\kappa > 0$, $\varepsilon > 0$, and $\delta > 0$ such that

$$\Gamma(x) \cap \mathbb{U}_\varepsilon(\bar{y}) \subset \bar{y} + \kappa \|x - \bar{x}\| \mathbb{B}_1(0) \quad \forall x \in \mathbb{U}_\delta(\bar{x}).$$

It is well known that Γ is *isolatedly calm* at $(\bar{x}, \bar{y})$ if and only if $D\Gamma(\bar{x}, \bar{y})(0) = \{0\}$ holds, and the latter has been named *Levy-Rockafellar criterion* in the literature.

The main purpose of introducing inner calmness* was its role in certain calculus rules, particularly, the so-called image rule for tangents, see [4, Section 4]. In this paper, we further pursue these developments as inner calmness* is the essential assumption we rely on to derive calculus rules for second subderivatives (it is essential for one (of two) pattern of related results while the other one requires only the most basic assumptions).

More precisely, in Section 4, we will study a marginal function rule for the second subderivative, and we now state some preparatory results. Therefore, we choose a proper, lower semicontinuous function $\varphi: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$ and consider the associated marginal function $h: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ given by

$$h(x) := \inf_{y \in \mathbb{R}^m} \varphi(x, y) \quad \forall x \in \mathbb{R}^n. \quad (1)$$

Closely related to h are the set-valued mappings $\Upsilon: \mathbb{R}^n \times \mathbb{R} \rightrightarrows \mathbb{R}^m$ and $\Psi: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given by

$$\begin{aligned} \Upsilon(x, \alpha) &:= \{y \in \mathbb{R}^m \mid (x, y, \alpha) \in \text{epi } \varphi\} & \forall x \in \mathbb{R}^n, \alpha \in \mathbb{R}, \\ \Psi(x) &:= \arg \min_{y \in \mathbb{R}^m} \varphi(x, y) & \forall x \in \mathbb{R}^n. \end{aligned} \quad (2)$$

We further specify that $\Psi(x) := \emptyset$ is used for each $x \in \mathbb{R}^n$ such that $\varphi(x, y) = \infty$ holds for all $y \in \mathbb{R}^m$. Let us note that Ψ is often called solution mapping in the literature and a tool of major interest in parametric optimization. To the best of our knowledge, Υ is rarely used in this regard, see e.g. [9, Section 5.2]. However, we observe that $\Psi(x) = \Upsilon(x, h(x))$ is valid for all $x \in \mathbb{R}^n$ with $|h(x)| < \infty$, i.e., Υ and Ψ are related via the value function h . Additionally, one can easily check that $\text{dom } \Upsilon \subset \text{epi } h$ is valid. Finally, by lower semicontinuity of φ , $\text{gph } \Upsilon$ is naturally closed.

Note that, given a closed set $\Omega \subset \mathbb{R}^n$, the distance function fits into (1) by observing

$$\text{dist}(x, \Omega) = \inf_{y \in \mathbb{R}^n} (\|y - x\| + \delta_\Omega(y)) \quad \forall x \in \mathbb{R}^n \quad (3)$$

where $\delta_\Omega: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is the so-called indicator function of Ω which vanishes on Ω and is set to ∞ on $\mathbb{R}^n \setminus \Omega$. Clearly, the argmin mapping Ψ from (2) corresponds to the projection mapping $P_\Omega: \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ given by $P_\Omega(x) := \arg \min_{y \in \Omega} \|y - x\|$, $x \in \mathbb{R}^n$, in this case. Interestingly, P_Ω is inner calm* at every point belonging to Ω .

Proposition 2.2. *Given a closed set $\Omega \subset \mathbb{R}^n$, the projection mapping P_Ω is inner calm* at every $\bar{x} \in \Omega$. Moreover, for sequences $\bar{x} \neq x_k \rightarrow \bar{x}$ satisfying $\text{dist}(x_k, \Omega)/\|x_k - \bar{x}\| \rightarrow 0$, we even get for all $y_k \in P_\Omega(x_k)$,*

$$\frac{y_k - \bar{x}}{\|x_k - \bar{x}\|} - \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \rightarrow 0,$$

i.e., along a subsequence, y_k converges to $\bar{x}$ from the same direction as x_k .

Proof. Consider a sequence $x_k \rightarrow \bar{x} = P_\Omega(\bar{x})$. By [38, Example 1.20], $P_\Omega(x_k) \neq \emptyset$, and for every $y_k \in P_\Omega(x_k)$, we easily get

$$\|y_k - \bar{x}\| \leq \|y_k - x_k\| + \|x_k - \bar{x}\| \leq 2\|x_k - \bar{x}\|,$$

showing inner calmness* of P_Ω at $\bar{x}$. Moreover, the second claim now follows from $(y_k - \bar{x}) - (x_k - \bar{x}) = y_k - x_k$, which means $\|(y_k - \bar{x}) - (x_k - \bar{x})\| = \text{dist}(x_k, \Omega)$. $\square$

To see that the statement of Proposition 2.2 does not remain valid if $\bar{x} \notin \Omega$, one can consider $\bar{x}$ to be the center of the unit sphere $\Omega \subset \mathbb{R}^2$. Then, the sequence x_k can be chosen to approach $\bar{x}$ rapidly along slowly changing rays. For instance $x_k := 1/k^2(\cos(1/k), \sin(1/k)) \rightarrow 0$, while we have $y_k = (\cos(1/k), \sin(1/k)) \rightarrow (1, 0) =: \bar{y}$, $\|x_k - \bar{x}\| = 1/k^2$, and $\|y_k - \bar{y}\| \approx 1/k$. Since this is not essential in the further parts of this paper, we omit the details.

We proceed with some basic results for the marginal function and the mappings from (2).

Lemma 2.3. *Let $\varphi: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ be a proper, lower semicontinuous function and consider the associated mappings defined in (1) and (2). Fix $\bar{x} \in \text{dom } \Psi$ and let Υ be inner semicompact at $(\bar{x}, h(\bar{x}))$ w.r.t. $\text{dom } \Upsilon$. Then h is lower semicontinuous at $\bar{x}$ and, locally around $(\bar{x}, h(\bar{x}))$, the sets $\text{epi } h$ and $\text{dom } \Upsilon$ coincide and are closed.*

Proof. From Lemma 2.1 we know that $\text{dom } \Upsilon$ is closed locally around $(\bar{x}, h(\bar{x}))$ and $\text{dom } \Upsilon \subset \text{epi } h$ holds always true. Thus, all the claims follow once we show that near $(\bar{x}, h(\bar{x}))$ also the opposite inclusion holds. This is, however, an easy consequence of the local closedness of $\text{dom } \Upsilon$. Indeed, consider a closed neighborhood $U \subset \mathbb{R}^n \times \mathbb{R}$ of $(\bar{x}, h(\bar{x}))$ such that $\text{dom } \Upsilon \cap U$ is closed and let $(x, \alpha) \in \text{epi } h \cap U$. Clearly, by definition, $(x, \alpha) \in \text{dom } \Upsilon$ if $\alpha > h(x)$. If $\alpha = h(x)$, the definition of h guarantees the existence of a sequence y_k such that $\alpha = \lim_{k \rightarrow \infty} \varphi(x, y_k)$. Particularly, $(x, \varphi(x, y_k)) \in \text{dom } \Upsilon$ and since $(x, \varphi(x, y_k)) \rightarrow (x, \alpha)$, the local closedness of $\text{dom } \Upsilon$ yields $(x, \alpha) \in \text{dom } \Upsilon$. $\square$

Lemma 2.4. *Let $\varphi: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ be a proper, lower semicontinuous function, consider the associated mappings defined in (1) and (2), and fix $\bar{x} \in \text{dom } \Psi$ as well as $u \in \mathbb{R}^n$. If Ψ is inner semicompact at $\bar{x}$ (inner calm* at $\bar{x}$ in direction u), then Υ is inner semicompact at $(\bar{x}, \alpha)$ w.r.t. $\text{dom } \Upsilon$ for all $\alpha \in \mathbb{R}$ (inner calm* at $(\bar{x}, \alpha)$ in direction (u, μ) w.r.t. $\text{dom } \Upsilon$ for all $\alpha, \mu \in \mathbb{R}$).*

Proof. Assume that Ψ is inner semicompact at $\bar{x}$ and consider $\alpha \in \mathbb{R}$ as well as a sequence $(x_k, \alpha_k) \rightarrow (\bar{x}, \alpha)$ such that $(x_k, \alpha_k) \in \text{dom } \Upsilon$ for each $k \in \mathbb{N}$. By passing to a subsequence (without relabeling), inner semicompactness of Ψ yields the existence of $\bar{y} \in \mathbb{R}^m$ and, for each $k \in \mathbb{N}$, $y_k \in \Psi(x_k) = \Upsilon(x_k, h(x_k))$ such that $y_k \rightarrow \bar{y}$. Since $(x_k, \alpha_k) \in \text{dom } \Upsilon$, we have $\alpha_k \geq h(x_k)$ and, thus, $y_k \in \Upsilon(x_k, \alpha_k)$ for each $k \in \mathbb{N}$, i.e., inner semicompactness of Υ follows.

Assume now that Ψ is inner calm* at $\bar{x}$ in direction u and consider $\alpha \in \mathbb{R}$ as well as sequences $t_k \downarrow 0$ and $(u_k, \mu_k) \rightarrow (u, \mu)$ with $(\bar{x}, \alpha) + t_k(u_k, \mu_k) \in \text{dom } \Upsilon$ for each $k \in \mathbb{N}$. By passing to a subsequence (without relabeling), the directional inner calmness* of Ψ yields the existence of $\bar{y} \in \mathbb{R}^m$, $\kappa > 0$, and, for each $k \in \mathbb{N}$, $y_k \in \Psi(x_k) = \Upsilon(x_k, h(x_k))$ such that $y_k \rightarrow \bar{y}$ and

$$\|y_k - \bar{y}\| \leq t_k \kappa \|u_k\| \leq t_k \kappa \|(u_k, \mu_k)\| \quad \forall k \in \mathbb{N}.$$

For each natural number $k \in \mathbb{N}$, $\alpha + t_k \mu_k \geq h(\bar{x} + t_k u_k)$ follows as before, and we obtain $y_k \in \Upsilon(\bar{x} + t_k u_k, \alpha + t_k \mu_k)$, showing the claimed inner calmness* of Υ at $(\bar{x}, h(\bar{x}))$ in direction (u, μ) . $\square$

Due to the above lemmas, inner semicompactness of Υ (or Ψ) ensures lower semicontinuity of h and so it will be used as a standing assumption in our analysis.

It is worth noting that the assumptions of Lemma 2.4 imposed on Ψ are more coarse, since they apply whenever $x_k \rightarrow \bar{x}$ (from direction u) regardless of what happens with $h(x_k)$. They imply the corresponding assumptions on Υ for any $\alpha \in \mathbb{R}$. Note also that the statements are trivially satisfied for $\alpha < h(\bar{x})$ since there are no sequences $(x_k, \alpha_k) \in \text{dom } \Upsilon \subset \text{epi } h$ approaching such points $(\bar{x}, \alpha)$. Similar arguments apply to the issue of directional convergence in the image space. Since

$$T_{\text{dom } \Upsilon}(\bar{x}, h(\bar{x})) \subset T_{\text{epi } h}(\bar{x}, h(\bar{x})) = \text{epi } d h(\bar{x}),$$

only for $\mu \geq d h(\bar{x})(u)$, the inner calmness* of Υ is not trivially satisfied. In Section 4, we propose two analogous results, the coarser one based on Ψ and the finer one based on Υ .

Observe that Lemma 2.4 assumes inner semicompactness/inner calmness* of Ψ which might be stronger than the restrictions of these properties to $\text{dom } \Psi$. Exemplary, the following example illustrates that inner semicompactness of Ψ w.r.t. $\text{dom } \Psi$ is not enough to guarantee lower semicontinuity of the marginal function h at the point of interest, i.e., this assumption is too weak in order to imply the requirements of Lemma 2.3.

Example 2.5. Consider $\varphi: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ given by $\varphi(x, y) := e^{xy}$ for all $x, y \in \mathbb{R}$. Then we find

$$h(x) = \begin{cases} 1 & x = 0, \\ 0 & x \neq 0, \end{cases} \quad \Psi(x) = \begin{cases} \mathbb{R} & x = 0, \\ \emptyset & x \neq 0 \end{cases} \quad \forall x \in \mathbb{R}.$$

Particularly, Ψ is inner semicompact at $\bar{x} := 0$ w.r.t. $\text{dom } \Psi = \{\bar{x}\}$, but h fails to be lower semicontinuous at $\bar{x}$.

Note that the associated mapping Υ indeed fails to be inner semicompact at $(\bar{x}, 1)$ w.r.t. $\text{dom } \Upsilon$. In order to see this, choose $x_k := -1/k^2$ and $\alpha_k := 1 - 1/k$ for each $k \in \mathbb{N}$. Then we have $(x_k, \alpha_k) \rightarrow (\bar{x}, 1)$ and $\Upsilon(x_k, \alpha_k) = [x_k^{-1} \ln \alpha_k, \infty)$ for each $k \in \mathbb{N}$. Due to $x_k^{-1} \ln \alpha_k \rightarrow \infty$ as $k \rightarrow \infty$, Υ is not inner semicompact at $(\bar{x}, 1)$. $\square$

Finally, we note that lower semicontinuity of h is also implied by a related (restricted) inf-compactness assumption, see e.g. [14, Hypothesis 6.5.1] and [23, Definition 3.8]. Without stating precise definitions of these properties, let us mention that inf-compactness is clearly implied by inner semicompactness of Ψ , see also [2, Section 4.1]. On the other hand, our refined semicompactness assumption imposed on Υ turns out to be milder than (perhaps equivalent to) restricted inf-compactness. Indeed, for $\bar{x}$ with $|h(\bar{x})| < \infty$, let a sequence $(x_k, \alpha_k) \rightarrow (\bar{x}, h(\bar{x}))$ satisfy $(x_k, \alpha_k) \in \text{dom } \Upsilon$ for each $k \in \mathbb{N}$. Then, for each $k \in \mathbb{N}$ and by definition of Υ , we get the existence of y_k with $h(x_k) \leq \varphi(x_k, y_k) \leq \alpha_k$. Combining this with $\alpha_k \rightarrow h(\bar{x})$ and taking the restricted inf-compactness into consideration yields the existence of a nonempty, compact set A and, for large enough $k \in \mathbb{N}$, some $\hat{y}_k \in \Psi(x_k) \cap A \subset \Upsilon(x_k, \alpha_k) \cap A$. Thus, Υ is inner semicompact at $(\bar{x}, h(\bar{x}))$.

2.3. Second subderivative

The following definition of the second subderivative of a function is taken from [38, Definition 13.3].

Definition 2.6. Let $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be a lower semicontinuous function and fix $\bar{z} \in \mathbb{R}^n$ with $|h(\bar{z})| < \infty$ as well as $z^* \in \mathbb{R}^n$. The *second subderivative* of h at $\bar{z}$ for z^* is the function $d^2h(\bar{z}; z^*): \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ defined by

$$d^2h(\bar{z}; z^*)(w) := \liminf_{t \downarrow 0, w' \rightarrow w} \frac{h(\bar{z} + tw') - h(\bar{z}) - t\langle z^*, w' \rangle}{\frac{1}{2}t^2} \quad \forall w \in \mathbb{R}^n.$$

Right from Definition 2.6, we readily obtain the homogeneity properties

$$d^2(\alpha h)(\bar{z}; z^*)(w) = \alpha d^2h(\bar{z}; z^*/\alpha)(w) \quad \forall \alpha \in (0, \infty) \quad (4)$$

as well as the relations

$$\begin{aligned} d h(\bar{z})(w) > \langle z^*, w \rangle &\implies d^2h(\bar{z}; z^*)(w) = \infty, \\ d h(\bar{z})(w) < \langle z^*, w \rangle &\implies d^2h(\bar{z}; z^*)(w) = -\infty. \end{aligned} \quad (5)$$

In Section 3.2, we provide some more details regarding when $d^2h(\bar{z}; z^*)(w)$ is finite. Observe that whenever h is twice differentiable and $z^* = \nabla h(\bar{z})$, then the lower limit can be replaced by a limit in Definition 2.6 since

$$h(\bar{z} + tw') - h(\bar{z}) = t\nabla h(\bar{z})w' + \frac{1}{2}t^2\nabla^2h(\bar{z})(w', w') + o(t^2).$$

The following lemma will be important for some proofs.

Lemma 2.7. Let $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be a lower semicontinuous function and fix $\bar{z} \in \mathbb{R}^n$ with $|h(\bar{z})| < \infty$, $z^* \in \mathbb{R}^n$, and $w \in \mathbb{R}^n$. Then there are sequences $t_k \downarrow 0$ and $w_k \rightarrow w$ such that

$$d h(\bar{z})(w) = \lim_{k \rightarrow \infty} \frac{h(\bar{z} + t_k w_k) - h(\bar{z})}{t_k}, \quad (6a)$$

$$d^2 h(\bar{z}; z^*)(w) = \lim_{k \rightarrow \infty} \frac{h(\bar{z} + t_k w_k) - h(\bar{z}) - t_k \langle z^*, w_k \rangle}{\frac{1}{2} t_k^2}. \quad (6b)$$

Proof. Fix sequences $t_k \downarrow 0$ and $w_k \rightarrow w$ which satisfy (6b). By passing to a subsequence (without relabeling) we may assume that

$$\mu_k := (h(\bar{z} + t_k w_k) - h(\bar{z}))/t_k \rightarrow \mu \in \overline{\mathbb{R}}$$

with $\mu \geq d h(\bar{z})(w)$. If we have $\mu = d h(\bar{z})(w)$, the sequences t_k and w_k have the desired properties. If $\mu > d h(\bar{z})(w)$, consider sequences $\tilde{t}_k \downarrow 0$ and $\tilde{w}_k \rightarrow w$ such that

$$\tilde{\mu}_k := (h(\bar{z} + \tilde{t}_k \tilde{w}_k) - h(\bar{z}))/\tilde{t}_k \rightarrow d h(\bar{z})(w).$$

Note that, for some $\varepsilon > 0$ and all sufficiently large $k, \ell \in \mathbb{N}$, $\tilde{\mu}_k + \varepsilon < \mu_\ell$ holds. We define $\ell(1) := \min\{\ell \in \mathbb{N} \mid t_\ell \leq \tilde{t}_1\}$ and $\ell(k+1) := \min\{\ell \in \mathbb{N} \mid t_\ell \leq \tilde{t}_{k+1}, \ell > \ell(k)\}$ for each $k \in \mathbb{N}$. Thus, we have $t_{\ell(k)} \leq \tilde{t}_k$ for all $k \in \mathbb{N}$. Furthermore, due to $\langle z^*, w_{\ell(k)} - \tilde{w}_k \rangle \rightarrow 0$, the inequality $\tilde{\mu}_k - \langle z^*, \tilde{w}_k \rangle + \varepsilon/2 \leq \mu_{\ell(k)} - \langle z^*, w_{\ell(k)} \rangle$ is valid for large enough $k \in \mathbb{N}$. Together, this gives

$$d^2 h(\bar{z}; z^*)(w) = \lim_{k \rightarrow \infty} \frac{\mu_{\ell(k)} - \langle z^*, w_{\ell(k)} \rangle}{\frac{1}{2} t_{\ell(k)}} \geq \lim_{k \rightarrow \infty} \frac{\tilde{\mu}_k - \langle z^*, \tilde{w}_k \rangle}{\frac{1}{2} \tilde{t}_k} \geq d^2 h(\bar{z}; z^*)(w),$$

showing $(h(\bar{z} + \tilde{t}_k \tilde{w}_k) - h(\bar{z}) - \tilde{t}_k \langle z^*, \tilde{w}_k \rangle)/(\frac{1}{2} \tilde{t}_k^2) \rightarrow d^2 h(\bar{z}; z^*)(w)$ and completing the proof. $\square$

Sequences $t_k \downarrow 0$ and $w_k \rightarrow w$ in the sense of Lemma 2.7 will be said to recover $d h(\bar{z})(w)$ and $d^2 h(\bar{z}; z^*)(w)$ simultaneously.

In the proposition below, we summarize some elementary sum rules which address the second subderivative.

Proposition 2.8. (a) For a twice differentiable function $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$, a lower semicontinuous function $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, some point $\bar{z} \in \mathbb{R}^n$ with $|f(\bar{z})| < \infty$, and $z^* \in \mathbb{R}^n$, we have

$$d^2(f_0 + f)(\bar{z}; z^*)(w) = \nabla^2 f_0(\bar{z})(w, w) + d^2 f(\bar{z}; z^* - \nabla f_0(\bar{z}))(w) \quad \forall w \in \mathbb{R}^n.$$

(b) For two lower semicontinuous functions $h_1: \mathbb{R}^{n_1} \rightarrow \overline{\mathbb{R}}$ and $h_2: \mathbb{R}^{n_2} \rightarrow \overline{\mathbb{R}}$, $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ for $n := n_1 + n_2$ defined by $h(z) := h_1(z_1) + h_2(z_2)$ for all $z := (z_1, z_2) \in \mathbb{R}^n$, some point $\bar{z} := (\bar{z}_1, \bar{z}_2) \in \mathbb{R}^n$ with $|h(\bar{z})| < \infty$, and $z^* := (z_1^*, z_2^*) \in \mathbb{R}^n$, we have

$$d^2 h(\bar{z}; z^*)(w) \geq d^2 h_1(\bar{z}_1; z_1^*)(w_1) + d^2 h_2(\bar{z}_2; z_2^*)(w_2) \quad \forall w := (w_1, w_2) \in \mathbb{R}^n$$

provided the right-hand side is not $\infty - \infty$.

Proof. (a) Exploiting a second-order Taylor expansion of f_0 at $\bar{z}$, we find

$$\begin{aligned} d^2(f_0 + f)(\bar{z}; z^*)(w) &= \liminf_{t \downarrow 0, w' \rightarrow w} \left(\frac{\nabla f_0(\bar{z})w' + \frac{1}{2}t \nabla^2 f_0(\bar{z})(w', w') + o(t)}{\frac{1}{2}t} + \frac{f(\bar{z} + tw') - f(\bar{z}) - t\langle z^*, w' \rangle}{\frac{1}{2}t^2} \right) \\ &= \nabla^2 f_0(\bar{z})(w, w) + \liminf_{t \downarrow 0, w' \rightarrow w} \frac{f(\bar{z} + tw') - f(\bar{z}) - t\langle z^* - \nabla f_0(\bar{z}), w' \rangle}{\frac{1}{2}t^2} \\ &= \nabla^2 f_0(\bar{z})(w, w) + d^2 f(\bar{z}; z^* - \nabla f_0(\bar{z}))(w) \end{aligned}$$

which yields the claim.

(b) This follows immediately from the definition of the second subderivative, taking into account that $\liminf_{k \rightarrow \infty} (a_k + b_k) \geq \liminf_{k \rightarrow \infty} a_k + \liminf_{k \rightarrow \infty} b_k$ holds except for the indeterminate case $\infty - \infty$. $\square$

Applying Proposition 2.8 (a) with f being constantly zero, we find

$$d^2 f_0(\bar{z}; z^*)(w) = \begin{cases} \nabla^2 f_0(\bar{z})(w, w) & z^* = \nabla f_0(\bar{z}), \\ \infty & \langle z^* - \nabla f_0(\bar{z}), w \rangle < 0, \\ -\infty & \langle z^* - \nabla f_0(\bar{z}), w \rangle \geq 0, z^* \neq \nabla f_0(\bar{z}) \end{cases} \quad \forall w \in \mathbb{R}^n \quad (7)$$

for each twice differentiable function $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$, $\bar{z} \in \mathbb{R}^n$, and $z^* \in \mathbb{R}^n$, and this recovers [38, Exercise 13.8].

We end this section by providing formulas for the second subderivative of two convex functions which we will need later, namely the Euclidean norm and the maximum function. In both cases, we only consider z^* and w such that $\langle z^*, w \rangle$ equals the subderivative and z^* belongs to the subdifferential at the reference point, respectively, for otherwise the second subderivative attains only the values $\pm\infty$, see (5), Propositions 3.10 and Corollary 3.14, or [38, Proposition 13.5]. Moreover, noting that the norm is twice differentiable at all non-zero points and keeping (7) in mind, we restrict ourselves to the origin in the next lemma.

Lemma 2.9. *For each point $w \in \mathbb{R}^n$, we have $d\|\cdot\|(0)(w) = \|w\|$, and for each $z^* \in \partial\|\cdot\|(0) = \mathbb{B}_1(0)$ satisfying $\langle z^*, w \rangle = \|w\|$, we find $d^2\|\cdot\|(0; z^*)(w) = 0$.*

Proof. The formula for the subderivative is straightforward. Let us note that

$$d^2\|\cdot\|(0; z^*)(w) = \liminf_{t \downarrow 0, w' \rightarrow w} \frac{2(\|w'\| - \langle z^*, w' \rangle)}{t}.$$

Choosing $w' := w$, we recover 0, showing that $d^2\|\cdot\|(0; z^*)(w) \leq 0$. Since $\|z^*\| \leq 1$, the Cauchy-Schwarz inequality gives the estimate

$$\|w'\| - \langle z^*, w' \rangle \geq \|w'\|(1 - \|z^*\|) \geq 0,$$

i.e., $d^2\|\cdot\|(0; z^*)(w) = 0$. $\square$

Lemma 2.10. Let $\text{vecmax}: \mathbb{R}^n \rightarrow \mathbb{R}$ be the function given by

$$\text{vecmax}(z) := \max\{z_1, \dots, z_n\} \quad \forall z \in \mathbb{R}^n.$$

For fixed $\bar{z} \in \mathbb{R}^n$ and $w \in \mathbb{R}^n$, we define

$$I(\bar{z}) := \{i \in \{1, \dots, n\} \mid \bar{z}_i = \text{vecmax}(\bar{z})\}.$$

Then we have $d \text{vecmax}(\bar{z})(w) = \max\{w_i \mid i \in I(\bar{z})\}$, and for each $z^* \in \partial \text{vecmax}(\bar{z})$ satisfying $\langle z^*, w \rangle = \max\{w_i \mid i \in I(\bar{z})\}$, we find $d^2 \text{vecmax}(\bar{z}; z^*)(w) = 0$.

Proof. The formula for the first subderivative is well known and can be distilled from [38, Exercise 8.31]. Thus, let us pick $\bar{z}, w \in \mathbb{R}^n$ and

$$z^* \in \partial \text{vecmax}(\bar{z}) = \{z^* \in \mathbb{R}_+^n \mid \sum_{i=1}^n z_i^* = 1, z_i^* = 0 \ \forall i \notin I(\bar{z})\}$$

satisfying $\langle z^*, w \rangle = \max\{w_i \mid i \in I(\bar{z})\}$. For an arbitrary index $i_0 \in I(\bar{z})$, we find

$$\begin{aligned} d^2 \text{vecmax}(\bar{z}; z^*)(w) &= \liminf_{w' \rightarrow w, t \downarrow 0} \frac{\text{vecmax}(\bar{z} + tw') - \bar{z}_{i_0} - t\langle z^*, w' \rangle}{\frac{1}{2}t^2} \\ &= \liminf_{w' \rightarrow w, t \downarrow 0} \frac{\max\{tw'_i \mid i \in I(\bar{z})\} - t\langle z^*, w' \rangle}{\frac{1}{2}t^2} \\ &= \liminf_{w' \rightarrow w, t \downarrow 0} \frac{\max\{w'_i \mid i \in I(\bar{z})\} - \langle z^*, w' \rangle}{\frac{1}{2}t}. \end{aligned}$$

Choosing $w' := w$, we recover 0, showing that $d^2 \text{vecmax}(\bar{z}; z^*)(w) \leq 0$. On the other hand, we have

$$\begin{aligned} \max\{w'_i \mid i \in I(\bar{z})\} - \langle z^*, w' \rangle &= \max\{w'_i \mid i \in I(\bar{z})\} - \sum_{i \in I(\bar{z})} z_i^* w'_i \\ &\geq \max\{w'_i \mid i \in I(\bar{z})\} \left(1 - \sum_{i \in I(\bar{z})} z_i^*\right) = 0, \end{aligned}$$

and $d^2 \text{vecmax}(\bar{z}; z^*)(w) = 0$ follows. $\square$

2.4. Second-order optimality conditions

The following result, which can be found in [12, Proposition 3.100], provides second-order necessary and sufficient optimality conditions for the unconstrained minimization of proper functions.

Proposition 2.11. Given a proper, lower semicontinuous function $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ and $\bar{z} \in \mathbb{R}^n$ with $|h(\bar{z})| < \infty$, the following statements hold.

- (a) If $\bar{z}$ is a local minimizer of h , then $0 \in \widehat{\partial}h(\bar{z})$ and $d^2h(\bar{z}; 0)(w) \geq 0$ for all $w \in \mathbb{R}^n$.
- (b) Having $d^2h(\bar{z}; 0)(w) > 0$ for all $w \in \mathbb{R}^n \setminus \{0\}$ is equivalent to having the existence of $\varepsilon > 0$ and $\delta > 0$ such that

$$h(z) \geq h(\bar{z}) + \varepsilon \|z - \bar{z}\|^2 \quad \forall z \in \mathbb{B}_\delta(\bar{z}). \quad (8)$$

Particularly, $\bar{z}$ is a strict local minimizer of h .

Note that the assumptions of Proposition 2.11(b) also imply $dh(\bar{z})(w) \geq \langle 0, w \rangle = 0$ for each $w \in \mathbb{R}^n \setminus \{0\}$, see [38, Proposition 13.5]. From (5), $dh(\bar{z})(w) > 0$ for some $w \in \mathbb{R}^n$ yields $d^2h(\bar{z}; 0)(w) = \infty$. In Section 3.2, we obtain a stronger statement, namely that $d^2h(\bar{z}; 0)(w) > 0$ implies $0 \in \tilde{\partial}^p h(\bar{z}; w)$ (which in turn yields $dh(\bar{z})(w) \geq 0$), where $\tilde{\partial}^p h(\bar{z}; w)$ denotes the directional proximal pre-subdifferential of h at $\bar{z}$ in direction w , see Definition 3.7 below.

Let us mention that whenever (8) holds for some $\varepsilon > 0$ and $\delta > 0$, then we say that h satisfies a second-order growth condition at $\bar{z}$.

Consider now the problem of minimizing a twice differentiable function $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$ over a closed set $S \subset \mathbb{R}^n$ and set $h := f_0 + \delta_S$ for the indicator function δ_S of S . By closedness of S , δ_S is lower semicontinuous, and, obviously, δ_S is proper. Taking into account Proposition 2.8, we find

$$d^2h(\bar{z}; z^*)(w) = \nabla^2 f_0(\bar{z})(w, w) + d^2\delta_S(\bar{z}; z^* - \nabla f_0(\bar{z}))(w)$$

for all $\bar{z} \in S$, $z^* \in \mathbb{R}^n$, and $w \in \mathbb{R}^n$. Proposition 2.11 thus yields the following result.

Proposition 2.12. *Given a twice differentiable function $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$ and a closed set $S \subset \mathbb{R}^n$, the following statements hold.*

(a) *If $\bar{z} \in \mathbb{R}^n$ is a local minimizer of f_0 over S , then $0 \in \nabla f_0(\bar{z}) + \widehat{N}_S(\bar{z})$ and*

$$\nabla^2 f_0(\bar{z})(w, w) + d^2\delta_S(\bar{z}; -\nabla f_0(\bar{z}))(w) \geq 0 \quad \forall w \in \mathbb{R}^n.$$

(b) *Having*

$$\nabla^2 f_0(\bar{z})(w, w) + d^2\delta_S(\bar{z}; -\nabla f_0(\bar{z}))(w) > 0 \quad \forall w \in \mathbb{R}^n \setminus \{0\}$$

is equivalent to having the existence of $\varepsilon > 0$ and $\delta > 0$ such that

$$f_0(z) \geq f_0(\bar{z}) + \varepsilon \|z - \bar{z}\|^2 \quad \forall z \in S \cap \mathbb{B}_\delta(\bar{z}). \quad (9)$$

Particularly, $\bar{z}$ is a strict local minimizer of f_0 over S .

Note again that $0 \in \widehat{\partial}(f_0 + \delta_S)(\bar{z}) = \nabla f_0(\bar{z}) + \widehat{N}_S(\bar{z})$ is implied by the requirements of Proposition 2.12 (b) and [38, Exercise 8.8]. Moreover, Proposition 3.6 from below, equalling [7, Proposition 2.18], in this case yields that, for each non-zero $w \in \mathbb{R}^n$, either $w \notin T_S(\bar{z})$ or $0 \in \nabla f_0(\bar{z}) + \widehat{N}_S^p(\bar{z}; w)$, which also gives $\langle \nabla f_0(\bar{z}), w \rangle \geq 0$. Here, $\widehat{N}_S^p(\bar{z}; w)$ stands for the proximal pre-normal cone to S at $\bar{z}$ in direction w , see Definition 3.4 below. Following the proof of Proposition 2.11, we even find $d^2\delta_S(\bar{z}; -\nabla f_0(\bar{z}))(0) = 0$ which gives $0 \in \nabla f_0(\bar{z}) + \widehat{N}_S^p(\bar{z})$, see Proposition 3.6 again.

Again, whenever there are $\varepsilon > 0$ and $\delta > 0$ such that (9) holds, we say that the optimization problem $\min\{f_0(z) \mid z \in S\}$ satisfies the second-order growth condition at $\bar{z}$.

3. Directional proximal normal cones and subdifferentials

3.1. Second subderivative of the indicator function and the directional proximal normal cone

Motivated by the second-order optimality conditions for constrained optimization problems from Proposition 2.12, we look deeper into the second subderivative of the

indicator function $\delta_S: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ for a closed set $S \subset \mathbb{R}^n$. As noted in [7], given $\bar{z} \in S$ and $z^* \in \mathbb{R}^n$, we get

$$\begin{aligned} d^2\delta_S(\bar{z}; z^*)(w) &= \liminf_{t \downarrow 0, w' \rightarrow w} \frac{\delta_S(\bar{z} + tw') - \delta_S(\bar{z}) - t\langle z^*, w' \rangle}{\frac{1}{2}t^2} \\ &= \liminf_{\substack{t \downarrow 0, w' \rightarrow w, \\ \bar{z} + tw' \in S}} \frac{-2\langle z^*, w' \rangle}{t} \quad \forall w \in T_S(\bar{z}) \end{aligned} \quad (10)$$

directly from Definition 2.6. It is immediate that $w \notin T_S(\bar{z})$ or $\langle z^*, w \rangle < 0$ implies $d^2\delta_S(\bar{z}; z^*)(w) = \infty$, see Proposition 3.6 as well. Additionally, for $\bar{z} \in \text{int } S$, we obviously have

$$d^2\delta_S(\bar{z}; z^*)(w) = \begin{cases} \infty & \langle z^*, w \rangle < 0, \\ 0 & z^* = 0, \\ -\infty & \text{otherwise} \end{cases} \quad \forall z^* \in \mathbb{R}^n, \forall w \in \mathbb{R}^n.$$

Moreover, since $\delta_S = \alpha\delta_S$ for $\alpha > 0$, the homogeneity property (4) can be restated as

$$d^2\delta_S(\bar{z}; \alpha z^*)(w) = \alpha d^2\delta_S(\bar{z}; z^*)(w) \quad \forall \alpha \in (0, \infty). \quad (11)$$

Let us investigate a simple example where the second subderivative of an indicator function at some boundary point can be calculated easily.

Example 3.1. A direct calculation gives

$$d^2\delta_{\mathbb{R}_-}(0; z^*)(w) = \begin{cases} \infty & w > 0 \text{ or } z^*w < 0, \\ 0 & z^* \geq 0, w \leq 0, z^*w = 0, \\ -\infty & \text{otherwise} \end{cases} \quad \forall z^* \in \mathbb{R}, \forall w \in \mathbb{R}.$$

The following simple calculus rules are consequences of (10).

Lemma 3.2. Let $S_1, \dots, S_\ell \subset \mathbb{R}^n$ be closed sets, set $S := \bigcup_{i=1}^\ell S_i$, and fix $\bar{z} \in S$ as well as $z^* \in \mathbb{R}^n$. Then we have

$$d^2\delta_S(\bar{z}; z^*)(w) = \inf_{i \in J(\bar{z}; w)} d^2\delta_{S_i}(\bar{z}; z^*)(w) \quad \forall w \in \mathbb{R}^n$$

where $J(\bar{z}; w) := \{i \in \{1, \dots, \ell\} \mid \bar{z} \in S_i, w \in T_{S_i}(\bar{z})\}$. If $S_1, \dots, S_\ell$ are convex polyhedral sets, we have

$$d^2\delta_S(\bar{z}; z^*)(w) = \inf_{i \in J(\bar{z}; w)} \delta_{K_{S_i}(\bar{z}; z^*)}(w) \quad \forall w \in \mathbb{R}^n$$

for all $z^* \in \bigcap_{i \in J(\bar{z}; w)} N_{S_i}(\bar{z}; w)$, where $K_{S_i}(\bar{z}; z^*) := T_{S_i}(\bar{z}) \cap \{z^*\}^\perp$ is used for each $i \in J(\bar{z}; w)$.

Proof. Let us define $J(\bar{z}) := \{i \in \{1, \dots, \ell\} \mid \bar{z} \in S_i\}$. Then, from (10), we find

$$d^2\delta_S(\bar{z}; z^*)(w) = \inf_{i \in J(\bar{z})} d^2\delta_{S_i}(\bar{z}; z^*)(w) \quad \forall w \in \mathbb{R}^n.$$

Note that for each $i \in J(\bar{z}) \setminus J(\bar{z}; w)$, we have $w \notin T_{S_i}(\bar{z})$ and, thus, $d^2\delta_{S_i}(\bar{z}; z^*)(w) = \infty$, i.e., the indices from $J(\bar{z}) \setminus J(\bar{z}; w)$ can be discarded from the infimum.

In order to prove the formula for the union of convex polyhedral sets, we fix $w \in \mathbb{R}^n$ arbitrarily. Applying [38, Exercise 13.17] gives

$$d^2\delta_{S_i}(\bar{z}; z^*)(w) = \delta_{K_{S_i}(\bar{z}; z^*)}(w)$$

for all $z^* \in N_{S_i}(\bar{z})$ and $i \in J(\bar{z}; w)$. Due to $N_{S_i}(\bar{z}; w) \subset N_{S_i}(\bar{z})$ for all $i \in J(\bar{z}; w)$, the stated formula is valid for all $z^* \in \bigcap_{i \in J(\bar{z}; w)} N_{S_i}(\bar{z}; w)$ which gives the claim. $\square$

A related result can be found in [40, Proposition 3.2]. Let us point out that Lemma 3.2 together with Proposition 3.6 from below shows that the second subderivative of an indicator associated with a union of finitely many convex polyhedral sets takes only value 0 as soon as it is finite.

Lemma 3.3. *For $\ell \in \mathbb{N}$ and $n_1, \dots, n_\ell \in \mathbb{N}$, we fix closed sets $S_i \subset \mathbb{R}^{n_i}$ and points $\bar{z}_i \in S_i$, $i = 1, \dots, \ell$. Then, for arbitrary $z_i^*, w_i \in \mathbb{R}^{n_i}$, $i = 1, \dots, \ell$, we have*

$$d^2\delta_{S_1 \times \dots \times S_\ell}((\bar{z}_1, \dots, \bar{z}_\ell); (z_1^*, \dots, z_\ell^*))(w_1, \dots, w_\ell) \geq \sum_{i=1}^{\ell} d^2\delta_{S_i}(\bar{z}_i; z_i^*)(w_i)$$

provided the right-hand side does not contain summands of type $-\infty$ and ∞ simultaneously.

It turns out that there is a connection between the second subderivative of a set indicator and a so-called directional proximal normal cone to the same set, which has been defined in [7, Definition 2.8].

Definition 3.4. For a given closed set $S \subset \mathbb{R}^n$, a point $\bar{z} \in S$, and a direction $w \in T_S(\bar{z})$, we define the *proximal pre-normal cone* to S in direction w at $\bar{z}$ as

$$\widehat{\mathcal{N}}_S^p(\bar{z}; w) := \left\{ z^* \in \mathbb{R}^n \left| \begin{array}{l} \exists \gamma > 0, \forall t_k \downarrow 0, \forall w_k \rightarrow w \text{ such that } \bar{z} + t_k w_k \in S \forall k \in \mathbb{N}: \\ \langle z^*, w_k \rangle \leq \gamma t_k \|w_k\|^2 \text{ for all sufficiently large } k \in \mathbb{N} \end{array} \right. \right\},$$

and the proximal normal cone to S at $\bar{z}$ in direction w as

$$\widehat{N}_S^p(\bar{z}; w) := \widehat{\mathcal{N}}_S^p(\bar{z}; w) \cap \{w\}^\perp.$$

In case where $w \notin T_S(\bar{z})$, we set $\widehat{\mathcal{N}}_S^p(\bar{z}; w) := \widehat{N}_S^p(\bar{z}; w) := \emptyset$.

In [7], the proximal pre-normal cone has been defined in a different way without sequences in order to visualize its close relationship to the standard proximal normal cone. In the lemma below, we show that both definitions are equivalent.

Lemma 3.5. *Given a closed set $S \subset \mathbb{R}^n$, a point $\bar{z} \in S$, and a direction $w \in T_S(\bar{z})$, we have*

$$\widehat{\mathcal{N}}_S^p(\bar{z}; w) = \left\{ z^* \in \mathbb{R}^n \left| \begin{array}{l} \exists \gamma, \delta, \rho > 0, \forall z \in S \cap (\bar{z} + \mathbb{V}_{\delta, \rho}(w)) : \\ \langle z^*, z - \bar{z} \rangle \leq \gamma \|z - \bar{z}\|^2 \end{array} \right. \right\},$$

where, for each $\delta, \rho > 0$, $\mathbb{V}_{\delta, \rho}(w) \subset \mathbb{R}^n$ is a so-called *directional neighborhood* of w and given by means of

$$\mathbb{V}_{\delta, \rho}(w) := \{w' \in \mathbb{B}_\delta(0) \mid \|\|w\|w' - \|w'\|w\| \leq \rho \|w'\| \|w\|\}.$$

Proof. For $w := 0$, the equivalence follows easily since the directional neighborhood coincides with a conventional neighborhood. In this case, both formulas recover the classical proximal normal cone to S at $\bar{z}$. Indeed, the inclusion $\supset$ can be easily shown by a direct calculation.

On the other hand, suppose that, for some $z^* \in \widehat{\mathcal{N}}_S^p(\bar{z}; 0)$, there is a sequence $z_k \rightarrow \bar{z}$ such that $z_k \in S$ and $\langle z^*, z_k - \bar{z} \rangle > k \|z_k - \bar{z}\|^2$ for all $k \in \mathbb{N}$. Then $z_k \neq \bar{z}$ for all $k \in \mathbb{N}$, and we can set $t_k := \|z_k - \bar{z}\|^{1/2}$ and $w_k := (z_k - \bar{z})/t_k$ for each $k \in \mathbb{N}$ in order to get $t_k \downarrow 0$, $w_k \rightarrow 0$, as well as $\bar{z} + t_k w_k \in S$ and $\langle z^*, w_k \rangle > k t_k \|w_k\|^2$ for each $k \in \mathbb{N}$ contradicting $z^* \in \widehat{\mathcal{N}}_S^p(\bar{z}; 0)$. Thus, let us assume that $w \neq 0$.

Fix $z^* \in \widehat{\mathcal{N}}_S^p(\bar{z}; w)$. Suppose that for each $k \in \mathbb{N}$, there is $z_k \in S \cap (\bar{z} + \mathbb{V}_{1/k, 1/k}(w))$ such that $z_k \rightarrow \bar{z}$ and $\langle z^*, z_k - \bar{z} \rangle > k \|z_k - \bar{z}\|^2$. This is only possible if $z_k \neq \bar{z}$ for each $k \in \mathbb{N}$, so we can set $t_k := \|z_k - \bar{z}\|/\|w\|$ and $w_k := \|w\|(z_k - \bar{z})/\|z_k - \bar{z}\|$ which gives $z_k = \bar{z} + t_k w_k$ for each $k \in \mathbb{N}$. Furthermore, we find $t_k \downarrow 0$ and $w_k \rightarrow w$ by definition of the directional neighborhood. By construction, we have $t_k \langle z^*, w_k \rangle > k t_k^2 \|w_k\|^2$ for each $k \in \mathbb{N}$ contradicting the definition of the proximal pre-normal cone. Thus, the inclusion $\subset$ has been shown. The proof of the converse inclusion is analogous. $\square$

It is clear from Lemma 3.5 that the proximal pre-normal cone is, in general, larger than the classical proximal normal cone, i.e., $\widehat{\mathcal{N}}_S^p(\bar{z}) \subset \widehat{\mathcal{N}}_S^p(\bar{z}; w)$ holds for each $\bar{z} \in S$ and $w \in T_S(\bar{z})$. Moreover, for any such $\bar{z}$ and w as well as each vector $z^* \in \mathbb{R}^n$ satisfying $\langle z^*, w \rangle < 0$, Definition 3.4 yields $z^* \in \widehat{\mathcal{N}}_S^p(\bar{z}; w)$. In fact, we even have the estimates

$$\{z^* \in \mathbb{R}^n \mid \langle z^*, w \rangle < 0\} \subset \widehat{\mathcal{N}}_S^p(\bar{z}; w) \subset \{z^* \in \mathbb{R}^n \mid \langle z^*, w \rangle \leq 0\}. \quad (12)$$

Consequently, if a vector $z^* \in \widehat{\mathcal{N}}_S^p(\bar{z}; w)$ satisfies $\langle z^*, w \rangle < 0$, it does not provide much useful information. Hence, it is natural to intersect the directional proximal pre-normal cone with the annihilator of w in order to define the directional proximal normal cone.

As shown in [7, Proposition 2.9], both $\widehat{\mathcal{N}}_S^p(\bar{z}; w)$ and $\widehat{N}_S^p(\bar{z}; w)$ are convex cones for $\bar{z} \in S$ and $w \in T_S(\bar{z})$. Furthermore, one has the estimates

$$\widehat{\mathcal{N}}_S^p(\bar{z}) \cap \{w\}^\perp \subset \widehat{N}_S^p(\bar{z}; w) \subset \widehat{N}_{T_S(\bar{z})}^p(w) \subset N_{T_S(\bar{z})}(w) \subset N_S(\bar{z}; w). \quad (13)$$

In particular, when S is a closed convex set, we get

$$\widehat{N}_S^p(\bar{z}; w) = N_S(\bar{z}; w) = N_S(\bar{z}) \cap \{w\}^\perp = N_{T_S(\bar{z})}(w) \quad (14)$$

for any such $\bar{z}$ and w .

In [7, Proposition 2.18], the following result was proven.

Proposition 3.6. *Consider a closed set $S \subset \mathbb{R}^n$, $\bar{z} \in S$, $z^* \in \mathbb{R}^n$, and $w \in \mathbb{R}^n$. Then the following statements hold.*

- (a) *If $w \notin T_S(\bar{z})$ or $\langle z^*, w \rangle < 0$, then $d^2 \delta_S(\bar{z}; z^*)(w) = \infty$.*
- (b) *For $w \in T_S(\bar{z})$, we have $d^2 \delta_S(\bar{z}; z^*)(w) > -\infty$ if and only if $z^* \in \widehat{\mathcal{N}}_S^p(\bar{z}; w)$.*
- (c) *If $d^2 \delta_S(\bar{z}; z^*)(w)$ is finite, then $z^* \in \widehat{N}_S^p(\bar{z}; w)$.*

3.2. The directional proximal subdifferential

In this subsection, we interrelate the second subderivative of a lower semicontinuous function with a new directional proximal subdifferential which is introduced below.

Definition 3.7. Given a lower semicontinuous function $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, a point $\bar{z} \in \mathbb{R}^n$ such that $|h(\bar{z})| < \infty$, and a direction $w \in \mathbb{R}^n$ such that $|\mathrm{d}h(\bar{z})(w)| < \infty$, we define the *proximal pre-subdifferential* of h at $\bar{z}$ in direction w as

$$\tilde{\partial}^p h(\bar{z}; w) := \{z^* \in \mathbb{R}^n \mid \mathrm{d}^2 h(\bar{z}; z^*)(w) > -\infty\},$$

and the *proximal subdifferential* of h at $\bar{z}$ in direction w as

$$\hat{\partial}^p h(\bar{z}; w) := \tilde{\partial}^p h(\bar{z}; w) \cap \{z^* \in \mathbb{R}^n \mid \mathrm{d}h(\bar{z})(w) = \langle z^*, w \rangle\}.$$

If $|\mathrm{d}h(\bar{z})(w)| = \infty$, we set $\tilde{\partial}^p h(\bar{z}; w) := \hat{\partial}^p h(\bar{z}; w) := \emptyset$. Finally, for some $\omega \in \mathbb{R}$, the sets

$$\begin{aligned} \tilde{\partial}_g^p h(\bar{z}; (w, \omega)) &:= \{z^* \in \mathbb{R}^n \mid (z^*, -1) \in \hat{\mathcal{N}}_{\text{epi } h}^p((\bar{z}, h(\bar{z})); (w, \omega))\}, \\ \hat{\partial}_g^p h(\bar{z}; (w, \omega)) &:= \{z^* \in \mathbb{R}^n \mid (z^*, -1) \in \hat{N}_{\text{epi } h}^p((\bar{z}, h(\bar{z})); (w, \omega))\} \end{aligned}$$

are referred to as the geometric proximal pre-subdifferential and subdifferential of h at $\bar{z}$ in direction (w, ω) , respectively.

Note that in case $\mathrm{d}h(\bar{z})(w) = -\infty$, we have $\langle z^*, w \rangle > \mathrm{d}h(\bar{z})(w)$, so that (5) gives $\mathrm{d}^2 h(\bar{z}; z^*)(w) = -\infty$, i.e., $\tilde{\partial}^p h(\bar{z}; w) = \emptyset$ would also follow from the defining relation of the proximal pre-subdifferential. On the other hand, in case $\mathrm{d}h(\bar{z})(w) = \infty$, we find $\mathrm{d}^2 h(\bar{z}; z^*)(w) = \infty$ for each $z^* \in \mathbb{R}^n$ from (5), and the defining relation of the proximal pre-subdifferential would give us the whole space, but we still define the set $\tilde{\partial}^p h(\bar{z}; w)$ to be empty. Note that this parallels the definition of the proximal pre-normal cone in a direction which is not tangent, see Definition 3.4. Particularly, if $\mathrm{d}h(\bar{z})(w) = \infty$, we get $\tilde{\partial}_g^p h(\bar{z}; (w, \omega)) = \hat{\partial}_g^p h(\bar{z}; (w, \omega)) = \emptyset$ for all $\omega \in \mathbb{R}$ since $T_{\text{epi } h}(\bar{z}, h(\bar{z})) = \text{epi } \mathrm{d}h(\bar{z})$, i.e., there is no $\omega \in \mathbb{R}$ with $(w, \omega) \in T_{\text{epi } h}(\bar{z}, h(\bar{z}))$.

Note that if $|\mathrm{d}h(\bar{z})(w)| < \infty$, then (5) yields

$$\{z^* \in \mathbb{R}^n \mid \mathrm{d}h(\bar{z})(w) > \langle z^*, w \rangle\} \subset \tilde{\partial}^p h(\bar{z}; w) \subset \{z^* \in \mathbb{R}^n \mid \mathrm{d}h(\bar{z})(w) \geq \langle z^*, w \rangle\}, \quad (15)$$

which motivates the definition of the proximal subdifferential of h at $\bar{z}$ in the direction w .

Similar as in Lemma 3.5, we can show that the proximal pre-subdifferential admits an alternative representation via directional neighborhoods which is free of sequences.

Lemma 3.8. Given a lower semicontinuous function $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, a point $\bar{z} \in \mathbb{R}^n$ such that $|h(\bar{z})| < \infty$, and a direction $w \in \mathbb{R}^n$ such that $|\mathrm{d}h(\bar{z})(w)| < \infty$, we have

$$\tilde{\partial}^p h(\bar{z}; w) = \left\{ z^* \in \mathbb{R}^n \mid \begin{array}{l} \exists \gamma, \delta, \rho > 0, \forall z \in \bar{z} + \mathbb{V}_{\delta, \rho}(w): \\ h(z) \geq h(\bar{z}) + \langle z^*, z - \bar{z} \rangle - \gamma \|z - \bar{z}\|^2 \end{array} \right\}.$$

Proof. For $w := 0$, both definitions are equivalent by definition of the second subderivative since the directional neighborhood reduces to a classical neighborhood. In this case, both definitions recover the classical proximal subdifferential of h at $\bar{z}$. Indeed, the inclusion $\supset$ can be easily shown by a direct calculation. On the other hand, suppose that, for some $z^* \in \tilde{\partial}^p h(\bar{z}; 0)$, there is a sequence $z_k \rightarrow \bar{z}$ such that

$$h(z_k) - h(\bar{z}) - \langle z^*, z_k - \bar{z} \rangle < -k \|z_k - \bar{z}\|^2 \quad \forall k \in \mathbb{N}.$$

Then $z_k \neq \bar{z}$ for all $k \in \mathbb{N}$, and we can set either $t_k := \|z_k - \bar{z}\|^{1/2}$ and $w_k := (z_k - \bar{z})/t_k$ provided $k\|z_k - \bar{z}\| \rightarrow \infty$ holds at least along a subsequence or $t_k := k^{1/3}\|z_k - \bar{z}\|$ and $w_k := (z_k - \bar{z})/t_k$ provided the expression $k\|z_k - \bar{z}\|$ remains bounded. In both cases, $t_k \downarrow 0$, $w_k \rightarrow 0$, as well as

$$(h(\bar{z} + t_k w_k) - h(\bar{z}) - t_k \langle z^*, w_k \rangle) / t_k^2 < -k \|z_k - \bar{z}\|^2 / t_k^2 \quad \forall k \in \mathbb{N}.$$

Since $k\|z_k - \bar{z}\|^2 / t_k^2 \rightarrow \infty$ in both cases (at least along a subsequence), this contradicts $z^* \in \tilde{\partial}^p h(\bar{z}; 0)$. Thus, let us assume that $w \neq 0$.

Fix $z^* \in \tilde{\partial}^p h(\bar{z}; w)$. Suppose that for each $k \in \mathbb{N}$, there is $z_k \in \bar{z} + \mathbb{V}_{1/k, 1/k}(w)$ such that $z_k \rightarrow \bar{z}$ and $h(z_k) < h(\bar{z}) + \langle z^*, z_k - \bar{z} \rangle - k\|z_k - \bar{z}\|^2$. This gives $z_k \neq \bar{z}$ for each $k \in \mathbb{N}$, so we can set $t_k := \|z_k - \bar{z}\|/\|w\|$ and $w_k := \|w\|(z_k - \bar{z})/\|z_k - \bar{z}\|$ which gives $z_k = \bar{z} + t_k w_k$ for each $k \in \mathbb{N}$. By construction, we find

$$\frac{h(\bar{z} + t_k w_k) - h(\bar{z}) - t_k \langle z^*, w_k \rangle}{t_k^2 \|w_k\|^2} < -k$$

for each $k \in \mathbb{N}$. Multiplying this with $2\|w_k\|^2$ while noting that $\|w_k\| \rightarrow \|w\| \neq 0$, taking the limit $k \rightarrow \infty$ gives $d^2 h(\bar{z}; z^*)(w) = -\infty$ which contradicts the definition of the proximal pre-subdifferential. Thus, the inclusion $\subset$ has been shown. The converse relation follows in analogous fashion. $\square$

By construction, the directional proximal pre-normal and normal cone from Definition 3.4 to some closed set correspond to the directional proximal pre-subdifferential and subdifferential from Definition 3.7 of the associated indicator function, respectively.

Example 3.9. Let $S \subset \mathbb{R}^n$ be a closed set and fix $\bar{z} \in S$. One can easily check that for arbitrary $w \in \mathbb{R}^n$, we have

$$|d\delta_S(\bar{z})(w)| < \infty \iff d\delta_S(\bar{z})(w) = 0 \iff w \in T_S(\bar{z}),$$

and $\{z^* \in \mathbb{R}^n \mid d\delta_S(\bar{z})(w) = \langle z^*, w \rangle\} = \{w\}^\perp$ follows for each $w \in T_S(\bar{z})$. Putting this together with Lemmas 3.5 and 3.8, we immediately see

$$\tilde{\partial}^p \delta_S(\bar{z}; w) = \hat{\mathcal{N}}_S^p(\bar{z}; w), \quad \hat{\partial}^p \delta_S(\bar{z}; w) = \hat{N}_S^p(\bar{z}; w) \quad \forall w \in T_S(\bar{z}),$$

and $\tilde{\partial}^p \delta_S(\bar{z}; w) = \hat{\partial}^p \delta_S(\bar{z}; w) = \emptyset$ if $w \notin T_S(\bar{z})$. $\square$

Subsequently, we state an analogy of Proposition 3.6 which is an immediate consequence of (5) and Definition 3.7.

Proposition 3.10. *Given a lower semicontinuous function $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, a point $\bar{z} \in \mathbb{R}^n$ such that $|h(\bar{z})| < \infty$, $z^* \in \mathbb{R}^n$, and a direction $w \in \mathbb{R}^n$, the following statements hold.*

- (a) *If $d h(\bar{z})(w) > \langle z^*, w \rangle$, then $d^2 h(\bar{z}; z^*)(w) = \infty$.*
- (b) *If $d h(\bar{z})(w) < \infty$, we have $d^2 h(\bar{z}; z^*)(w) > -\infty$ if and only if $z^* \in \tilde{\partial}^p h(\bar{z}; w)$.*
- (c) *If $d^2 h(\bar{z}; z^*)(w)$ is finite, then $z^* \in \widehat{\partial}^p h(\bar{z}; w)$.*

Observe that Example 3.9 and Proposition 3.10 precisely recover Proposition 3.6.

The upcoming results discuss the calculus of second subderivatives of indicator functions associated with graphs and epigraphs of single-valued mappings. Furthermore, we carve out the role of the directional proximal subdifferential in this context.

Proposition 3.11. *For a continuous function $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $(\bar{x}, \bar{y}) \in \text{gph } F$, and pairs $(x^*, y^*), (u, v) \in \mathbb{R}^n \times \mathbb{R}^m$, we have*

$$\begin{aligned} d^2 \delta_{\text{gph } F}((\bar{x}, \bar{y}); (x^*, y^*))(u, v) &= d^2 \delta_{\text{gph } F^{-1}}((\bar{y}, \bar{x}); (y^*, x^*))(v, u) \\ &\geq d^2 \langle -y^*, F \rangle(\bar{x}; x^*)(u), \end{aligned}$$

and the last estimate holds as equality for some $v \in DF(\bar{x})(u)$ satisfying the condition $d \langle -y^, F \rangle(\bar{x})(u) = \langle -y^*, v \rangle$ whenever F is calm at $\bar{x}$ in direction u .*

Proof. The first equality is trivial. Observing that $\bar{y} = F(\bar{x})$, consider sequences $t_k \downarrow 0$ and $(u_k, v_k) \rightarrow (u, v)$ with $F(\bar{x}) + t_k v_k = F(\bar{x} + t_k u_k)$ such that

$$d^2 \delta_{\text{gph } F}((\bar{x}, \bar{y}); (x^*, y^*))(u, v) = \lim_{k \rightarrow \infty} \frac{-2 \langle (x^*, y^*), (u_k, v_k) \rangle}{t_k}.$$

Then we have

$$\begin{aligned} d^2 \delta_{\text{gph } F}((\bar{x}, \bar{y}); (x^*, y^*))(u, v) &= \lim_{k \rightarrow \infty} \frac{\langle -y^*, F \rangle(\bar{x} + t_k u_k) - \langle -y^*, F \rangle(\bar{x}) - t_k \langle x^*, u_k \rangle}{\frac{1}{2} t_k^2} \\ &\geq d^2 \langle -y^*, F \rangle(\bar{x}; x^*)(u). \end{aligned}$$

If such sequences $t_k \downarrow 0$ and $(u_k, v_k) \rightarrow (u, v)$ do not exist, $(u, v) \notin T_{\text{gph } F}(\bar{x}, \bar{y})$ is valid and the inequality holds trivially.

To prove the converse inequality, consider $t_k \downarrow 0$ and $u_k \rightarrow u$ which recover $d \langle -y^*, F \rangle(\bar{x})(u)$ and $d^2 \langle -y^*, F \rangle(\bar{x}; x^*)(u)$ simultaneously, see Lemma 2.7, and set $v_k := (F(\bar{x} + t_k u_k) - F(\bar{x}))/t_k$ for each $k \in \mathbb{N}$. By the assumed calmness of F at $\bar{x}$ in direction u , we know that there is $v \in DF(\bar{x})(u)$ satisfying the condition $d \langle -y^*, F \rangle(\bar{x})(u) = \langle -y^*, v \rangle$ such that by passing to a subsequence (without relabeling) we may assume $v_k \rightarrow v$. Thus,

$$d^2 \langle -y^*, F \rangle(\bar{x}; x^*)(u) = \lim_{k \rightarrow \infty} \frac{-2 \langle (x^*, y^*), (u_k, v_k) \rangle}{t_k} \geq d^2 \delta_{\text{gph } F}((\bar{x}, \bar{y}); (x^*, y^*))(u, v)$$

and the proof is complete. $\square$

Given the connections between second subderivatives and directional proximal normal cones as well as subdifferentials, see Proposition 3.6 as well as Definition 3.7 and Proposition 3.10, estimates for second subderivatives automatically contain certain estimates for the directional proximal normals and subdifferentials. We demonstrate this in the following corollary.

Corollary 3.12. *In the setting of Proposition 3.11, let $v \in DF(\bar{x})(u)$. We have*

$$\begin{aligned} x^* \in \tilde{\partial}^p \langle -y^*, F \rangle(\bar{x}; u) &\implies (x^*, y^*) \in \hat{\mathcal{N}}_{\text{gph } F}^p((\bar{x}, \bar{y}); (u, v)) \\ x^* \in \hat{\partial}^p \langle -y^*, F \rangle(\bar{x}; u) &\implies (x^*, y^*) \in \hat{\mathcal{N}}_{\text{gph } F}^p((\bar{x}, \bar{y}); (u, v)) \text{ provided} \\ &\text{that } d \langle -y^*, F \rangle(\bar{x})(u) = \langle -y^*, v \rangle \end{aligned}$$

and the reverse implications also hold for some $v \in DF(\bar{x})(u)$ satisfying the condition $d \langle -y^*, F \rangle(\bar{x})(u) = \langle -y^*, v \rangle$ whenever F is calm at $\bar{x}$ in direction u .

Proof. For $x^* \in \tilde{\partial}^p \langle -y^*, F \rangle(\bar{x}; u)$, Definition 3.7 and Proposition 3.11 imply

$$-\infty < d^2 \langle -y^*, F \rangle(\bar{x}; x^*)(u) \leq d^2 \delta_{\text{gph } F}((\bar{x}, \bar{y}); (x^*, y^*))(u, v)$$

for all $v \in \mathbb{R}^m$. Consequently, Proposition 3.6 yields $(x^*, y^*) \in \hat{\mathcal{N}}_{\text{gph } F}^p((\bar{x}, \bar{y}); (u, v))$ for all $v \in DF(\bar{x})(u)$. If we have additionally $\langle x^*, u \rangle = d \langle -y^*, F \rangle(\bar{x})(u)$, that is, $x^* \in \hat{\partial}^p \langle -y^*, F \rangle(\bar{x}; u)$, then for v satisfying the relation $d \langle -y^*, F \rangle(\bar{x})(u) = \langle -y^*, v \rangle$, we get $\langle (x^*, y^*), (u, v) \rangle = \langle x^*, u \rangle - \langle -y^*, v \rangle = 0$ and $(x^*, y^*) \in \hat{\mathcal{N}}_{\text{gph } F}^p((\bar{x}, \bar{y}); (u, v))$ follows.

Next, let F be calm at $\bar{x}$ in direction u . Consider $(x^*, y^*) \in \hat{\mathcal{N}}_{\text{gph } F}^p((\bar{x}, \bar{y}); (u, v))$ with $v \in DF(\bar{x})(u)$ and $d \langle -y^*, F \rangle(\bar{x})(u) = \langle -y^*, v \rangle$ satisfying

$$d^2 \delta_{\text{gph } F}((\bar{x}, \bar{y}); (x^*, y^*))(u, v) = d^2 \langle -y^*, F \rangle(\bar{x}; x^*)(u).$$

Such v exists by Proposition 3.11. Both of these second subderivatives are greater than $-\infty$ by Proposition 3.6, and Proposition 3.10 implies $x^* \in \tilde{\partial}^p \langle -y^*, F \rangle(\bar{x}; u)$. Again, if $\langle (x^*, y^*), (u, v) \rangle = 0$, i.e., $(x^*, y^*) \in \hat{\mathcal{N}}_{\text{gph } F}^p((\bar{x}, \bar{y}); (u, v))$, we directly get $\langle x^*, u \rangle = \langle -y^*, v \rangle = d \langle -y^*, F \rangle(\bar{x})(u)$ and $x^* \in \hat{\partial}^p \langle -y^*, F \rangle(\bar{x}; u)$ follows. $\square$

The above corollary can be viewed as a form of the scalarization formula for coderivatives of single-valued mappings, see [38, Proposition 9.24].

Proposition 3.13. *Let $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be a lower semicontinuous function and fix $\bar{z} \in \mathbb{R}^n$ with $|h(\bar{z})| < \infty$ as well as $z^* \in \mathbb{R}^n$ and $(w, \omega) \in \mathbb{R}^n \times \mathbb{R}$. We have*

$$d^2 \delta_{\text{epi } h}((\bar{z}, h(\bar{z})); (z^*, -1))(w, \omega) \geq d^2 h(\bar{z}; z^*)(w),$$

and the estimate holds as equality for each ω satisfying

$$\omega \in \begin{cases} \mathbb{R} & \text{if } dh(\bar{z})(w) = \infty, \\ (-\infty, \langle z^*, w \rangle) & \text{if } dh(\bar{z})(w) = -\infty, \\ \{dh(\bar{z})(w)\} & \text{if } dh(\bar{z})(w) \in \mathbb{R}. \end{cases}$$

Proof. Consider some sequences $t_k \downarrow 0$ and $(w_k, \omega_k) \rightarrow (w, \omega)$ with the property $(\bar{z} + t_k w_k, h(\bar{z}) + t_k \omega_k) \in \text{epi } h$ for each $k \in \mathbb{N}$ such that

$$d^2 \delta_{\text{epi } h}((\bar{z}, h(\bar{z})); (z^*, -1))(w, \omega) = \lim_{k \rightarrow \infty} \frac{-2\langle (z^*, -1), (w_k, \omega_k) \rangle}{t_k} = \lim_{k \rightarrow \infty} \frac{\omega_k - \langle z^*, w_k \rangle}{\frac{1}{2}t_k}.$$

Due to $\omega_k \geq (h(\bar{z} + t_k w_k) - h(\bar{z}))/t_k$ for each $k \in \mathbb{N}$, we have

$$d^2 \delta_{\text{epi } h}((\bar{z}, h(\bar{z})); (z^*, -1))(w, \omega) \geq \lim_{k \rightarrow \infty} \frac{h(\bar{z} + t_k w_k) - h(\bar{z}) - t_k \langle z^*, w_k \rangle}{\frac{1}{2}t_k^2} \geq d^2 h(\bar{z}; z^*)(w).$$

If such sequences $t_k \downarrow 0$ and $(w_k, \omega_k) \rightarrow (w, \omega)$ do not exist, $(w, \omega) \notin T_{\text{epi } h}(\bar{z}, h(\bar{z}))$ is valid and the inequality holds trivially.

The converse relation will be shown by a distinction of cases. If $d h(\bar{z})(w) = \infty$, we get $d h(\bar{z})(w) > \langle z^*, w \rangle$ and

$$d^2 \delta_{\text{epi } h}((\bar{z}, h(\bar{z})); (z^*, -1))(w, \omega) \geq d^2 h(\bar{z}; z^*)(w) = \infty$$

for each $\omega \in \mathbb{R}$ from the first part of the proof and (5). If $d h(\bar{z})(w) = -\infty$, we find sequences $t'_k \downarrow 0$ and $w'_k \rightarrow w$ such that $\omega'_k := (h(\bar{z} + t'_k w'_k) - h(\bar{z}))/t'_k \rightarrow -\infty$. Thus, for each $\omega < \langle z^*, w \rangle$, we have $\omega'_k < \omega$ and, thus, $(\bar{z} + t'_k w'_k, h(\bar{z}) + t'_k \omega) \in \text{epi } h$ for large enough $k \in \mathbb{N}$. Together with (5), this gives

$$d^2 \delta_{\text{epi } h}((\bar{z}, h(\bar{z})); (z^*, -1))(w, \omega) \leq \liminf_{k \rightarrow \infty} \frac{\omega - \langle z^*, w'_k \rangle}{\frac{1}{2}t'_k} = -\infty = d^2 h(\bar{z}; z^*)(w).$$

Finally, if $d h(\bar{z})(w) \in \mathbb{R}$ let us pick some sequences $t_k \downarrow 0$ and $w_k \rightarrow w$ recovering $d h(\bar{z})(w)$ and $d^2 h(\bar{z}; z^*)(w)$ simultaneously, according to Lemma 2.7. Now, setting $\omega_k := (h(\bar{z} + t_k w_k) - h(\bar{z}))/t_k$ for each $k \in \mathbb{N}$, we obtain

$$\begin{aligned} d^2 h(\bar{z}; z^*)(w) &= \lim_{k \rightarrow \infty} \frac{-2\langle (z^*, -1), (w_k, \omega_k) \rangle}{t_k} \\ &\geq d^2 \delta_{\text{epi } h}((\bar{z}, h(\bar{z})); (z^*, -1))(w, d h(\bar{z})(w)), \end{aligned}$$

and this completes the proof. $\square$

Taking into account Proposition 3.6, the above result actually clarifies the close relationship between the directional proximal subdifferential of a function and its geometric counterpart.

Corollary 3.14. *Given a lower semicontinuous function $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, a point $\bar{z} \in \mathbb{R}^n$ such that $|h(\bar{z})| < \infty$, and a direction $w \in \mathbb{R}^n$ such that $|d h(\bar{z})(w)| < \infty$, we have*

$$\tilde{\partial}^p h(\bar{z}; w) = \tilde{\partial}_g^p h(\bar{z}; (w, d h(\bar{z})(w))), \quad \widehat{\partial}^p h(\bar{z}; w) = \widehat{\partial}_g^p h(\bar{z}; (w, d h(\bar{z})(w))).$$

If h is convex, we have

$$\widehat{\partial}^p h(\bar{z}; w) = \partial h(\bar{z}) \cap \{z^* \in \mathbb{R}^n \mid \langle z^*, w \rangle = d h(\bar{z})(w)\}.$$

Proof. Since

$$\widehat{\partial}_g^p h(\bar{z}; (w, d h(\bar{z})(w))) = \widetilde{\partial}_g^p h(\bar{z}; (w, d h(\bar{z})(w))) \cap \{z^* \in \mathbb{R}^n \mid d h(\bar{z})(w) = \langle z^*, w \rangle\},$$

the second statement follows immediately from the first one.

Let $z^* \in \widetilde{\partial}_g^p h(\bar{z}; w)$. Since $d h(\bar{z})(w) \in \mathbb{R}$, Definition 3.7 and Proposition 3.13 yield the relations

$$-\infty < d^2 h(\bar{z}; z^*)(w) = d^2 \delta_{\text{epi } h}((\bar{z}, h(\bar{z})); (z^*, -1))(w, d h(\bar{z})(w)),$$

and since we have $(w, d h(\bar{z})(w)) \in \text{epi } d h(\bar{z}) = T_{\text{epi } h}(\bar{z}, h(\bar{z}))$, Proposition 3.6 implies $z^* \in \widetilde{\partial}_g^p h(\bar{z}; (w, d h(\bar{z})(w)))$.

The opposite inclusion follows from the same arguments. The convex case follows from the definition of $\widehat{\partial}_g^p h(\bar{z}; (w, d h(\bar{z})(w)))$ and (14). $\square$

4. Calculus for second subderivatives

This section is devoted to the calculus of second subderivatives. First, we propose two very general calculus rules for second subderivatives, namely a chain rule, i.e., the rule for compositions, and a rule for marginal functions. Afterwards, we apply these rules to derive some other calculus principles for the second subderivative of set indicators where the set is given as the smooth image or pre-image of a closed set.

4.1. Compositions and marginal functions

Let us start with the consideration of a very general chain rule.

Theorem 4.1. *Consider a lower semicontinuous function $g: \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ and a continuous mapping $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$, and, for $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ given by $h := g \circ F$, let $\bar{x} \in \mathbb{R}^n$ be chosen such that $|h(\bar{x})| < \infty$. If F is calm at $\bar{x}$ in direction $u \in \mathbb{R}^n$, then there exist $v \in DF(\bar{x})(u)$ such that for each $x^* \in \mathbb{R}^n$, one has*

$$\begin{aligned} d h(\bar{x})(u) &\geq d \langle y^*, F \rangle(\bar{x})(u) + d g(F(\bar{x}))(v) - \langle y^*, v \rangle, \\ d^2 h(\bar{x}; x^*)(u) &\geq d^2 \langle y^*, F \rangle(\bar{x}; x^*)(u) + d^2 g(F(\bar{x}); y^*)(v) \end{aligned}$$

for all $y^* \in \mathbb{R}^m$ such that the right-hand side does not contain the summands $-\infty$ and ∞ simultaneously.

Proof. By Lemma 2.7, there are sequences $t_k \downarrow 0$ and $u_k \rightarrow u$ which recover $d h(\bar{x})(u)$ and $d^2 h(\bar{x}; x^*)(u)$ simultaneously. Let us set $v_k := (F(\bar{x} + t_k u_k) - F(\bar{x}))/t_k$ for each $k \in \mathbb{N}$. Due to the postulated calmness assumption, we may assume that $v_k \rightarrow v \in DF(\bar{x})(u)$. Thus, we find

$$\begin{aligned} d h(\bar{x})(u) &= \lim_{k \rightarrow \infty} \frac{g(F(\bar{x} + t_k u_k)) - g(F(\bar{x}))}{t_k} \\ &= \lim_{k \rightarrow \infty} \left(\frac{g(F(\bar{x}) + t_k v_k) - g(F(\bar{x}))}{t_k} + \frac{\langle y^*, F \rangle(\bar{x} + t_k u_k) - \langle y^*, F \rangle(\bar{x})}{t_k} - \langle y^*, v_k \rangle \right) \\ &\geq d g(F(\bar{x}))(v) + d \langle y^*, F \rangle(\bar{x})(u) - \langle y^*, v \rangle \end{aligned}$$

and

$$\begin{aligned} d^2 h(\bar{x}; x^*)(u) &= \lim_{k \rightarrow \infty} \frac{g(F(\bar{x} + t_k u_k)) - g(F(\bar{x})) - t_k \langle x^*, u_k \rangle}{\frac{1}{2} t_k^2} \\ &= \lim_{k \rightarrow \infty} \left(\frac{g(F(\bar{x}) + t_k v_k) - g(F(\bar{x})) - t_k \langle y^*, v_k \rangle}{\frac{1}{2} t_k^2} + \frac{\langle y^*, F \rangle(\bar{x} + t_k u_k) - \langle y^*, F \rangle(\bar{x}) - t_k \langle x^*, u_k \rangle}{\frac{1}{2} t_k^2} \right) \\ &\geq d^2 g(F(\bar{x}); y^*)(v) + d^2 \langle y^*, F \rangle(\bar{x}; x^*)(u) \end{aligned}$$

where we used $\liminf_{k \rightarrow \infty} (a_k + b_k) \geq \liminf_{k \rightarrow \infty} a_k + \liminf_{k \rightarrow \infty} b_k$, which holds except for the indeterminate situation $\infty - \infty$, to obtain the last inequality. $\square$

Corollary 4.2. *In the setting of Theorem 4.1, assume that $d h(\bar{x})(u) < \infty$. Then there exists $v \in DF(\bar{x})(u)$ such that for each $x^* \in \mathbb{R}^n$ and $y^* \in \mathbb{R}^m$, we have the following implications:*

$$\left. \begin{aligned} y^* \in \tilde{\partial}^p g(F(\bar{x}); v), \quad x^* \in \tilde{\partial}^p \langle y^*, F \rangle(\bar{x}; u) &\implies x^* \in \tilde{\partial}^p h(\bar{x}; u), \\ d h(\bar{x})(u) = d \langle y^*, F \rangle(\bar{x})(u), \\ y^* \in \tilde{\partial}^p g(F(\bar{x}); v), \quad x^* \in \hat{\partial}^p \langle y^*, F \rangle(\bar{x}; u) &\implies x^* \in \hat{\partial}^p h(\bar{x}; u). \end{aligned} \right\}$$

Proof. For $y^* \in \tilde{\partial}^p g(F(\bar{x}); v)$ and $x^* \in \tilde{\partial}^p \langle y^*, F \rangle(\bar{x}; u)$, we get the following estimates by Definition 3.7:

$$\begin{aligned} d^2 \langle y^*, F \rangle(\bar{x}; x^*)(u) &> -\infty, \quad d^2 g(F(\bar{x}); y^*)(v) > -\infty, \\ |d \langle y^*, F \rangle(\bar{x})(u)| &< \infty, \quad |d g(F(\bar{x}))(v)| < \infty. \end{aligned}$$

Consequently, the estimates from Theorem 4.1 apply and yield that $d h(\bar{x})(u) > -\infty$ and $d^2 h(\bar{x}; x^*)(u) > -\infty$. This in turn gives us $x^* \in \tilde{\partial}^p h(\bar{x}; u)$ since we assumed $d h(\bar{x})(u) < \infty$.

In case when $d h(\bar{x})(u) = d \langle y^*, F \rangle(\bar{x})(u)$, $y^* \in \tilde{\partial}^p g(F(\bar{x}); v)$, and $x^* \in \hat{\partial}^p \langle y^*, F \rangle(\bar{x}; u)$, we can deduce $x^* \in \tilde{\partial}^p h(\bar{x}; u)$ as above. Due to $d \langle y^*, F \rangle(\bar{x})(u) = \langle x^*, u \rangle$, we have $d h(\bar{x})(u) = \langle x^*, u \rangle$, and $x^* \in \hat{\partial}^p h(\bar{x}; u)$ follows. $\square$

Under the assumptions of Corollary 4.2, the second implication also gives

$$\left. \begin{aligned} d h(\bar{x})(u) &= d \langle y^*, F \rangle(\bar{x})(u) + d g(F(\bar{x}))(v) - \langle y^*, v \rangle, \\ y^* \in \hat{\partial}^p g(F(\bar{x}); v), \quad x^* \in \hat{\partial}^p \langle y^*, F \rangle(\bar{x}; u) &\implies x^* \in \hat{\partial}^p h(\bar{x}; u), \end{aligned} \right\}$$

which demonstrates the close relationship to the subderivative chain rule from Theorem 4.1.

If F is twice continuously differentiable, we get the following corollary from Theorem 4.1 by taking into account

$$DF(\bar{x})(u) = \{\nabla F(\bar{x})u\}, \quad \nabla \langle y^*, F \rangle(\bar{x}) = \nabla F(\bar{x})^\top y^*,$$

$$\nabla^2 \langle y^*, F \rangle(\bar{x})(u, u) = \langle y^*, \nabla^2 F(\bar{x})(u, u) \rangle,$$

local Lipschitzness of F around $\bar{x}$, and (7).

Corollary 4.3. *Consider a lower semicontinuous function $g: \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ and a twice continuously differentiable mapping $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$, and, for $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ given by $h := g \circ F$, let $\bar{x} \in \mathbb{R}^n$ be chosen such that $|h(\bar{x})| < \infty$. Then for each $x^* \in \mathbb{R}^m$ and $u \in \mathbb{R}^n$, one has*

$$d h(\bar{x})(u) \geq d g(F(\bar{x}))(\nabla F(\bar{x})u),$$

$$d^2 h(\bar{x}; x^*)(u) \geq \sup_{\nabla F(\bar{x})^\top y^* = x^*} (\langle y^*, \nabla^2 F(\bar{x})(u, u) \rangle + d^2 g(F(\bar{x}); y^*)(\nabla F(\bar{x})u)).$$

If $\nabla F(\bar{x})$ possesses full row rank, for each $y^* \in \mathbb{R}^m$ and $u \in \mathbb{R}^n$, one has

$$d h(\bar{x})(u) = d g(F(\bar{x}))(\nabla F(\bar{x})u),$$

$$d^2 h(\bar{x}; \nabla F(\bar{x})^\top y^*)(u) = \langle y^*, \nabla^2 F(\bar{x})(u, u) \rangle + d^2 g(F(\bar{x}); y^*)(\nabla F(\bar{x})u).$$

Proof. The lower estimates are a direct consequence of Theorem 4.1. In order to show the second statement, choose sequences $t_k \downarrow 0$ and $v_k \rightarrow \nabla F(\bar{x})u$ which recover $d g(F(\bar{x}))(\nabla F(\bar{x})u)$ and $d^2 g(F(\bar{x}); y^*)(\nabla F(\bar{x})u)$ simultaneously, see Lemma 2.7. Furthermore, set $x^* := \nabla F(\bar{x})^\top y^*$. Since $\nabla F(\bar{x})$ possesses full row rank, [38, Exercise 9.44] implies the existence of a constant $\kappa > 0$ and neighborhoods $U \subset \mathbb{R}^n$ of $\bar{x}$ and $V \subset \mathbb{R}^m$ of $F(\bar{x})$ such that

$$\text{dist}(x, F^{-1}(y)) \leq \kappa \text{dist}(y, F(x)) \quad \forall x \in U, \forall y \in V,$$

i.e., F is so-called metrically regular at $(\bar{x}, F(\bar{x}))$. For sufficiently large $k \in \mathbb{N}$, we may apply this estimate with $x := \bar{x} + t_k u$ and $y := F(\bar{x}) + t_k v_k$ to find $x_k \in \mathbb{R}^n$ such that $F(x_k) = F(\bar{x}) + t_k v_k$ and $\|x_k - \bar{x} - t_k u\| \leq \kappa \|F(\bar{x} + t_k u) - F(\bar{x}) - t_k v_k\|$. Let us set $u_k := (x_k - \bar{x})/t_k$ for each $k \in \mathbb{N}$ sufficiently large. Then we find

$$\|u_k - u\| \leq \kappa \left\| \frac{F(\bar{x} + t_k u) - F(\bar{x})}{t_k} - v_k \right\| \rightarrow 0$$

from $v_k \rightarrow \nabla F(\bar{x})u$, i.e., $u_k \rightarrow u$. Thus, we can exploit $x_k = \bar{x} + t_k u_k$ for all sufficiently large $k \in \mathbb{N}$ in order to find

$$d g(F(\bar{x}))(\nabla F(\bar{x})u) = \lim_{k \rightarrow \infty} \frac{g(F(\bar{x} + t_k u_k)) - g(F(\bar{x}))}{t_k} \geq d h(\bar{x})(u)$$

and

$$d^2 g(F(\bar{x}); y^*)(\nabla F(\bar{x})u) = \lim_{k \rightarrow \infty} \frac{g(F(\bar{x} + t_k u_k)) - g(F(\bar{x})) - t_k \langle y^*, v_k \rangle}{\frac{1}{2} t_k^2}$$

$$= \lim_{k \rightarrow \infty} \left(\frac{g(F(\bar{x} + t_k u_k)) - g(F(\bar{x})) - t_k \langle x^*, u_k \rangle}{\frac{1}{2} t_k^2} + \frac{\langle x^*, t_k u_k \rangle - \langle y^*, t_k v_k \rangle}{\frac{1}{2} t_k^2} \right)$$

$$\geq d^2 h(\bar{x}; x^*)(u) + \lim_{k \rightarrow \infty} \frac{\langle y^*, \nabla F(\bar{x})(x_k - \bar{x}) + F(\bar{x}) - F(x_k) \rangle}{\frac{1}{2} t_k^2}$$

$$\begin{aligned}
&= d^2 h(\bar{x}; x^*)(u) + \lim_{k \rightarrow \infty} \langle y^*, -\nabla^2 F(\bar{x})(u_k, u_k) \rangle \\
&= d^2 h(\bar{x}; x^*)(u) - \langle y^*, \nabla^2 F(\bar{x})(u, u) \rangle,
\end{aligned}$$

where we used a second-order Taylor expansion of F at $\bar{x}$ in the last but one equality. Noting that the converse relations hold due to the general estimates, the proof is complete. $\square$

Let us note that the result in Corollary 4.3 is essentially different from the chain rule which can be found in [38, Theorem 13.14]. Therein, the authors exploit a less restrictive qualification condition than the full row rank of $\nabla F(\bar{x})$ in order to derive a general lower estimate of the second subderivative, and in order to get equality, they additionally assume that g is a so-called fully amenable function locally around $F(\bar{x})$, i.e., the composition of a twice continuously differentiable inner and a piecewise linear-quadratic outer function. In contrast, Corollary 4.3 yields a general lower estimate even in the absence of a qualification condition. Since an upper estimate is not so important for our purposes, we show equality only under the comparatively strong full rank condition without focusing on minimal assumptions.

Corollary 4.4. *For two lower semicontinuous functions $h_1, h_2: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, and $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ defined by $h(x) := h_1(x) + h_2(x)$ for all $x \in \mathbb{R}^n$, some point $\bar{x} \in \mathbb{R}^n$ with $|h(\bar{x})| < \infty$, and $x^*, u \in \mathbb{R}^n$, we have*

$$d^2 h(\bar{x}; x^*)(u) \geq d^2 h_1(\bar{x}; x_1^*)(u) + d^2 h_2(\bar{x}; x_2^*)(u)$$

for all $x_1^*, x_2^* \in \mathbb{R}^n$ such that $x_1^* + x_2^* = x^*$ provided the right-hand side is not $\infty - \infty$.

Proof. For the proof, we first apply the chain rule from Corollary 4.3 to the function $F(x) := (x, x)$, $x \in \mathbb{R}^n$, and $g(x_1, x_2) := h_1(x_1) + h_2(x_2)$, $x_1, x_2 \in \mathbb{R}^n$, and then Proposition 2.8 (b). $\square$

Next, we will study a marginal function rule for the second subderivative. Therefore, recall from Section 2.2 that we choose a proper, lower semicontinuous function $\varphi: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ and consider the associated marginal function $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ given as in (1) together with the set-valued mappings $\Upsilon: \mathbb{R}^n \times \mathbb{R} \rightrightarrows \mathbb{R}^m$ and $\Psi: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ from (2). Recall that $\Psi(x) := \emptyset$ is used for each $x \in \mathbb{R}^n$ such that $\varphi(x, y) = \infty$ holds for all $y \in \mathbb{R}^m$. Furthermore, we would like to mention again that inner semicompactness of Υ or Ψ yields lower semicontinuity of h so that the consideration of subderivatives is reasonable, see Lemmas 2.3 and 2.4.

In the next theorem, we address the second subderivative of marginal functions with the aid of the solution mapping Ψ .

Theorem 4.5. *Consider a proper, lower semicontinuous function $\varphi: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ and fix $\bar{x} \in \text{dom } \Psi$ for the mapping $\Psi: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given in (2), and let Ψ be inner semicompact at $\bar{x}$. Then for each $x^* \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$, one has*

$$d h(\bar{x})(u) \leq \inf_{y \in \Psi(\bar{x}), v \in \mathbb{R}^m} d \varphi(\bar{x}, y)(u, v), \quad (16a)$$

$$d^2 h(\bar{x}; x^*)(u) \leq \inf_{y \in \Psi(\bar{x}), v \in \mathbb{R}^m} d^2 \varphi((\bar{x}, y); (x^*, 0))(u, v) \quad (16b)$$

where $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is the marginal function defined in (1).

On the other hand, suppose that Ψ is inner calm* at $\bar{x}$ in direction u . Then the estimates (16) hold as equalities, and whenever $dh(\bar{x})(u)$ is finite, both infima therein are attained at some pair $(y, v) \in \mathbb{R}^m \times \mathbb{R}^m$ with $y \in \Psi(\bar{x})$ and $v \in D\Psi(\bar{x}, y)(u)$.

Proof. For arbitrary $y \in \Psi(\bar{x})$ and $v \in \mathbb{R}^m$, by Lemma 2.7, there are sequences $t_k \downarrow 0$ and $(u_k, v_k) \rightarrow (u, v)$ satisfying

$$\begin{aligned} d\varphi(\bar{x}, y)(u, v) &= \lim_{k \rightarrow \infty} \frac{\varphi(\bar{x}, y) + t_k(u_k, v_k) - \varphi(\bar{x}, y)}{t_k} \\ &\geq \lim_{k \rightarrow \infty} \frac{h(\bar{x} + t_k u_k) - h(\bar{x})}{t_k} \geq dh(\bar{x})(u), \\ d^2\varphi((\bar{x}, y); (x^*, 0))(u, v) &= \lim_{k \rightarrow \infty} \frac{\varphi(\bar{x}, y) + t_k(u_k, v_k) - \varphi(\bar{x}, y) - t_k\langle x^*, 0 \rangle, (u_k, v_k)\rangle}{\frac{1}{2}t_k^2} \\ &\geq \lim_{k \rightarrow \infty} \frac{h(\bar{x} + t_k u_k) - h(\bar{x}) - t_k\langle x^*, u_k \rangle}{\frac{1}{2}t_k^2} \geq d^2h(\bar{x}; x^*)(u) \end{aligned}$$

since $h(\bar{x}) = \varphi(\bar{x}, y)$. This shows the first claim.

To prove the second claim, consider sequences $t_k \downarrow 0$ and $u_k \rightarrow u$ which recover $dh(\bar{x})(u)$ and $d^2h(\bar{x}; x^*)(u)$ simultaneously, see Lemma 2.7. By passing to a subsequence (without relabeling), the assumed inner calmness* of Ψ in direction u yields the existence of $\kappa > 0$, a point $y \in \mathbb{R}^m$, and a sequence $y_k \rightarrow y$ such that

$$y_k \in \Psi(\bar{x} + t_k u_k), \quad (17a)$$

$$\|y_k - y\| \leq t_k \kappa \|u_k\| \quad (17b)$$

hold for each $k \in \mathbb{N}$. Due to (17b), $v_k := (y_k - y)/t_k$ remains bounded as $k \rightarrow \infty$, and we may assume $v_k \rightarrow v$ for some $v \in \mathbb{R}^m$ with $\|v\| \leq \kappa \|u\|$.

Again, by passing to some subsequence (without relabeling), we may assume that $h(\bar{x} + t_k u_k) \rightarrow \alpha \geq h(\bar{x})$, taking into account lower semicontinuity of h , see Lemmas 2.3 and 2.4. Moreover, if $\alpha > h(\bar{x})$, we get $dh(\bar{x})(u) = d^2h(\bar{x}; x^*)(u) = \infty$ and the converse inequalities in (16) are trivial.

Thus, we assume that $\varphi(\bar{x} + t_k u_k, y_k) = h(\bar{x} + t_k u_k) \rightarrow h(\bar{x})$, and lower semicontinuity of φ gives $\varphi(\bar{x}, y) \leq h(\bar{x})$, i.e., $y \in \Psi(\bar{x})$. Moreover, (17a) yields $v \in D\Psi(\bar{x}, y)(u)$ since $y_k = y + t_k v_k$ holds for all $k \in \mathbb{N}$. Finally, we obtain

$$\begin{aligned} dh(\bar{x})(u) &= \lim_{k \rightarrow \infty} \frac{h(\bar{x} + t_k u_k) - h(\bar{x})}{t_k} = \lim_{k \rightarrow \infty} \frac{\varphi(\bar{x} + t_k u_k, y + t_k v_k) - \varphi(\bar{x}, y)}{t_k} \\ &\geq d\varphi(\bar{x}, y)(u, v), \\ d^2h(\bar{x}; x^*)(u) &= \lim_{k \rightarrow \infty} \frac{h(\bar{x} + t_k u_k) - h(\bar{x}) - t_k\langle x^*, u_k \rangle}{\frac{1}{2}t_k^2} \\ &= \lim_{k \rightarrow \infty} \frac{\varphi(\bar{x} + t_k u_k, y + t_k v_k) - \varphi(\bar{x}, y) - t_k\langle x^*, 0 \rangle, (u_k, v_k)\rangle}{\frac{1}{2}t_k^2} \\ &\geq d^2\varphi((\bar{x}, y); (x^*, 0))(u, v), \end{aligned}$$

which completes the proof. $\square$

Corollary 4.6. *In the setting of Theorem 4.5, let $|dh(\bar{x})(u)| < \infty$. For all $y \in \Psi(\bar{x})$ and $v \in \mathbb{R}^m$, one has*

$$\begin{aligned} x^* \in \tilde{\partial}^p h(\bar{x}; u) &\implies (x^*, 0) \in \tilde{\partial}^p \varphi((\bar{x}, y); (u, v)) \text{ provided } d\varphi(\bar{x}, y)(u, v) < \infty \\ x^* \in \hat{\partial}^p h(\bar{x}; u) &\implies (x^*, 0) \in \hat{\partial}^p \varphi((\bar{x}, y); (u, v)) \text{ provided } d\varphi(\bar{x}, y)(u, v) = dh(\bar{x})(u). \end{aligned}$$

Moreover, the reverse implications hold for some $y \in \Psi(\bar{x})$ and $v \in D\Psi(\bar{x}, y)(u)$ with $d\varphi(\bar{x}, y)(u, v) = dh(\bar{x})(u)$ provided Ψ is inner calm* at $\bar{x}$ in direction u .

Proof. For $x^* \in \tilde{\partial}^p h(\bar{x}; u)$, Definition 3.7 and Theorem 4.5 imply

$$-\infty < d^2 h(\bar{x}; x^*)(u) \leq d^2 \varphi((\bar{x}, y); (x^*, 0))(u, v)$$

for all $y \in \Psi(\bar{x})$ and $v \in D\Psi(\bar{x}, y)(u)$, and this also gives $d\varphi(\bar{x}, y)(u, v) > -\infty$ due to (5). Consequently, $(x^*, 0) \in \tilde{\partial}^p \varphi((\bar{x}, y); (u, v))$ if $d\varphi(\bar{x}, y)(u, v) < \infty$. If, additionally, $\langle x^*, u \rangle = dh(\bar{x})(u)$, i.e., $x^* \in \hat{\partial}^p h(\bar{x}; u)$, then for $y \in \Psi(\bar{x})$ and $v \in \mathbb{R}^m$ satisfying $d\varphi(\bar{x}, y)(u, v) = dh(\bar{x})(u)$, we get

$$\langle (x^*, 0), (u, v) \rangle = \langle x^*, u \rangle = dh(\bar{x})(u) = d\varphi(\bar{x}, y)(u, v) \quad (18)$$

and $(x^*, 0) \in \hat{\partial}^p \varphi((\bar{x}, y); (u, v))$ follows.

Let us now argue for the reverse implications. Consider $(x^*, 0) \in \tilde{\partial}^p \varphi((\bar{x}, y); (u, v))$ with $y \in \Psi(\bar{x})$ and $v \in D\Psi(\bar{x}, y)(u)$ satisfying

$$d\varphi(\bar{x}, y)(u, v) = dh(\bar{x})(u), \quad d^2 \varphi((\bar{x}, y); (x^*, 0))(u, v) = d^2 h(\bar{x}; x^*)(u).$$

Such y, v exist by Theorem 4.5. Both second subderivatives are thus greater than $-\infty$ by Definition 3.7, and so we get $x^* \in \tilde{\partial}^p h(\bar{x}; u)$. If $d\varphi(\bar{x}, y)(u, v) = \langle (x^*, 0), (u, v) \rangle$, we obtain again the equalities (18) and $x^* \in \hat{\partial}^p h(\bar{x}; u)$ follows. $\square$

To better demonstrate Theorem 4.5, we use it to compute the second subderivative of the distance function to a closed set $\Omega \subset \mathbb{R}^n$, given via the representation (3).

Corollary 4.7. *For a closed set $\Omega \subset \mathbb{R}^n$, $\bar{x} \in \Omega$, $x^* \in \mathbb{R}^n$ with $\|x^*\| \leq 1$, and $u \in T_\Omega(\bar{x})$, we have*

$$d^2 \text{dist}(\cdot, \Omega)(\bar{x}; x^*)(u) \geq d^2 \delta_\Omega(\bar{x}; x^*)(u).$$

Proof. First, we claim that

$$d^2 \text{dist}(\cdot, \Omega)(\bar{x}; x^*)(u) = d^2(\varphi_1 + \varphi_2)((\bar{x}, \bar{x}); (x^*, 0))(u, u)$$

for $\varphi_1(x, y) := \|y - x\|$ and $\varphi_2(x, y) := \delta_\Omega(y)$, $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$. By Proposition 2.2, the mapping $\Psi := P_\Omega$ is inner calm* at $\bar{x}$, and so we can apply Theorem 4.5. Since, $P_\Omega(\bar{x}) = \bar{x}$, there is no need for the infimum over $\bar{y} \in P_\Omega(\bar{x})$. Let us now argue why we can omit also the infimum over directions $v \in \mathbb{R}^n$ and take just $v = u$. Note that the proof of Theorem 4.5 yields that we can take v with $\|v\| \leq \kappa\|u\|$ for some $\kappa \geq 0$, which covers the case $u = 0$. If $u \neq 0$, the proof of Theorem 4.5 uses Lemma 2.7 in order to only consider the sequences t_k and u_k which recover also $d \text{dist}(\cdot, \Omega)(\bar{x})(u)$.

This means that $\text{dist}(\bar{x} + t_k u_k, \Omega)/(t_k \|u_k\|) \rightarrow d \text{dist}(\cdot, \Omega)(\bar{x})(u)/\|u\| = 0$ due to $u \in T_\Omega(\bar{x})$ and [38, Example 8.53]. Thus, the second statement of Proposition 2.2 yields that the sequence v_k from the proof of Theorem 4.5 indeed converges to u along a subsequence.

Next, in order to estimate $d^2(\varphi_1 + \varphi_2)((\bar{x}, \bar{x}); (x^*, 0))(u, u)$, we apply the sum rule from Corollary 4.4 first and then the chain rule from Corollary 4.3 twice, taking into account that $\varphi_1 = \|\cdot\| \circ h_1$ and $\varphi_2 = \delta_\Omega \circ h_2$ for the simple smooth mappings $h_1(x, y) := x - y$ and $h_2(x, y) := y$, $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, with vanishing second derivatives. In the sum rule, we choose to split $(x^*, 0)$ into $(x^*, -x^*) + (0, x^*)$ due to the structure of h_1 and h_2 . Consequently, taking into account Lemma 2.9, we conclude

$$\begin{aligned} d^2(\varphi_1 + \varphi_2)((\bar{x}, \bar{x}); (x^*, 0))(u, u) \\ \geq d^2\varphi_1((\bar{x}, \bar{x}); (x^*, -x^*))(u, u) + d^2\varphi_2((\bar{x}, \bar{x}); (0, x^*))(u, u) \\ = d^2\|\cdot\|(0; x^*)(0) + d^2\delta_\Omega(\bar{x}; x^*)(u) = d^2\delta_\Omega(\bar{x}; x^*)(u) \end{aligned}$$

to complete the proof. $\square$

Let us now turn our focus on the general marginal function rule again. With the additional assumption $d h(\bar{x})(u) \geq \langle x^*, u \rangle$, we can derive estimates, similar to those ones in Theorem 4.5, under the milder inner semicompactness/inner calmness* requirements imposed on Υ , see Lemma 2.4.

Theorem 4.8. *Consider a proper, lower semicontinuous function $\varphi: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ and fix $\bar{x} \in \text{dom } \Psi$ for the mapping $\Psi: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given in (2). Suppose that $\Upsilon: \mathbb{R}^n \times \mathbb{R} \rightrightarrows \mathbb{R}^m$ given in (2) is inner semicompact at $(\bar{x}, h(\bar{x}))$ w.r.t. $\text{dom } \Upsilon$. Then for each $x^* \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$, the estimates (16) hold.*

On the other hand, suppose that $d h(\bar{x})(u) \geq \langle x^, u \rangle$. Then the following statements hold.*

- (a) *If $d h(\bar{x})(u) > \langle x^*, u \rangle$, then equality holds in (16b).*
- (b) *If $d h(\bar{x})(u) = \langle x^*, u \rangle$ and if Υ is inner calm* at $(\bar{x}, h(\bar{x}))$ in the direction $(u, \langle x^*, u \rangle)$ w.r.t. $\text{dom } \Upsilon$, then the estimates (16) hold as equalities, and whenever $d^2 h(\bar{x}; x^*)(u)$ is finite, both infima therein are attained at some pair $(y, v) \in \mathbb{R}^m \times \mathbb{R}^m$ with $y \in \Psi(\bar{x})$ and $v \in D\Upsilon((\bar{x}, h(\bar{x})), y)(u, \langle x^*, u \rangle)$.*

Proof. The first claim follows by the same arguments as in the previous theorem.

To prove the second claim, consider sequences $t_k \downarrow 0$ and $u_k \rightarrow u$ recovering the subderivatives $d h(\bar{x})(u)$ and $d^2 h(\bar{x}; x^*)(u)$ simultaneously, see Lemma 2.7. By passing to a subsequence (without relabeling) and taking into account lower semicontinuity of h again, see Lemma 2.3, we may assume that $h(\bar{x} + t_k u_k) \rightarrow \alpha \geq h(\bar{x})$, and that

$$\mu_k := (h(\bar{x} + t_k u_k) - h(\bar{x}))/t_k \rightarrow d h(\bar{x})(u) \geq \langle x^*, u \rangle$$

due to the postulated assumptions. If $\alpha > h(\bar{x})$, we get $d h(\bar{x})(u) = d^2 h(\bar{x}; x^*)(u) = \infty$ and the converse inequalities in (16) follow trivially. If $d h(\bar{x})(u) > \langle x^*, u \rangle$, we find $d^2 h(\bar{x}; x^*)(u) = \infty$, and the converse estimate in (16b) is trivial. Thus, we assume that $\alpha = h(\bar{x})$ and $d h(\bar{x})(u) = \langle x^*, u \rangle$.

The postulated inner semicompactness of Υ at $(\bar{x}, h(\bar{x}))$ yields that, locally around that point, $\text{epi } h = \text{dom } \Upsilon$ and so $(\bar{x} + t_k u_k, h(\bar{x} + t_k u_k)) \in \text{dom } \Upsilon$ for sufficiently large $k \in \mathbb{N}$, see Lemma 2.3. The assumed inner calmness* of Υ at $(\bar{x}, h(\bar{x}))$ in direction $(u, \langle x^*, u \rangle)$ then yields the existence of $\kappa > 0$, a point $y \in \mathbb{R}^m$, and a sequence $y_k \rightarrow y$ such that

$$\begin{aligned} y_k &\in \Upsilon(\bar{x} + t_k u_k, h(\bar{x}) + t_k \mu_k) = \Psi(\bar{x} + t_k u_k), \\ \|y_k - y\| &\leq t_k \kappa \|(u_k, \mu_k)\| \end{aligned}$$

hold for each $k \in \mathbb{N}$. The remainder of the proof follows the same arguments as used to show Theorem 4.5. $\square$

Similar as in Corollary 4.6, one can also derive hidden calculus results on the directional proximal (pre-) subdifferential of h in the setting of Theorem 4.8. For brevity, however, we leave this straightforward task to the interested reader.

Let us recall that the inner calmness* of Ψ at $\bar{x}$ (or Υ at $(\bar{x}, h(\bar{x}))$) in Theorem 4.5 (or Theorem 4.8) is inherent whenever Ψ (or Υ) is isolatedly calm at all points $(\bar{x}, y) \in \text{gph } \Psi$ (or $((\bar{x}, h(\bar{x})), y) \in \text{gph } \Upsilon$) since we already claimed that inner semicompactness holds at $\bar{x}$ (or $(\bar{x}, h(\bar{x}))$), and isolated calmness can be checked in terms of the associated graphical derivative, i.e., in terms of problem data, see e.g. [20].

4.2. Other calculus rules

In order to apply our results to interesting optimization problems, let us derive some additional calculus rules for second subderivatives. More precisely, we look into the term $d^2 \delta_S(x; x^*)(u)$ assuming that S has a pre-image or image structure.

Applying Corollary 4.3 with $h := \delta_C$ for a closed set $C \subset \mathbb{R}^m$ immediately yields the following pre-image rule.

Proposition 4.9. *Consider a closed set $C \subset \mathbb{R}^m$ and a twice continuously differentiable mapping $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$, and let $\bar{x} \in S := F^{-1}(C)$. Then for each $x^* \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$, one has*

$$d^2 \delta_S(\bar{x}; x^*)(u) \geq \sup_{\nabla F(\bar{x})^\top y^* = x^*} (\langle y^*, \nabla^2 F(\bar{x})(u, u) \rangle + d^2 \delta_C(F(\bar{x}); y^*)(\nabla F(\bar{x})u)).$$

If $\nabla F(\bar{x})$ possesses full row rank, for each $y^* \in \mathbb{R}^m$ and $u \in \mathbb{R}^n$, one has

$$d^2 \delta_S(\bar{x}; \nabla F(\bar{x})^\top y^*)(u) = \langle y^*, \nabla^2 F(\bar{x})(u, u) \rangle + d^2 \delta_C(F(\bar{x}); y^*)(\nabla F(\bar{x})u).$$

On the basis of Theorem 4.8, we now aim to derive an image rule. In order to do that, we assume that $S := G(Q) := \{x \in \mathbb{R}^n \mid \exists y \in Q: G(y) = x\}$ holds for a twice continuously differentiable mapping $G: \mathbb{R}^m \rightarrow \mathbb{R}^n$ and a closed set $Q \subset \mathbb{R}^m$. Associated with the data is the set-valued mapping $\Phi: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given by

$$\Phi(x) := Q \cap G^{-1}(x) \quad \forall x \in \mathbb{R}^n. \quad (19)$$

Note that $\text{dom } \Phi = S$.

Proposition 4.10. *Consider a closed set $Q \subset \mathbb{R}^m$ and a twice continuously differentiable mapping $G: \mathbb{R}^m \rightarrow \mathbb{R}^n$, and let $\bar{x} \in S := G(Q)$ be chosen such that Φ is inner semicompact at $\bar{x}$ w.r.t. S . Then for each $x^* \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$ as well as each pair $(y, v) \in \mathbb{R}^m \times \mathbb{R}^m$ with $G(y) = \bar{x}$ and $\nabla G(y)v = u$, one has*

$$d^2\delta_S(\bar{x}; x^*)(u) \leq -\langle x^*, \nabla^2 G(y)(v, v) \rangle + d^2\delta_Q(y; \nabla G(y)^\top x^*)(v), \quad (20)$$

and the converse holds always true if, additionally, $u \notin T_S(\bar{x})$ or $\langle x^*, u \rangle < 0$ in which cases both sides in (20) equal ∞ .

Given $u \in T_S(\bar{x})$ such that $\langle x^*, u \rangle = 0$, assume that the mapping Φ is inner calm* at $\bar{x}$ in direction u w.r.t. S . Then we have

$$d^2\delta_S(\bar{x}; x^*)(u) = \inf_{G(y)=\bar{x}, \nabla G(y)v=u} \left(-\langle x^*, \nabla^2 G(y)(v, v) \rangle + d^2\delta_Q(y; \nabla G(y)^\top x^*)(v) \right).$$

Furthermore, if $d^2\delta_S(\bar{x}; x^*)(u)$ is finite, then there exists a pair $(y, v) \in \mathbb{R}^m \times \mathbb{R}^m$ with $G(y) = \bar{x}$ and $\nabla G(y)v = u$ such that the estimate (20) holds as equality, i.e., the infimum is attained.

Proof. Setting $\varphi(x, y) := \delta_{(\mathbb{R}^n \times Q) \cap \text{gph } G^{-1}}(x, y)$ for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ yields $\delta_S(x) = \inf_{y \in \mathbb{R}^m} \varphi(x, y)$, i.e., we may apply Theorem 4.8 in order to verify the claim (note that φ is obviously proper, lower semicontinuous, and lower bounded). First, let us show that

$$d^2\varphi((\bar{x}, y); (x^*, 0))(u, v) = -\langle x^*, \nabla^2 G(y)(v, v) \rangle + d^2\delta_Q(y; \nabla G(y)^\top x^*)(v) \quad (21)$$

holds for $(y, v) \in \mathbb{R}^m \times \mathbb{R}^m$ satisfying $G(y) = \bar{x}$ and $\nabla G(y)v = u$.

Taking into account (10), consider sequences $t_k \downarrow 0$ and $(u_k, v_k) \rightarrow (u, v)$ such that $y + t_k v_k \in Q$,

$$\bar{x} + t_k u_k = G(y + t_k v_k) = G(y) + t_k \nabla G(y)v_k + t_k^2/2 \nabla^2 G(y)(v, v) + o(t_k^2),$$

i.e., $u_k = \nabla G(y)v_k + t_k/2 \nabla^2 G(y)(v, v) + o(t_k) \rightarrow \nabla G(y)v$, and

$$d^2\varphi((\bar{x}, y); (x^*, 0))(u, v) = \lim_{k \rightarrow \infty} -2\langle (x^*, 0), (u_k, v_k) \rangle / t_k.$$

Then we immediately find

$$\begin{aligned} d^2\varphi((\bar{x}, y); (x^*, 0))(u, v) &= -\langle x^*, \nabla^2 G(y)(v, v) \rangle + \lim_{k \rightarrow \infty} -2\langle x^*, \nabla G(y)v_k \rangle / t_k \\ &\geq -\langle x^*, \nabla^2 G(y)(v, v) \rangle + d^2\delta_Q(y; \nabla G(y)^\top x^*)(v). \end{aligned}$$

We note that this estimate holds trivially whenever such sequences $t_k \downarrow 0$ and $(u_k, v_k) \rightarrow (u, v)$ do not exist since this gives $(u, v) \notin T_{(\mathbb{R}^n \times Q) \cap \text{gph } G^{-1}}(\bar{x}, y)$ and, thus, $d^2\varphi((\bar{x}, y); (x^*, 0))(u, v) = \infty$.

To show the opposite inequality in (21), consider now sequences $t_k \downarrow 0$ and $v_k \rightarrow v$ such that $y + t_k v_k \in Q$ as well as

$$d^2\delta_Q(y; \nabla G(y)^\top x^*)(v) = \lim_{k \rightarrow \infty} -2\langle \nabla G(y)^\top x^*, v_k \rangle / t_k.$$

Particularly, $G(y) = \bar{x} \in S$ and $G(y + t_k v_k) = \bar{x} + t_k u_k \in S$ for $u_k \in \mathbb{R}^n$ given by

$$u_k := (G(y + t_k v_k) - G(y))/t_k = \nabla G(y)v_k + t_k^2/2 \nabla^2 G(y)(v, v) + o(t_k^2).$$

Similar computations as above thus yield the opposite inequality, taking into account $\langle \nabla G(y)^\top x^*, v_k \rangle = \langle x^*, \nabla G(y)v_k \rangle$ and $u_k \rightarrow \nabla G(y)v = u$. Clearly, the converse inequality in (21) holds trivially if there are no such sequences $t_k \downarrow 0$ and $v_k \rightarrow v$ since this gives $v \notin T_Q(y)$ and, thus, $d^2 \delta_Q(y; \nabla G(y)^\top x^*)(v) = \infty$.

In order to apply Theorem 4.8, we interrelate the inner semicompactness and inner calmness* of $\Upsilon: \mathbb{R}^n \times \mathbb{R} \rightrightarrows \mathbb{R}^m$ given as in (2) with the respective properties of Φ from (19). By definition of φ , we find

$$\Upsilon(x, \alpha) = \begin{cases} \{y \in Q \mid G(y) = x\} & \alpha \geq 0, \\ \emptyset & \alpha < 0 \end{cases} \quad \forall x \in \mathbb{R}^n, \forall \alpha \in \mathbb{R}.$$

Note that we have $\text{dom } \Upsilon = S \times \mathbb{R}_+$. A simple calculation reveals that inner semicompactness of Φ at $\bar{x}$ w.r.t. S gives inner semicompactness of Υ at $(\bar{x}, 0)$ w.r.t. $\text{dom } \Upsilon$ (and vice versa). Furthermore, inner calmness* of Φ at $\bar{x}$ w.r.t. $\text{dom } \Phi$ in direction u gives inner calmness* of Υ at $(\bar{x}, 0)$ w.r.t. $\text{dom } \Upsilon$ in direction $(u, \langle x^*, u \rangle)$.

Now, by means of Theorem 4.8, the general upper estimate in (20) is valid for each pair $(y, v) \in \mathbb{R}^m \times \mathbb{R}^m$ with $G(y) = \bar{x}$ and $\nabla G(y)v = u$ since Υ from above is assumed to be inner semicompact at $(\bar{x}, 0)$ w.r.t. $\text{dom } \Upsilon$. If $u \notin T_S(\bar{x})$, we find $v \notin T_Q(y)$ from [38, Theorem 6.43] and Proposition 3.6 (a) gives that both sides of the estimate equal ∞ . In case $\langle x^*, u \rangle < 0$, we also have $\langle \nabla G(y)^\top x^*, v \rangle = \langle x^*, \nabla G(y)v \rangle = \langle x^*, u \rangle < 0$ and, again, Proposition 3.6 (a) yields that both sides of the estimate equal ∞ .

Now, we address the converse relation. Observing that the marginal function associated with φ satisfies $h := \delta_S$ in our present setting, we find $dh(\bar{x})(u) = 0$ for each $u \in T_S(\bar{x})$, see Example 3.9. Hence, the final statements follow from Theorem 4.8 as well. $\square$

Remark 4.11. We exploit the notation from Proposition 4.10. In [9, Section 5.1.3], it has been pointed out that $\Phi: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ from (19) is inner semicompact at some point $\bar{x} \in S$ w.r.t. its domain S whenever there is a neighborhood $U \subset \mathbb{R}^n$ of $\bar{x}$ such that $\Phi(U)$ is bounded, and this is trivially satisfied whenever Q is compact.

Additionally, Φ is isolatedly calm at $(\bar{x}, y)$ for some $y \in \Phi(\bar{x})$ whenever we have

$$\nabla G(y)v = 0, \quad v \in T_Q(y) \implies v = 0,$$

see [9, Section 5.1.3] again. Whenever this is satisfied for all $y \in \Phi(\bar{x})$, and if Φ is inner semicompact at $\bar{x}$ w.r.t. S , then Φ is inner calm* at $\bar{x}$ w.r.t. S , see [9, Lemma 4.3(ii)].

5. Second-order sufficient conditions in constrained optimization

In this section, we aim to derive second-order sufficient optimality conditions for the constrained optimization problem

$$(P) \quad \min\{f_0(x) \mid F(x) \in C\},$$

for twice continuously differentiable mappings $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$ and $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ as well as a closed set $C \subset \mathbb{R}^m$. The feasible set of (P) will be denoted by $S := F^{-1}(C)$ throughout the section. Clearly, this could be done on the basis of Proposition 2.12 and the pre-image rule stated in Proposition 4.9. This approach improves [30, Theorem 7.1(ii) and Proposition 7.3] as it drops a constraint qualification and/or structural assumptions on C while yielding the quadratic growth condition (9) associated with (P) under the same second-order condition (corresponding to (26) below with $\alpha = 1$). In [7, Theorem 3.3], however, a result was shown, which uses the milder second-order condition (see, again, (26) below), yet gives a stronger statement in terms of the following notion introduced by Penot in [32].

Definition 5.1. A point $\bar{x} \in \mathbb{R}^n$ is said to be an *essential local minimizer of second order* for problem (P) if $\bar{x}$ is feasible and there exist $\varepsilon > 0$ and $\delta > 0$ such that

$$f(x) := \max\{f_0(x) - f_0(\bar{x}), \text{dist}(F(x), C)\} \geq \varepsilon \|x - \bar{x}\|^2 \quad \forall x \in \mathbb{B}_\delta(\bar{x}). \quad (22)$$

In the presence of a mild constraint qualification, (22) is equivalent to the quadratic growth condition at $\bar{x}$, see [7, Lemma 3.5]. In general, (22) is stronger; in fact, it is equivalent to the quadratic growth of the function f (which vanishes at $\bar{x}$) and thus fully characterized by its second subderivative by Proposition 2.11. Hence, we estimate the second subderivative of f , showing versatility of the calculus rules in the process, and fully recover [7, Theorem 3.3]. To this end, for each constant $\alpha \geq 0$, let us introduce the associated Lagrangian function $L^\alpha: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ given by

$$L^\alpha(x, \lambda) := \alpha f_0(x) + \langle \lambda, F(x) \rangle \quad \forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m. \quad (23)$$

Furthermore, for each $\bar{x} \in S$, we introduce the critical cone to S at $\bar{x}$ by means of

$$\mathcal{C}(\bar{x}) := \{u \in \mathbb{R}^n \mid \nabla F(\bar{x})u \in T_C(F(\bar{x})), \langle \nabla f_0(\bar{x}), u \rangle \leq 0\}. \quad (24)$$

We use the multiplier set $\Lambda^\alpha(\bar{x})$ defined by

$$\Lambda^\alpha(\bar{x}) := \{\lambda \in \mathbb{R}^m \mid \nabla_x L^\alpha(\bar{x}, \lambda) = 0\}. \quad (25)$$

Theorem 5.2. Let the point $\bar{x} \in \mathbb{R}^n$ be a feasible point of (P). Assume that for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, there are $\alpha \geq 0$ and a multiplier $\lambda \in \Lambda^\alpha(\bar{x})$ such that

$$\nabla_{xx}^2 L^\alpha(\bar{x}, \lambda)(u, u) + d^2 \delta_C(F(\bar{x}); \lambda)(\nabla F(\bar{x})u) > 0. \quad (26)$$

Then $\bar{x}$ is an essential local minimizer of second order for (P).

Proof. The statement follows from Proposition 2.11 once we show $d^2 f(\bar{x}; 0)(u) > 0$ for all $u \in \mathbb{R}^n \setminus \{0\}$, where f is defined in (22). Note that $f = \text{vecmax} \circ G$ for $G: \mathbb{R}^n \rightarrow \mathbb{R}^2$ given by $G(x) := (G_1(x), G_2(x)) := (f_0(x) - f_0(\bar{x}), \text{dist}(F(x), C))$ for all $x \in \mathbb{R}^n$.

If $d f(\bar{x})(u) > \langle 0, u \rangle = 0$, we get $d^2 f(\bar{x}; 0)(u) = \infty$ by (5). Thus, let us assume that $d f(\bar{x})(u) \leq 0$ holds for $u \neq 0$. The chain rule Theorem 4.1 yield the existence of $v \in DG(\bar{x})(u)$ such that

$$d f(\bar{x})(u) \geq d \langle y^*, G \rangle(\bar{x})(u) + d \text{vecmax}((0, 0))(v) - \langle y^*, v \rangle, \quad (27a)$$

$$d^2 f(\bar{x}; 0)(u) \geq d^2 \langle y^*, G \rangle(\bar{x}; 0)(u) + d^2 \text{vecmax}((0, 0); y^*)(v) \quad (27b)$$

holds for each $y^* \in \mathbb{R}^2$ (which gives well-defined expressions). Let us point out that $v = (v_1, v_2) \in DG(\bar{x})(u)$ means $v_1 = \nabla f_0(\bar{x})u$ and

$$\begin{aligned} v_2 &\geq d \operatorname{dist}(F(\cdot), C)(\bar{x})(u) \geq d \operatorname{dist}(\cdot, C)(F(\bar{x}))(\nabla F(\bar{x})u) \\ &= \operatorname{dist}(\nabla F(\bar{x})u, T_C(F(\bar{x}))) \geq 0 \end{aligned} \quad (28)$$

by the chain rule from Theorem 4.1 (with $y^* := 0$), [38, Example 8.53], and the following argument: From $v_2 \in DG_2(\bar{x})(u)$, we find sequences $u_k \rightarrow u$, $v_{2k} \rightarrow v_2$, and $t_k \downarrow 0$ such that $v_{2k} = \operatorname{dist}(F(\bar{x} + t_k u_k), C)/t_k$ for all $k \in \mathbb{N}$, so taking the limit as $k \rightarrow \infty$ while respecting the definition of the subderivative gives the first estimate.

Using (27a) with $y^* = 0$, Lemma 2.10, and Lipschitzianity of the distance function, we obtain

$$0 \geq d f(\bar{x})(u) \geq d \operatorname{vecmax}((0, 0))(v) = \max\{\nabla f_0(\bar{x})u, v_2\},$$

which together with (28) gives $v_2 = 0$ and $u \in \mathcal{C}(\bar{x})$.

Now, our assumptions guarantee the existence of $\alpha \geq 0$ and $\lambda \in \Lambda^\alpha(\bar{x})$ satisfying (26), which also implies that α and λ cannot be zero simultaneously. Moreover, if $\nabla f_0(\bar{x})u < 0$, (26) together with $\lambda \in \Lambda^\alpha(\bar{x})$ also imply $\alpha = 0$, for otherwise $\langle \lambda, \nabla F(\bar{x})u \rangle > 0$ and $d^2 \delta_C(F(\bar{x}); \lambda)(\nabla F(\bar{x})u) = -\infty$ by Proposition 3.6 (b). Thus, we have $\alpha \nabla f_0(\bar{x})u = 0$. We will show that one has

$$d^2 f(\bar{x}; 0)(u) \geq \frac{1}{\alpha + \|\lambda\|} \left(\nabla_{xx}^2 L^\alpha(\bar{x}, \lambda)(u, u) + d^2 \delta_C(F(\bar{x}); \lambda)(\nabla F(\bar{x})u) \right). \quad (29)$$

Set $\hat{\alpha} := \alpha/(\alpha + \|\lambda\|)$, note that $1 - \hat{\alpha} = \|\lambda\|/(\alpha + \|\lambda\|)$, and apply (27b) with $y^* := (\hat{\alpha}, 1 - \hat{\alpha})$ as well as Lemma 2.10 to obtain

$$\begin{aligned} d^2 f(\bar{x}; 0)(u) &\geq d^2 (\hat{\alpha} G_1 + (1 - \hat{\alpha}) G_2)(\bar{x}; 0)(u) + d^2 \operatorname{vecmax}((0, 0); (\hat{\alpha}, 1 - \hat{\alpha}))(\nabla f_0(\bar{x})u, 0) \\ &= d^2 (\hat{\alpha} G_1 + (1 - \hat{\alpha}) G_2)(\bar{x}; 0)(u). \end{aligned}$$

If $\lambda = 0$, we get $\alpha > 0$, $\hat{\alpha} = 1$, and $\nabla f_0(\bar{x})u = 0$ from above, and (29) follows from (7). Otherwise, if $\lambda \neq 0$, we set $\hat{\lambda} := \lambda/\|\lambda\|$ and continue with the sum rule from Proposition 2.8 (a) as well as the homogeneity property stated in (4). Additionally, applying the chain rule from Theorem 4.1 with $y^* := \hat{\lambda}$ and Corollary 4.7 then yields

$$\begin{aligned} d^2 (\hat{\alpha} G_1 + (1 - \hat{\alpha}) G_2)(\bar{x}; 0)(u) &\geq \hat{\alpha} \nabla^2 f_0(\bar{x})(u, u) + (1 - \hat{\alpha}) d^2 G_2 \left(\bar{x}; -\frac{\hat{\alpha}}{1 - \hat{\alpha}} \nabla f_0(\bar{x}) \right) (u) \\ &\geq \hat{\alpha} \nabla^2 f_0(\bar{x})(u, u) + (1 - \hat{\alpha}) \nabla^2 \langle \hat{\lambda}, F \rangle(\bar{x})(u, u) \\ &\quad + (1 - \hat{\alpha}) d^2 \operatorname{dist}(\cdot, C)(F(\bar{x}); \hat{\lambda})(\nabla F(\bar{x})u) \\ &\geq \frac{1}{\alpha + \|\lambda\|} \left(\nabla_{xx}^2 L^\alpha(\bar{x}, \lambda)(u, u) + d^2 \delta_C(F(\bar{x}); \lambda)(\nabla F(\bar{x})u) \right), \end{aligned}$$

noting that we have $\nabla F(\bar{x})^\top \hat{\lambda} = -\alpha/\|\lambda\| \nabla f_0(\bar{x}) = -\hat{\alpha}/(1 - \hat{\alpha}) \nabla f_0(\bar{x})$ as well as $(1 - \hat{\alpha}) \hat{\lambda} = \lambda/(\alpha + \|\lambda\|)$ and taking into account (11). This completes the proof. $\square$

In [7, Theorem 3.3], this result has been proven via a classical contradiction argument, i.e., via the standard approach to obtain second-order sufficient optimality conditions, while our proof is direct, relying on the calculus rules. In general, the calculus-based approach is certainly very convenient as one just applies the formulas, but it may not always be the best since it uses artificial steps, which can be accompanied with artificial assumptions. In case of second subderivatives, however, it seems that the calculus works very well since it can handle even the complicated structure of the function f from (22) without adding superfluous requirements or losing valuable information.

As pointed out in [7, Proposition 3.4], there is some additional information about the multipliers available in Theorem 5.2 which can be distilled from (26).

Remark 5.3. Let the assumptions of Theorem 5.2 be valid. Then we can conclude that for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, (26) yields $d^2\delta_C(F(\bar{x}); \lambda)(\nabla F(\bar{x})u) > -\infty$ for some $\lambda \in \Lambda^\alpha(\bar{x})$ where $\alpha \geq 0$. By means of Proposition 3.6, this immediately yields $\lambda \in \widehat{N}_C^p(F(\bar{x}); \nabla F(\bar{x})u)$. Due to (12), we have $\langle \lambda, \nabla F(\bar{x})u \rangle \leq 0$. On the other hand, the definitions of $\mathcal{C}(\bar{x})$ and $\Lambda^\alpha(\bar{x})$ give

$$0 \leq \langle -\alpha \nabla f_0(\bar{x}), u \rangle = \langle \nabla F(\bar{x})^\top \lambda, u \rangle = \langle \lambda, \nabla F(\bar{x})u \rangle,$$

and $\langle \lambda, \nabla F(\bar{x})u \rangle = 0$ follows. Therefore, choosing $\lambda \in \widehat{N}_C^p(F(\bar{x}); \nabla F(\bar{x})u)$ is always possible.

In the following subsections, we discuss various constrained optimization problems of the form (P). More precisely, we investigate three settings differing from each other by the particular structure of the set C :

Section 5.1: C is polyhedral, i.e., it has no curvature (standard nonlinear or disjunctive programs);

Section 5.2: C is curved, but simple (nonlinear second-order cone programs);

Section 5.3: C is an image of a pre-image of a simple set (bilevel programs, programs with (quasi-) variational inequality constraints).

5.1. Disjunctive programs

Here, we apply Theorem 5.2 to so-called disjunctive programs where $C := \bigcup_{i=1}^\ell P_i$ holds for convex polyhedral sets $P_1, \dots, P_\ell \subset \mathbb{R}^m$. Then $S = F^{-1}(C)$ can be used to represent feasible sets modeled via complementarity-, cardinality-, switching-, or vanishing-type constraints, exemplary, but also standard nonlinear optimization problems, see e.g. [5, 17, 29] for an introduction to disjunctive programming and suitable references for more information on the aforementioned subclasses.

For $\bar{y} \in C$, we make use of the index set $J(\bar{y}) := \{i \in \{1, \dots, \ell\} \mid \bar{y} \in P_i\}$. Then, for some $\bar{x} \in S$, we find

$$\mathcal{C}(\bar{x}) = \left\{ w \in \mathbb{R}^n \mid \nabla F(\bar{x})u \in \bigcup_{i \in J(F(\bar{x}))} T_{P_i}(F(\bar{x})), \langle \nabla f_0(\bar{x}), u \rangle \leq 0 \right\}$$

e.g. from [1, Table 4.1]. Furthermore, for each $u \in \mathcal{C}(\bar{x})$, we make use of the index set

$$J(\bar{x}; u) := \{i \in J(F(\bar{x})) \mid \nabla F(\bar{x})u \in T_{P_i}(F(\bar{x}))\}.$$

Based on Theorem 5.2, we find the following second-order sufficient optimality condition for the associated optimization problem (P).

Theorem 5.4. *Let $\bar{x} \in \mathbb{R}^n$ be a feasible point of (P) where $C := \bigcup_{i=1}^{\ell} P_i$ holds for convex polyhedral sets $P_1, \dots, P_{\ell} \subset \mathbb{R}^m$. Assume that for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, there are $\alpha \geq 0$ and a multiplier $\lambda \in \Lambda^{\alpha}(\bar{x}) \cap \bigcap_{i \in J(\bar{x}; u)} N_{T_{P_i}(F(\bar{x}))}(\nabla F(\bar{x})u)$ such that*

$$\nabla_{xx}^2 L^{\alpha}(\bar{x}, \lambda)(u, u) > 0. \quad (30)$$

Then $\bar{x}$ is an essential local minimizer of second order for (P).

Proof. Theorem 5.2 shows that the assertion of the corollary is true whenever for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, we find $\alpha \geq 0$ and some $\lambda \in \Lambda^{\alpha}(\bar{x})$ satisfying (26). The assumptions guarantee that, for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, we find $\alpha \geq 0$ and a multiplier $\lambda \in \Lambda^{\alpha}(\bar{x}) \cap \bigcap_{i \in J(\bar{x}; u)} N_{T_{P_i}(F(\bar{x}))}(\nabla F(\bar{x})u)$ with (30). Lemma 3.2 shows that we have $d^2 \delta_C(F(\bar{x}); \lambda)(\nabla F(\bar{x})u) \in \{0, \infty\}$ in this case. Thus, (30) implies (26). $\square$

Let us note that under some additional assumptions and in case where $\ell := 1$ and $\alpha := 1$, a second-order sufficient condition similar to the one from Theorem 5.4 has been obtained in [38, Example 13.25]. The sufficient conditions from [40, Theorem 4.1] reduce to ours when applied to the present situation. However, the authors present them in the presence of an additional qualification condition. Furthermore, Theorem 5.4 recovers the sufficient conditions from [19, Theorem 3.17] and [7, Theorem 6.1], taking into account that we have

$$\begin{aligned} \widehat{N}_C^p(F(\bar{x}); \nabla F(\bar{x})u) &= \widehat{N}_{T_C(F(\bar{x}))}(\nabla F(\bar{x})u) \\ &= \widehat{N}_{\bigcup_{i \in J(F(\bar{x}))} T_{P_i}(F(\bar{x}))}(\nabla F(\bar{x})u) = \bigcap_{i \in J(\bar{x}; u)} N_{T_{P_i}(F(\bar{x}))}(\nabla F(\bar{x})u) \end{aligned}$$

from [7, Remark 5.2] and [5, formula (22)].

Simplicity of the second-order sufficient conditions from Theorem 5.4 is caused by the fact that, despite being variationally difficult and highly non-convex, unions of finitely many convex polyhedral sets are not *curved* causing the second subderivative of the associated indicator function to be zero if finite, see Lemma 3.2. In the next section, we consider the situation where C is an instance of the well-known second-order cone which possesses curvature.

5.2. Nonlinear second-order cone programming

Let us take a closer look at a popular situation where the abstract set C in the setting of (P) is curved. Therefore, recall that for a given integer $s \geq 3$, the set

$$\mathcal{Q}_s := \{y \in \mathbb{R}^s \mid (y_2^2 + \dots + y_s^2)^{1/2} \leq y_1\}$$

is referred to as second-order or ice-cream cone in $\mathbb{R}^s$.

First, let us compute the second subderivative of the indicator function associated with a second-order cone. Therefore, we coin some additional notation as follows:

$$\|v\| := (v_2^2 + \dots + v_s^2)^{1/2}, \quad \langle\langle y, v \rangle\rangle := \sum_{i=2}^s y_i v_i \quad \forall v, y \in \mathbb{R}^s.$$

Lemma 5.5. *Fix an integer $s \geq 3$. For each $\bar{y} \in \mathcal{Q}_s$, $y^* \in \mathbb{R}^s$, and $v \in \mathbb{R}^s$, the following assertions hold.*

(a) *For $\bar{y} \in \text{int } \mathcal{Q}_s$, we have*

$$d^2 \delta_{\mathcal{Q}_s}(\bar{y}; y^*)(v) = \begin{cases} \infty & \langle y^*, v \rangle < 0, \\ 0 & y^* = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

(b) *For $\bar{y} := 0$, we have*

$$d^2 \delta_{\mathcal{Q}_s}(\bar{y}; y^*)(v) = \begin{cases} \infty & \langle y^*, v \rangle < 0 \text{ or } v \notin \mathcal{Q}_s, \\ 0 & y^* \in -\mathcal{Q}_s \cap \{v\}^\perp, v \in \mathcal{Q}_s, \\ -\infty & \text{otherwise.} \end{cases}$$

(c) *For $\bar{y} \in \text{bdry } \mathcal{Q}_s \setminus \{0\}$, we have*

$$d^2 \delta_{\mathcal{Q}_s}(\bar{y}; y^*)(v) \geq \begin{cases} \frac{\|y^*\|}{\|\bar{y}\|} (\|v\|^2 - v_1^2) & y^* = \beta q(\bar{y}), \beta \geq 0, \\ & \langle\langle \bar{y}, v \rangle\rangle \leq \bar{y}_1 v_1, \langle y^*, v \rangle = 0, \\ \infty & y^* = \beta q(\bar{y}) \text{ and} \\ & (\langle\langle \bar{y}, v \rangle\rangle > \bar{y}_1 v_1 \text{ or } \langle y^*, v \rangle < 0), \\ -\infty & \text{otherwise,} \end{cases}$$

where $q(\bar{y}) := \|\bar{y}\|^{-1} \sum_{i=2}^s \bar{y}_i \mathbf{e}_i - \mathbf{e}_1$. Furthermore, equality holds in the first two cases.

(d) *Whenever $y^* \in N_{\mathcal{Q}_s}(\bar{y}) \cap \{v\}^\perp$ holds, we find*

$$d^2 \delta_{\mathcal{Q}_s}(\bar{y}; y^*)(v) = \begin{cases} 0 & \bar{y} \in \text{int } \mathcal{Q}_s, \\ 0 & \bar{y} = 0, v \in \mathcal{Q}_s, \\ \frac{\|y^*\|}{\|\bar{y}\|} (\|v\|^2 - v_1^2) & \bar{y} \in \text{bdry } \mathcal{Q}_s \setminus \{0\}, \langle\langle \bar{y}, v \rangle\rangle \leq \bar{y}_1 v_1, \\ \infty & v \notin T_{\mathcal{Q}_s}(\bar{y}). \end{cases}$$

Proof. The assertion of the first statement is trivial.

For the proof of the second statement, we observe that, since $\mathcal{Q}_s$ is a cone, we find

$$d^2 \delta_{\mathcal{Q}_s}(\bar{y}; y^*)(v) = \liminf_{t \downarrow 0, v' \rightarrow v, v' \in \mathcal{Q}_s} \frac{-2\langle y^*, v' \rangle}{t}.$$

The cases where $\langle y^*, v \rangle < 0$ or $v \notin \mathcal{Q}_s = T_{\mathcal{Q}_s}(\bar{y})$ and $\langle y^*, v \rangle > 0$ are clear, so let us assume $\langle y^*, v \rangle = 0$ and $v \in \mathcal{Q}_s$. In case where $y^* \in -\mathcal{Q}_s$, we find $\langle y^*, v' \rangle \leq 0$ for all

$v' \in \mathcal{Q}_s$, and the second subderivative obviously vanishes (here, we used that the polar cone of $\mathcal{Q}_s$ is $-\mathcal{Q}_s$). Otherwise, there is $\tilde{v} \in \mathcal{Q}_s$ with $\langle y^*, \tilde{v} \rangle > 0$. By convexity of $\mathcal{Q}_s$, we have $v + \tilde{v}/k \in \mathcal{Q}_s$ for each $k \in \mathbb{N}$, and

$$-2\langle y^*, v + \tilde{v}/k \rangle / (1/k^2) = -2k\langle y^*, \tilde{v} \rangle \rightarrow -\infty$$

shows that the second subderivative is $-\infty$.

For the proof of the third statement, we introduce a function $\phi: \mathbb{R}^s \rightarrow \mathbb{R}$ by means of $\phi(y) := \|y\| - y_1$ for all $y \in \mathbb{R}^s$ and observe that, due to $\bar{y} \in \text{bdry } \mathcal{Q}_s \setminus \{0\}$, we have $\phi(\bar{y}) = 0$ and ϕ is twice continuously differentiable at $\bar{y}$ with non-vanishing gradient $q(\bar{y})$. Furthermore, we have $\mathcal{Q}_s = \phi^{-1}(\mathbb{R}_-)$, i.e., Proposition 4.9 can be applied to get a formula for the second subderivative. More precisely, using

$$Q(\bar{y}) := \frac{1}{\|\bar{y}\|} \text{diag}(\mathbf{e}_2 + \dots + \mathbf{e}_s) - \frac{1}{\|\bar{y}\|^3} \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & \bar{y}_2^2 & \dots & \bar{y}_2 \bar{y}_s \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \bar{y}_s \bar{y}_2 & \dots & \bar{y}_s^2 \end{pmatrix},$$

we find

$$d^2 \delta_{\mathcal{Q}_s}(\bar{y}; y^*)(v) \geq \begin{cases} \beta Q(\bar{y})(v, v) + d^2 \delta_{\mathbb{R}_-}(0; \beta)(\langle q(\bar{y}), v \rangle) & y^* = \beta q(\bar{y}), \\ -\infty & \text{otherwise,} \end{cases}$$

and equality holds in the first case. Taking Example 3.1 into account, this gives

$$d^2 \delta_{\mathcal{Q}_s}(\bar{y}; y^*)(v) \geq \begin{cases} \beta Q(\bar{y})(v, v) & y^* = \beta q(\bar{y}), \beta \geq 0, \langle q(\bar{y}), v \rangle \leq 0, \langle y^*, v \rangle = 0, \\ \infty & y^* = \beta q(\bar{y}) \text{ and } (\langle q(\bar{y}), v \rangle > 0 \text{ or } \langle y^*, v \rangle < 0), \\ -\infty & \text{otherwise,} \end{cases}$$

and equality holds in the first two cases. Finally, let us simplify the expression $\beta Q(\bar{y})(v, v)$ in the first of the appearing cases. Thus, let us assume that $y^* = \beta q(\bar{y})$ for some $\beta \geq 0$ such that $\langle q(\bar{y}), v \rangle \leq 0$ and $\langle y^*, v \rangle = 0$. The case where $\beta = 0$ is trivial. Therefore, let us assume $\beta > 0$. Then $\langle y^*, v \rangle = 0$ gives $\|\bar{y}\|v_1 = \sum_{i=2}^s \bar{y}_i v_i$. From $\bar{y} \in \text{bdry } \mathcal{Q}_s \setminus \{0\}$, we have $\|\bar{y}\|^2 = \bar{y}_1^2 + \|\bar{y}\|^2 = 2\|\bar{y}\|^2$. Additionally, $y^* = \beta q(\bar{y})$ gives $\|y^*\|^2 = \beta^2 \|q(\bar{y})\|^2 = \beta^2 (1 + \|\bar{y}\|^{-2} \sum_{i=2}^s \bar{y}_i^2) = 2\beta^2$. Thus, we end up with

$$\begin{aligned} \beta Q(\bar{y})(v, v) &= \beta \left(\frac{\|v\|^2}{\|\bar{y}\|} - \frac{1}{\|\bar{y}\|^3} \left(\sum_{i=2}^s \bar{y}_i v_i \right)^2 \right) \\ &= \beta \left(\frac{\|v\|^2}{\|\bar{y}\|} - \frac{v_1^2}{\|\bar{y}\|} \right) = \frac{\sqrt{2}\beta}{\sqrt{2}\|\bar{y}\|} (\|v\|^2 - v_1^2) = \frac{\|y^*\|}{\|\bar{y}\|} (\|v\|^2 - v_1^2), \end{aligned}$$

and the assertion follows from

$$\langle q(\bar{y}), v \rangle \leq 0 \iff -\|\bar{y}\|v_1 + \sum_{i=2}^s \bar{y}_i v_i \leq 0 \iff \langle \langle \bar{y}, v \rangle \rangle \leq \bar{y}_1 v_1$$

since $\bar{y}_1 = \|\bar{y}\|$. Note that the above equivalent expressions provide representations of the condition $v \in T_{\mathcal{Q}_s}(\bar{y})$ in the given situation.

The final statement follows from the first three. $\square$

In [24, Theorem 3.1], it has been shown that $\delta_{\mathcal{Q}_s}$ is already so-called twice epi-differentiable, and that its second epi-derivative can be calculated by means of the formula stated in the final statement of the above lemma. However, the proof of [24, Theorem 3.1] is much more technical than our proof of Lemma 5.5 which, particularly for the difficult case $\bar{y} \in \text{bdry } \mathcal{Q}_s \setminus \{0\}$, exploits the pre-image rule from Proposition 4.9.

Note that the simple pre-image structure $\mathcal{Q}_s = \phi^{-1}(\mathbb{R}_-)$ of the set $\mathcal{Q}_s$ for the function $\phi(y) = \|y\| - y_1$ is clearly valid for all points in $\mathcal{Q}_s$, but ϕ is not differentiable at the origin.

For given integers $m_1, \dots, m_\ell \geq 3$ such that $m := m_1 + \dots + m_\ell$, we consider

$$C := \prod_{i=1}^{\ell} \mathcal{Q}_{m_i} \quad (31)$$

in problem (P) with twice continuously differentiable data functions $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$ and $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Furthermore, let $F_i: \mathbb{R}^n \rightarrow \mathbb{R}^{m_i}$, $i = 1, \dots, \ell$, be the component mappings of F such that

$$F(x) \in C \iff F_i(x) \in \mathcal{Q}_{m_i} \quad \forall i \in \{1, \dots, \ell\}.$$

Based on Theorem 5.2, we find the following second-order sufficient optimality conditions for the associated problem (P).

Theorem 5.6. *Let $\bar{x} \in \mathbb{R}^n$ be a feasible point of (P) where C is given as in (31). Assume that for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, there are $\alpha \geq 0$ and multipliers $\lambda_i \in N_{\mathcal{Q}_{m_i}}(F_i(\bar{x})) \cap \{\nabla F_i(\bar{x})u\}^\perp$, $i = 1, \dots, \ell$, such that*

$$\begin{aligned} 0 &= \alpha \nabla f_0(\bar{x}) + \sum_{i=1}^{\ell} \nabla F_i(\bar{x})^\top \lambda_i, \\ 0 &< \left(\alpha \nabla^2 f_0(\bar{x}) + \sum_{i=1}^{\ell} \sum_{j=1}^{m_i} (\lambda_i)_j \nabla^2 (F_i)_j(\bar{x}) \right) (u, u) \\ &\quad + \sum_{i \in \mathcal{I}(\bar{x})} \frac{\|\lambda_i\|}{\|F_i(\bar{x})\|} (\| \nabla F_i(\bar{x})u \|^2 - (\nabla F_i(\bar{x})u)_1^2), \end{aligned}$$

where $\mathcal{I}(\bar{x}) := \{i \in \{1, \dots, \ell\} \mid F_i(\bar{x}) \in \text{bdry } \mathcal{Q}_{m_i} \setminus \{0\}\}$. Then $\bar{x}$ is an essential local minimizer of second order for (P).

Proof. We note that the convexity of the cones $\mathcal{Q}_{m_i}$, $i = 1, \dots, \ell$, yields

$$T_C(F(\bar{x})) = \prod_{i=1}^{\ell} T_{\mathcal{Q}_{m_i}}(F_i(\bar{x})).$$

Thus, $u \in \mathcal{C}(\bar{x})$ satisfies $\nabla F_i(\bar{x})u \in T_{\mathcal{Q}_{m_i}}(F_i(\bar{x}))$, and we find

$$\begin{aligned} \widehat{N}_C^p(F(\bar{x}); \nabla F(\bar{x})u) &= N_C(F(\bar{x}); \nabla F(\bar{x})u) = N_C(F(\bar{x})) \cap \{\nabla F(\bar{x})u\}^\perp \\ &= \left(\prod_{i=1}^{\ell} N_{\mathcal{Q}_{m_i}}(F_i(\bar{x})) \right) \cap \{\nabla F(\bar{x})u\}^\perp \supset \prod_{i=1}^{\ell} N_{\mathcal{Q}_{m_i}}(F_i(\bar{x})) \cap \{\nabla F_i(\bar{x})u\}^\perp. \end{aligned}$$

Keeping Lemmas 3.3 and 5.5 as well as the discussion right after Theorem 5.2 in mind, the assumptions of Theorem 5.6 imply validity of the assumptions of Theorem 5.2 and the assertion follows. $\square$

A related second-order condition, which comprises a non-vanishing term that incorporates the curvature of the second-order cone, has been obtained in [11, Theorem 29]. The latter result, however, has been formulated in the presence of a constraint qualification and makes use of the same Lagrange multiplier for each non-vanishing critical direction. As shown above, this is not necessary when dealing with sufficient optimality conditions. Similarly, our result enhances the sufficient optimality condition presented in [25, Proposition 2.1].

5.3. Structured geometric constraints

Observe that the set C in (P) on its own could be the image or pre-image of another closed set. More precisely, it has been mentioned in [7, Section 1] that the setting

$$C := H(Q), \quad Q := \{z \in \mathbb{R}^\ell \mid G(z) \in D\} \quad (32)$$

for twice continuously differentiable mappings $G: \mathbb{R}^\ell \rightarrow \mathbb{R}^p$ and $H: \mathbb{R}^\ell \rightarrow \mathbb{R}^m$ as well as a simple, closed set $D \subset \mathbb{R}^p$ (e.g., a polyhedral set) is of significant interest since it covers the special situations where (P) is a bilevel optimization problem or an optimization problem with (quasi-) variational inequality constraints, see e.g. [15, 16, 27, 31]. Note that essentially the same structure has also been recognized in [8], where the set C corresponds to the graph of a set-valued mapping which is a composition of two other mappings, and the intermediate variables were named *implicit variables* therein. It is well known that the graph of a composition possesses the structure (32), see also Example 5.9 below.

These considerations underline the need not only for the standard pre-image rule from Proposition 4.9, but also for the image rule from Proposition 4.10, which is valid under the comparatively mild inner calmness* assumption. The challenging setting from (32) will be investigated deeply in a larger context in a forthcoming paper by the authors of [7] so we do not provide many details here. Instead, we just want to emphasize that the calculus for second subderivatives is so satisfying that it allows to handle even such challenging structures with ease as long as second-order sufficient optimality conditions are under consideration.

Let us apply our results from Section 4.2 in order to find a lower estimate for the second subderivative of the indicator function associated with C from (32).

Lemma 5.7. *Fix $\bar{y} \in C$ given in (32), $y^* \in \mathbb{R}^m$, and $v \in \mathbb{R}^m$. Define $\Phi: \mathbb{R}^m \rightrightarrows \mathbb{R}^\ell$ by means of*

$$\Phi(y) := \{z \in \mathbb{R}^\ell \mid G(z) \in D, H(z) = y\} \quad \forall y \in \mathbb{R}^m, \quad (33)$$

and assume that Φ is inner semicompact at $\bar{y}$ w.r.t. $C = \text{dom } \Phi$ and, if $\langle y^, v \rangle = 0$, inner calm* at $\bar{y}$ in direction v w.r.t. C .*

Then we have

$$d^2\delta_C(\bar{y}; y^*)(v) \geq \inf_{z \in \Phi(\bar{y}), \nabla H(z)w=v} \sup_{\nabla G(z)^\top \eta = \nabla H(z)^\top y^*} \vartheta(z, w, y^*, \eta)$$

for the function $\vartheta: \mathbb{R}^\ell \times \mathbb{R}^\ell \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ given by

$$\begin{aligned} \vartheta(z, w, y^*, \eta) \\ := \langle \eta, \nabla^2 G(z)(w, w) \rangle - \langle y^*, \nabla^2 H(z)(w, w) \rangle + d^2\delta_D(G(z); \eta)(\nabla G(z)w) \end{aligned} \quad (34)$$

for arbitrary $z, w \in \mathbb{R}^\ell$, $y^* \in \mathbb{R}^m$, and $\eta \in \mathbb{R}^p$.

Proof. Due to Proposition 4.10, the assumptions of the lemma guarantee

$$d^2\delta_C(\bar{y}; y^*)(v) = \inf_{z \in \Phi(\bar{y}), \nabla H(z)w=v} \left(-\langle y^*, \nabla^2 H(z)(w, w) \rangle + d^2\delta_Q(z; \nabla H(z)^\top y^*)(w) \right).$$

Now, we apply Proposition 4.9 to obtain the fully explicit lower estimate. $\square$

Relying on Remark 4.11, we obtain that in case $\langle y^*, v \rangle = 0$, the necessary inner calmness* of the mapping Φ is inherent whenever the qualification condition

$$\nabla H(z)w = 0, \quad \nabla G(z)w \in T_D(G(z)) \implies w = 0 \quad \forall z \in \Phi(\bar{y}) \quad (35)$$

is valid since we have $T_Q(z) = T_{G^{-1}(D)}(z) \subset \nabla G(z)^{-1}T_D(G(z))$ for each $z \in \Phi(\bar{y})$ from [38, Theorem 6.31].

Let us note that the assumptions of Lemma 5.7 hold trivially if $H: \mathbb{R}^m \rightarrow \mathbb{R}^m$ is continuously invertible at $\bar{z} := H^{-1}(\bar{y})$ such that $\nabla H(\bar{z})$ is regular. In this case, the given lower estimate for the second subderivative simplifies to

$$d^2\delta_C(\bar{y}; y^*)(v) \geq \sup_{\nabla G(\bar{z})^\top \eta = \nabla H(\bar{z})^\top y^*} \vartheta(\bar{z}, \bar{w}, y^*, \eta)$$

where $\bar{w} := \nabla H(\bar{z})^{-1}v$. Particularly, whenever the function $H: \mathbb{R}^m \rightarrow \mathbb{R}^m$ is given by $H(z) := Az - b$ for all $z \in \mathbb{R}^m$ where $A \in \mathbb{R}^{m \times m}$ is a regular matrix and $b \in \mathbb{R}^m$ is a vector, i.e., whenever H models just a change of coordinates, the above formula applies and the term involving the second derivative of H vanishes. The situation where H is the identity map corresponds to the setting where C itself is just a pre-image.

Remark 5.8. Let us note that in the setting discussed in [7, Section 1, formula (3)], the mapping $\Phi: \mathbb{R}^m \rightrightarrows \mathbb{R}^\ell$ from (33) is closely related to the Lagrange multiplier mapping of another given variational problem. Inner semicompactness and inner calmness* of such mappings have been shown to be valid under reasonably mild constraint qualifications, see [4, Theorem 3.9] for details.

Example 5.9. For set-valued mappings $M_1: \mathbb{R}^{m_1} \rightrightarrows \mathbb{R}^{m_2}$ and $M_2: \mathbb{R}^{m_2} \rightrightarrows \mathbb{R}^{m_3}$ with a closed graph, we consider the situation where $C := \text{gph}(M_2 \circ M_1)$ holds. Here, $M_2 \circ M_1: \mathbb{R}^{m_1} \rightrightarrows \mathbb{R}^{m_3}$ is the composition of M_1 and M_2 given by

$$(M_2 \circ M_1)(y_1) := \bigcup_{y_2 \in M_1(y_1)} M_2(y_2) \quad \forall y_1 \in \mathbb{R}^{m_1}.$$

Setting $\ell := m_1 + m_2 + m_3$, $m := m_1 + m_3$, $p := m_1 + m_2 + m_2 + m_3$,

$$H(z_1, z_2, z_3) := (z_1, z_3), \quad G(z_1, z_2, z_3) := (z_1, z_2, z_2, z_3) \quad \forall (z_1, z_2, z_3) \in \mathbb{R}^\ell,$$

and $D := \text{gph } M_1 \times \text{gph } M_2$, we find $C = H(G^{-1}(D))$. In the present situation, we have $\Phi(y_1, y_3) = \{(y_1, y_2, y_3) \in \mathbb{R}^\ell \mid y_2 \in M_1(y_1), y_3 \in M_2(y_2)\}$, i.e., Φ is closely related to the so-called intermediate mapping $\Theta: \mathbb{R}^m \rightrightarrows \mathbb{R}^{m_2}$, given by

$$\Theta(y_1, y_3) := \{y_2 \in M_1(y_1) \mid y_3 \in M_2(y_2)\} \quad \forall (y_1, y_3) \in \mathbb{R}^m,$$

which is associated with the composition $M_2 \circ M_1$, see [9, Section 5.3]. Clearly, for given $(\bar{y}_1, \bar{y}_3) \in \text{dom } \Phi$, Φ is inner semicompact at $(\bar{y}_1, \bar{y}_3)$ w.r.t. $\text{dom } \Phi$ (inner calm* at $(\bar{y}_1, \bar{y}_3)$ in direction $(v_1, v_3) \in \mathbb{R}^m$ w.r.t. $\text{dom } \Phi$) if and only if Θ is inner semicompact at $(\bar{y}_1, \bar{y}_3)$ w.r.t. $\text{dom } \Theta = \text{dom } \Phi$ (inner calm* at $(\bar{y}_1, \bar{y}_3)$ in direction $(v_1, v_3) \in \mathbb{R}^m$ w.r.t. $\text{dom } \Theta$), and the latter is trivially satisfied if M_1 is single-valued and continuous (single-valued and calm in direction (v_1, v_3)). Observe that the qualification condition (35) is implied by

$$(0, w_2) \in T_{\text{gph } M_1}(\bar{y}_1, y_2), (w_2, 0) \in T_{\text{gph } M_2}(y_2, \bar{y}_3) \implies w_2 = 0 \quad \forall y_2 \in \Theta(\bar{y}_1, \bar{y}_3),$$

and this is obviously satisfied if, for each $y_2 \in \Theta(\bar{y}_1, \bar{y}_3)$, M_1 is isolatedly calm at $(\bar{y}_1, y_2)$ or M_2^{-1} is isolatedly calm at $(\bar{y}_3, y_2)$. $\square$

Finally, we apply Lemma 5.7 in order to find second-order sufficient optimality conditions for (P) where C is given as in (32).

Theorem 5.10. *Let $\bar{x} \in \mathbb{R}^n$ be a feasible point of (P) where C is given as in (32). Furthermore, let $\Phi: \mathbb{R}^m \rightrightarrows \mathbb{R}^\ell$ defined in (33) be inner semicompact and inner calm* at $F(\bar{x})$ w.r.t. C . Assume that for each $u \in \mathbb{R}^n \setminus \{0\}$ satisfying*

$$\nabla F(\bar{x})u \in \bigcup_{z \in \Phi(F(\bar{x}))} \{\nabla H(z)w \in \mathbb{R}^m \mid \nabla G(z)w \in T_D(G(z))\}, \quad \langle \nabla f_0(\bar{x}), u \rangle \leq 0,$$

there are $\alpha \geq 0$ and a multiplier $\lambda \in \Lambda^\alpha(\bar{x})$ such that, for each $z \in \Phi(F(\bar{x}))$ and $w \in \mathbb{R}^\ell$ satisfying $\nabla H(z)w = \nabla F(\bar{x})u$, there is a multiplier $\eta \in \Sigma(z, w, \lambda)$, where

$$\Sigma(z, w, \lambda) := \{\eta \in \widehat{N}_D^p(G(z); \nabla G(z)w) \mid \nabla G(z)^\top \eta = \nabla H(z)^\top \lambda\}$$

such that

$$\nabla_{xx}^2 L^\alpha(\bar{x}, \lambda)(u, u) + \vartheta(z, w, \lambda, \eta) > 0,$$

where ϑ has been defined in (34). Then $\bar{x}$ is an essential local minimizer of second order for (P).

Proof. For the proof, we are going to apply Theorem 5.2. Due to [4, Theorem 4.1], which is applicable since Φ is inner calm* at $F(\bar{x})$ w.r.t. C , we find

$$\mathcal{C}(\bar{x}) \subset \left\{ u \in \mathbb{R}^n \mid \begin{array}{l} \exists z \in \Phi(F(\bar{x})), \exists w \in \mathbb{R}^\ell: \\ \nabla F(\bar{x})u = \nabla H(z)w, \nabla G(z)w \in T_D(G(z)), \langle \nabla f_0(\bar{x}), u \rangle \leq 0 \end{array} \right\}.$$

Keeping Lemma 5.7 in mind, the assumptions of the theorem guarantee that for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, there are $\alpha \geq 0$ and a multiplier $\lambda \in \Lambda^\alpha(\bar{x})$ such that

$$\nabla_{xx}^2 L^\alpha(\bar{x}, \lambda)(u, u) + d^2 \delta_C(F(\bar{x}); \lambda)(\nabla F(\bar{x})u) > 0,$$

i.e., Theorem 5.2 shows that $\bar{x}$ is an essential local minimizer of second order. Note that the above inequality clearly holds if the infimum in Lemma 5.7 is attained, but it also holds if it is not, since in that case $d^2\delta_C(F(\bar{x}); \lambda)(\nabla F(\bar{x})u) = \infty$. $\square$

Note that incorporating the directional proximal normal cone $\widehat{N}_D^p(G(z); \nabla G(z)w)$ into the definition of the multiplier set $\Sigma(z, w, \lambda)$ in Theorem 5.10 is not restrictive. For fixed $z \in \Phi(F(\bar{x}))$, $w \in \mathbb{R}^\ell$, and $\eta \in \mathbb{R}^p$ such that $\nabla H(z)w = \nabla F(\bar{x})u$, $\lambda \in \Lambda^\alpha(\bar{x})$, and $\nabla G(z)^\top \eta = \nabla H(z)^\top \lambda$, the implicitly postulated lower estimate $d^2\delta_D(G(z); \eta)(\nabla G(z)w) > -\infty$ already implies $\eta \in \widehat{N}_D^p(G(z); \nabla G(z)w)$, see Proposition 3.6. By definition of the proximal pre-normal cone, this gives $\langle \eta, \nabla G(z)w \rangle \leq 0$. On the other hand, by choice of u , we also have

$$\langle \eta, \nabla G(z)w \rangle = \langle \nabla H(z)^\top \lambda, w \rangle = \langle \lambda, \nabla F(\bar{x})u \rangle = \langle -\nabla f_0(\bar{x}), u \rangle \geq 0$$

giving $\eta \in \{\nabla G(z)w\}^\perp$, i.e., $\eta \in \widehat{N}_D^p(G(z); \nabla G(z)w)$.

6. Second-order sufficient conditions in composite optimization

Let us investigate the composite minimization problem

$$(CP) \quad \min\{f_0(x) + g(F(x)) \mid x \in \mathbb{R}^n\}$$

where $f_0: \mathbb{R}^n \rightarrow \mathbb{R}$ and $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ are twice continuously differentiable and $g: \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ is proper and lower semicontinuous. In comparison with (P), this is a more general optimization problem, since (P) results from (CP) by setting $g := \delta_C$. We make use of the function $L: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ given via

$$L(x, \lambda) := f_0(x) + \langle \lambda, F(x) \rangle \quad \forall x \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}^m$$

and note that this notation is consistent with (23) since we have $L = L^1$. For each $\bar{x} \in \mathbb{R}^n$ such that $|g(F(\bar{x}))| < \infty$, it is, thus, also reasonable to work with the multiplier set $\Lambda(\bar{x}) := \Lambda^1(\bar{x})$, see (25), and we define the so-called critical cone given by means of

$$\mathcal{C}(\bar{x}) := \{u \in \mathbb{R}^n \mid \langle \nabla f_0(\bar{x}), u \rangle + d g(F(\bar{x}))(\nabla F(\bar{x})u) \leq 0\}.$$

Again, this is compatible with the definition of the critical cone from (24) in the setting of constrained optimization, see Example 3.9. Observe that

$$d(f_0 + g \circ F)(\bar{x})(u) \geq \langle \nabla f_0(\bar{x}), u \rangle + d g(F(\bar{x}))(\nabla F(\bar{x})u) \quad \forall u \in \mathbb{R}^n \quad (36)$$

by [38, Theorem 10.6, Corollary 10.9], see Corollary 4.3 as well.

Based on Proposition 2.11 and our chain rule from Corollary 4.3, we are in position to state second-order sufficient optimality conditions for (CP) very easily.

Theorem 6.1. *Let $\bar{x} \in \mathbb{R}^n$ be a point such that $|g(F(\bar{x}))| < \infty$. Furthermore, assume that, for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, there exists a multiplier $\lambda \in \Lambda(\bar{x})$ such that*

$$\nabla_{xx}^2 L(\bar{x}, \lambda)(u, u) + d^2 g(F(\bar{x}); \lambda)(\nabla F(\bar{x})u) > 0. \quad (37)$$

Then the second-order growth condition holds for $f_0 + g \circ F$ at $\bar{x}$. Particularly, $\bar{x}$ is a strict local minimizer of (CP).

Proof. We will show that $d^2(f_0 + g \circ F)(\bar{x}; 0)(u) > 0$ for each $u \in \mathbb{R}^n \setminus \{0\}$ in order to distill the result from Proposition 2.11. Thus, fix $u \in \mathbb{R}^n \setminus \{0\}$. If $u \notin \mathcal{C}(\bar{x})$, we obtain

$$d(f_0 + g \circ F)(\bar{x})(u) \geq \langle \nabla f_0(\bar{x}), u \rangle + d g(F(\bar{x}))(\nabla F(\bar{x})u) > 0$$

from (36), and (5) yields $d^2(f_0 + g \circ F)(\bar{x}; 0)(u) = \infty$. For $u \in \mathcal{C}(\bar{x})$, the assumptions of the theorem guarantee the existence of a multiplier $\lambda \in \Lambda(\bar{x})$ satisfying

$$\begin{aligned} d^2(f_0 + g \circ F)(\bar{x}; 0)(u) &= \nabla^2 f_0(\bar{x})(u, u) + d^2(g \circ F)(\bar{x}, -\nabla f_0(\bar{x}))(u) \\ &\geq \nabla_{xx}^2 L(\bar{x}, \lambda)(u, u) + d^2 g(F(\bar{x}); \lambda)(\nabla F(\bar{x})u) > 0 \end{aligned} \quad (38)$$

by Proposition 2.8, Corollary 4.3, and the definitions of L and $\Lambda(\bar{x})$. $\square$

As already seen in Section 5, there is additional information hidden in (37). Similarly as in Remark 5.3 and exploiting (15), one can show that, for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, the multiplier $\lambda \in \Lambda(\bar{x})$ which satisfies (37) is an element of $\hat{\partial}^p g(F(\bar{x}); \nabla F(\bar{x})u)$.

Let us mention that under some additional assumptions and in a more specific setting, second-order sufficient optimality conditions for composite optimization problems have been shown in [39, Proposition 2.5] and [38, Exercise 13.26].

When comparing Theorem 5.2 and Theorem 6.1, the natural question arises whether the assertion of Theorem 6.1 remains true when using the more general Lagrangian L^α and the more general multiplier set $\Lambda^\alpha(\bar{x})$ for some $\alpha \geq 0$, see (23) and (25). The following examples illustrates that this is indeed not the case.

Example 6.2. We consider (CP) with the data functions given by

$$f_0(x) := -\frac{1}{2}x^2, \quad F(x) := \frac{1}{2}x^2, \quad g(x) := x \quad \forall x \in \mathbb{R}$$

and fix the point $\bar{x} := 0$ which, obviously, is a (global) minimizer of $f_0 + g \circ F$ where the second-order growth condition fails. However, for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\} = \mathbb{R} \setminus \{0\}$, we find $1 \in \Lambda^0(\bar{x})$, $\nabla_{xx}^2 L^0(\bar{x}, 1)(u, u) = u^2 > 0$ and

$$d^2 g(F(\bar{x}); 1)(\nabla F(\bar{x})u) = d^2 g(0; 1)(0) = 0$$

which gives the estimate

$$\nabla_{xx}^2 L^0(\bar{x}, 1)(u, u) + d^2 g(F(\bar{x}); 1)(\nabla F(\bar{x})u) > 0.$$

Thus, a potential generalization of the second-order condition from Theorem 5.2 holds, but the second-order growth condition fails. $\square$

The following example visualizes the result of Theorem 6.1 in terms of so-called sparse optimization.

Example 6.3. We consider the problem (CP) with $g(y) := \|y\|_0$, $y \in \mathbb{R}^m$, where $\|\cdot\|_0: \mathbb{R}^m \rightarrow \mathbb{R}$ is the so-called ℓ_0 -quasi-norm which counts the non-zero entries of the argument vector. Thus, in the associated program (CP), those points $\bar{x} \in \mathbb{R}^n$ are preferred that come along with many zero entries in $F(\bar{x})$. This is of particular interest whenever the potential constraint system $F(x) = 0$ of equations does not possess a solution.

Introducing $\phi: \mathbb{R} \rightarrow \mathbb{R}$ by means of

$$\phi(t) := \begin{cases} 0 & t = 0, \\ 1 & t \neq 0 \end{cases} \quad \forall t \in \mathbb{R},$$

we have $\|y\|_0 = \sum_{i=1}^m \phi(y_i)$ for each $y \in \mathbb{R}^m$, i.e., the nonsmoothness of $\|\cdot\|_0$ is separable and we can exploit the sum rule from Proposition 2.8 (b) in order to compute the (first and) second subderivative of $\|\cdot\|_0$. Some easy calculations show

$$d\phi(t)(r) = \begin{cases} 0 & t \neq 0 \text{ or } t = r = 0, \\ \infty & t = 0, r \neq 0, \end{cases} \quad \forall t, r \in \mathbb{R}$$

and

$$d^2\phi(t; t^*)(r) = \begin{cases} \infty & t^*r < 0 \text{ or } t = 0, r \neq 0, \\ -\infty & t^*r > 0 \text{ and } t \neq 0, \\ 0 & \text{otherwise} \end{cases} \quad \forall t, t^*, r \in \mathbb{R}.$$

This can be used to show that

$$\mathcal{C}(\bar{x}) = \{u \in \mathbb{R}^n \mid \langle \nabla f_0(\bar{x}), u \rangle \leq 0, \langle \nabla F_i(\bar{x}), u \rangle = 0 \forall i \in I_0(\bar{x})\}$$

where, for some point $\bar{x} \in \mathbb{R}^n$, we use $I_0(\bar{x}) := \{i \in \{1, \dots, m\} \mid F_i(\bar{x}) = 0\}$ and $F_1, \dots, F_m: \mathbb{R}^n \rightarrow \mathbb{R}$ are the component functions of F .

Thus, whenever for each $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, there is a $\lambda \in \Lambda(\bar{x})$ such that $\lambda_i \langle \nabla F_i(\bar{x}), u \rangle = 0$ for all $i \notin I_0(\bar{x})$ and $\nabla_{xx}^2 L(\bar{x}, \lambda)(u, u) > 0$, then the second-order growth condition holds for $f_0 + \|\cdot\|_0 \circ F$ at $\bar{x}$. Note that the above is obviously less restrictive than a standard second-order sufficient optimality condition for the equality-constrained optimization problem

$$\min\{f_0(x) \mid F_i(x) = 0 \forall i \in I_0(\bar{x})\}$$

since the multiplier λ does not necessarily need to vanish on $\{1, \dots, m\} \setminus I_0(\bar{x})$. $\square$

7. Concluding remarks

In this paper, we derived calculus rules for the second subderivative of lower semicontinuous functions, comprising a composition rule, a marginal function rule, an image rule, and a pre-image rule. Our findings throw some new light on the results from [38, Section 13]. Moreover, we worked out the precise role of the comparatively new inner calmness* property from [4] in the context of the marginal function and image rule. We introduced the directional proximal subdifferential of a given lower semicontinuous function which captures the finiteness of the second subderivative. Based on the derived results, we were in position to easily obtain second-order sufficient optimality conditions in constrained and composite optimization which are given in terms of initial problem data. Exemplary, this has been illustrated in terms of disjunctive and nonlinear second-order cone optimization. Let us point out that our findings are applicable to inherently difficult problem classes such as optimization problems with (quasi-) variational inequality constraints (like abstract

complementarity constraints induced by non-polyhedral convex cones) or bilevel optimization problems as well. However, the calculations which are necessary to estimate the appearing curvature term from below still are a slightly laborious task which is why we abstained from presenting them here but leave them as a promising topic of future research. Keeping in mind that second-order sufficient optimality conditions guarantee local fast convergence of diverse numerical solution methods like Newton-type or multiplier-penalty-algorithms, it should be studied whether our new second-order conditions can be employed beneficially in this area. Some first steps in this direction have been done recently in [25, 26, 39], where the authors investigate augmented Lagrangian and sequential-quadratic-programming methods for nonlinear second-order cone programs and composite optimization problems with piecewise linear-quadratic nonsmooth terms. Finally, let us recall that our approach may not be suitable in order to obtain applicable second-order necessary optimality conditions.

Acknowledgments. We would like to thank the referees for their valuable comments which helped to improve the quality of this paper.

Funding. The research of Matúš Benko was supported by the Austrian Science Fund (FWF) under grant P32832-N as well as by the infrastructure of the Institute of Computational Mathematics, Johannes Kepler University Linz, Austria.

References

- [1] J.-P. Aubin, H. Frankowska: *Set-Valued Analysis*, Birkhäuser, Boston, 2009.
- [2] K. Bai, J. J. Ye: *Directional necessary optimality conditions for bilevel programs*, Math. Oper. Res. 47/2 (2022) 1169–1191.
- [3] A. Ben-Tal: *Second-order and related extremality conditions in nonlinear programming*, J. Optim. Theory Appl. 31/2 (1980) 143–165.
- [4] M. Benko: *On inner calmness*, generalized calculus, and derivatives of the normal cone mapping*, J. Nonsmooth Anal. Optim. 2 (2021), art. no. 5881.
- [5] M. Benko, M. Červinka, T. Hoheisel: *Sufficient conditions for metric subregularity of constraint systems with applications to disjunctive and ortho-disjunctive programs*, Set-Valued Var. Analysis 30/1 (2022) 143–177.
- [6] M. Benko, H. Gfrerer, J. V. Outrata: *Calculus for directional limiting normal cones and subdifferentials*, Set-Valued Var. Analysis 27/3 (2019) 713–745.
- [7] M. Benko, H. Gfrerer, J. J. Ye, J. Zhang, J. Zhou: *Second-order optimality conditions for general nonconvex optimization problems and variational analysis of disjunctive systems*, arXiv: 2203.10015 (2022).
- [8] M. Benko, P. Mehlitz: *On implicit variables in optimization theory*, J. Nonsmooth Anal. Optim. 2 (2021), art. no. 7215.
- [9] M. Benko, P. Mehlitz: *Calmness and calculus: two basic patterns*, Set-Valued Var. Analysis 30 (2022) 81–117.
- [10] J. F. Bonnans, R. Cominetti, A. Shapiro: *Second order optimality conditions based on parabolic second order tangent sets*, SIAM J. Optimization 9/2 (1999) 466–492.
- [11] J. F. Bonnans, C. H. Ramírez: *Perturbation analysis of second-order cone programming problems*, Math. Programming 104 (2005) 205–227.

- [12] J. F. Bonnans, A. Shapiro: *Perturbation Analysis of Optimization Problems*, Springer, New York (2000).
- [13] C. Christof, G. Wachsmuth: *No-gap second-order conditions via a directional curvature functional*, SIAM J. Optimization 28/3 (2018) 2097–2130.
- [14] F. H. Clarke: *Optimization and Nonsmooth Analysis*, Wiley Interscience, New York (1983).
- [15] S. Dempe: *Foundations of Bilevel Programming*, Kluwer Academic, Dordrecht (2002).
- [16] F. Facchinei, J.-S. Pang: *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York (2003).
- [17] M. L. Flegel, C. Kanzow, J. V. Outrata: *Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints*, Set-Valued Analysis 15/2 (2007) 139–162.
- [18] H. Gfrerer: *On directional metric regularity, subregularity and optimality conditions for nonsmooth mathematical programs*, Set-Valued Var. Analysis 21/2 (2013) 151–176.
- [19] H. Gfrerer: *Optimality conditions for disjunctive programs based on generalized differentiation with application to mathematical programs with equilibrium constraints*, SIAM J. Optimization 24/2 (2014) 898–931.
- [20] H. Gfrerer, J. V. Outrata: *On computation of generalized derivatives of the normal-cone mapping and their applications*, Math. Oper. Res. 41/4 (2016) 1535–1556.
- [21] H. Gfrerer, J. J. Ye, J. Zhou: *Second-order optimality conditions for nonconvex set-constrained optimization problems*, Math. Oper. Res. 47/3 (2022) 2344–2365.
- [22] I. Ginchev, B. S. Mordukhovich: *On directionally dependent subdifferentials*, Proc. Bulgarian Acad. Sci. 64/4 (2011) 497–508.
- [23] L. Guo, G.-H. Lin, J. J. Ye, J. Zhang: *Sensitivity analysis of the value function for parametric mathematical programs with equilibrium constraints*, SIAM J. Optimization 24/3 (2014) 1206–1237.
- [24] N. T. V. Hang, B. S. Mordukhovich, M. E. Sarabi: *Second-order variational analysis in second-order cone programming*, Math. Programming 180/1 (2020) 75–116.
- [25] N. T. V. Hang, B. S. Mordukhovich, M. E. Sarabi: *Augmented Lagrangian method for second-order cone programs under second-order sufficiency*, J. Global Optim. 82 (2022) 51–81.
- [26] N. T. V. Hang, M. E. Sarabi: *Local convergence analysis of augmented Lagrangian methods for piecewise linear-quadratic composite optimization problems*, SIAM J. Optimization 31/4 (2021) 2665–2694.
- [27] Z.-Q. Luo, J.-S. Pang, D. Ralph: *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge (1996).
- [28] G. McCormick: *Second order conditions for constrained minima*, SIAM J. Appl. Math. 15/3 (1967) 641–652.
- [29] P. Mehlitz: *On the linear independence constraint qualification in disjunctive programming*, Optimization 69/10 (2020) 2241–2277.
- [30] A. Mohammadi, B. S. Mordukhovich, M. E. Sarabi: *Parabolic regularity in geometric variational analysis*, Trans. Amer. Math. Soc. 374 (2021) 1711–1763.

- [31] J. V. Outrata, M. Kočvara, J. Zowe: *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Kluwer Academic, Dordrecht (1998).
- [32] J. Penot: *Second-order conditions for optimization problems with constraints*, SIAM J. Control Optim. 37/1 (1998) 303–318.
- [33] R. A. Poliquin, R. T. Rockafellar: *Tilt stability of a local minimum*, SIAM J. Optimization 8/2 (1998) 287–299.
- [34] S. M. Robinson: *Some continuity properties of polyhedral multifunctions*, in: *Mathematical Programming at Oberwolfach*, H. König et al. (eds.), Springer, Berlin (1981) 206–214.
- [35] R. T. Rockafellar: *Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives*, Math. Oper. Res. 14/3 (1989) 462–484
- [36] R. T. Rockafellar: *Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality*, Math. Programming (2022).
- [37] R. T. Rockafellar: *Convergence of augmented Lagrangian methods in extensions beyond nonlinear programming*, Math. Programming (2022).
- [38] R. T. Rockafellar, R. J.-B. Wets: *Variational Analysis*, Springer, Berlin (1998).
- [39] M. E. Sarabi: *Primal superlinear convergence of SQP methods in piecewise linear-quadratic composite optimization*, Set-Valued Var. Analysis 30 (2022) 1–37.
- [40] V. D. Thinh, T. D. Chuong, N. L. H. Anh: *Second order variational analysis of disjunctive constraint sets and its applications to optimization problems*, Optimization Letters 15 (2021) 2201–2224.
- [41] D. Wachsmuth, G. Wachsmuth: *Second-order conditions for non-uniformly convex integrands: quadratic growth in L^1* , J. Nonsmooth Anal. Optim. 3 (2022), art. no. 8733.
- [42] D. W. Walkup, R. J.-B. Wets: *A Lipschitz characterization of convex polyhedra*, Proc. Amer. Math. Soc. 23/1 (1969) 167–173.