Mathematical Programming 198 (2023), 159–194

Augmented Lagrangians and Hidden Convexity in Sufficient Conditions for Local Optimality

R. Tyrrell Rockafellar[1]

Abstract

Second-order sufficient conditions for local optimality have long been central to designing solution algorithms and justifying claims about their convergence. Here a far-reaching extension of such conditions, called variational sufficiency, is explored in territory beyond just classical nonlinear programming. Variational sufficiency is already known to support multiplier methods that are able, even without convexity, to achieve problem decomposition, but further insight has been needed into how it coordinates with other sufficient conditions. In the framework of this paper, it is shown to characterize local optimality in terms of having a convex-concave-type local saddle point of an augmented Lagrangian function. A stronger version of variational sufficiency is tied in turn to local strong convexity in the primal argument of that function and a property of augmented tilt stability that offers crucial aid to Lagrange multiplier methods at a fundamental level of analysis. Moreover, that strong version is translated here through second-order variational analysis into statements that can readily be compared to existing sufficient conditions in nonlinear programming, second-order cone programming, and other problem formulations which can incorporate nonsmooth objectives and regularization terms.

Keywords: *sufficient conditions for local optimality, generalized nonlinear programming, second-order cone programming, nonsmooth optimization, generalized augmented Lagrangians, ALM, local duality, variational convexity, local maximal monotonicity, tilt stability, metric regularity, second-order variational analysis*

(finalized version after acceptance for publication)

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@uw.edu*, URL: sites.math.washington.edu/∼rtr/mypage.html

1 Introduction

The classical sufficient condition for local optimality in the unconstrained minimization of a $\mathcal{C}^2$ function on $I\!R^n$ combines the vanishing of the gradient with the positive-definiteness of the Hessian matrix. That positive-definiteness makes the function be strongly convex around the minimizing point and effectively reduces the problem locally to one of convex optimization. In minimizing a function subject to constraints, there seems no hope of a local reduction to convex optimization, because the feasible set is usually not convex. But this turns out to be a misconception. As will be shown here, hidden convexity can emerge when localization is taken not only in the primal variables but also in the Lagrange multipliers as dual variables. Moreover the typical second-order sufficient conditions for local optimality in nonlinear programming and its extensions are truly anchored in that.

This is a new observation, but there have been some hints in the past in the theory of augmented Lagrangians, where the so-called strong second-order sufficient condition in nonlinear programming leads to a kind of local duality that explains the workings of the multiplier method of Hestenes and Powell and its decendents. In this duality, the primal-dual pair in the first-order condition is a local *saddle point* of the augmented Lagrangian — for a high enough level of augmentation. All that was pointed out by Bertsekas in his 1982 book [2] and other works. But actually, the augmented Lagrangian in this case is *locally convex-concave* in the primal and dual arguments, so the saddle point corresponds to solving a localized primal-dual pair of optimization problems in the duality framework of convex analysis. Again, there is effectively a local reduction to convex optimization, but it needs to be "elicited" through augmentation.

Our goal is to develop this picture of sufficient conditions versus local duality in a much broader setting. The generalized nonlinear programming problem we take up is to

$$\begin{array}{l}\text{minimize } \; f_0(x) + g(F(x)) \; \text{ for } \; F(x) = (f_1(x), \ldots, f_m(x)), \; \text{ where} \\ f_0, f_1, \ldots, f_m \text{ are } \mathcal{C}^1 \text{ on } I\!R^n \text{ and } g \text{ is closed proper convex on } I\!R^m.\end{array} \tag{1.1}$$

In particular, g could be the indicator δ_K of a closed convex set K of vectors $u = (u_1, \ldots, u_m) \in I\!R^m$, and then the g term in the objective would stand for the constraint $F(x) \in K$. Classical nonlinear programming with equality and inequality constraints would correspond to

$$g = \delta_K \;\text{ for }\; K = \{\, u \mid u_i \leq 0 \text{ for } 1 \leq i \leq s, \text{ but } u_i = 0 \text{ for } s+1 \leq i \leq m \,\}. \tag{1.2}$$

Taking K instead to be the Lorenz cone, would yield second-order cone programming in (1.1). Semidefinite programming comes up when the vectors x stand for matrices and K is the cone of symmetric, positive-semidefinite matrices. However, other choices of g, which we call the *modeling function* in (1.1), cover problems even with nonsmooth objectives. For instance, g could be a norm like $||\cdot||_1$ or $||\cdot||_\infty$ with a role in "regularizing" solutions. On the other hand, nonsmoothness is paramount when

$$g(F(x)) = \max\{f_1(x), \ldots, f_m(x)\} \quad \text{in the case of} \quad g(u) = \text{vecmax}(u) = \max\{u_1, \ldots, u_m\}. \tag{1.3}$$

And of course, there can be a mixture of the various possibilities: the vector u could be partitioned into subvectors u^j with a different modeling function g^j attached to each, and then $g(u)$ would be the sum of the expressions $g^j(u^j)$, some capturing constraints and others representing composite terms with possible nonsmoothness or regularization effects.

It may seem odd that, for work on sufficient conditions, we have taken the functions f_i in (1.1) only to be $\mathcal{C}^1$ instead of the usual $\mathcal{C}^2$. That is because we fundamentally rely on convexity-type properties, hidden or explicit, and these can fully be articulated without assistance from assumptions about second derivatives. This definitely distinguishes our approach from that of most others in the subject.

However, after our main results have been established, the $\mathcal{C}^2$ case of (1.1), where the functions f_i are $\mathcal{C}^2$, will nonetheless be scrutinized to see what more can be said in terms of that structure in connecting up with previously known conditions for local optimality in special cases of problem (1.1). But even in the $\mathcal{C}^2$ case, the augmented Lagrangian functions that come into play will generally not be twice differentiable but only $\mathcal{C}^{1+}$, i.e., continuously differentiable with gradient mappings that are locally Lipschitz continous.

Although Lagrange multipliers are believed by many to be an accompaniment just of constraints, they can be attached also to the composite expression in (1.1), as has long been known, cf. [19]. The multipliers are induced by incorporating a perturbation parameter, and we do that by recasting (1.1) in the form

$$(P) \qquad \text{minimize } \varphi(x,u) \text{ subject to } u=0, \text{ where } \varphi(x,u)=f_0(x)+g(F(x)+u).$$

Problems of minimizing $\varphi(x,u)$ in x for different choices of u than $u=0$ are viewed as perturbations of problem (1.1). The Lagrangian function corresponding to the perturbation scheme in (P) is

$$\begin{aligned} &l(x,y)=\inf_u\{\,\varphi(x,u)-y{\cdot}u\}=L(x,y)-g^*(y) \text{ in the notation} \\ &L(x,y)=f_0(x)+y{\cdot}F(x)=f_0(x)+y_1f_1(x)+\cdots+y_mf_m(x), \end{aligned} \tag{1.4}$$

where g^* is the convex function conjugate to g. If $g=\delta_K$ for a set K, then g^* is the positively homogeneous support function of K. If g itself is positively homogeneous, then g^* is the indicator of a set Y. For instance, if g is a norm, g^* is the indicator of the unit ball of the dual norm. For the vecmax function g in (1.3), g^* is δ_Y for the unit simplex, $Y=\{\,y=(y_1,\ldots,y_m)\,|\,y_i\geq 0,\ y_1+\cdots+y_m=1\}$. If g is the indicator of a *cone* K, then g^* is the indicator of the polar cone $Y=K^*$. The cone K in (1.2) has Y being the familiar space of multiplier vectors for the classical constraint system in nonlinear programming, namely $Y=\mathbb{R}^s_+\times\mathbb{R}^{m-s}$.

Also associated with problem (P) is the *augmented* Lagrangian with parameter r, defined by

$$l_r(x,y)=\inf_u\Big\{\,\varphi(x,u)-y{\cdot}u+\frac{r}{2}|u|^2\Big\} \text{ for } r>0, \text{ where } |u|=||u||_2. \tag{1.5}$$

The minimization here can be carried out in different ways. With the help of the convex functions

$$\begin{aligned} &g^r(u)=\min_{u'}\Big\{\,g(u')+\tfrac{r}{2}|u'-u|^2\Big\} \text{ with conjugate } g^{r\,*}(y)=g^*(y)+\tfrac{1}{2r}|y|^2, \\ &g_r(u)\,=g(u)+\tfrac{r}{2}|u|^2 \text{ with conjugate } g_r^*(y)\,=\,\min_{y'}\Big\{\,g^*(y')+\tfrac{1}{2r}|y'-y|^2\Big\}, \end{aligned} \tag{1.6}$$

one gets

$$\begin{aligned} &l_r(x,y)=f_0(x)+g^r(F(x)+\tfrac{1}{r}y)-\tfrac{1}{2r}|y|^2, \text{ or} \\ &l_r(x,y)=L(x,y)+\tfrac{r}{2}|F(x)|^2-g_r^*(y+rF(x)). \end{aligned} \tag{1.7}$$

In the cone case with $g=\delta_K$ and $g^*=\delta_Y$, the functions in these expressions come down to $g^r(u)=\frac{r}{2}d_K^2(u)$ and $g_r^*(y)=\frac{1}{2r}d_Y^2(y)$, where d_K and d_Y are the distance functions for K and Y.

An important feature in (1.7) is that the functions g^r and g_r^*, as prox-regularizations of g and g^*, are differentiable in y with gradient mappings that are Lipschitz continous *globally*; the modulus for ∇g^r is r, while for ∇g_r^* it is r^{-1}. In fact, because the formula for g^r in (1.6) yields $\nabla g^r(u)=r(u-u')$ for the unique u' giving the minimum, which itself is characterized by $0\in\partial g(u')+r(u'-u)$, one has

$$\begin{aligned} &\nabla g^r=I-(I+r^{-1}\partial g)^{-1}=r(I-\mathrm{prox}_{r^{-1}g}), \text{ where} \\ &\mathrm{prox}_{r^{-1}g}(u)=\mathrm{argmin}_{u'}\Big\{\,r^{-1}g(u')+\tfrac{1}{2}|u'-u|^2\Big\}. \end{aligned} \tag{1.8}$$

The prox mappings associated with closed proper convex functions have a valuable characteristic called firm expansiveness, and they are seen increasingly nowadays in numerical schemes in data science. If $g = \delta_K$, then $\mathrm{prox}_{r^{-1}g}$ is P_K, the projection mapping onto K.

Observe that the augmented Lagrangian l_r, as generated in (1.5) from φ, can equally be regarded as the ordinary Lagrangian generated as in (1.4) from the *augmented objective* function

$$\varphi_r(x,u) = \varphi(x,u) + \frac{r}{2}|u|^2, \tag{1.9}$$

which has the extra property of being *strongly* convex in u. Since (1.4) and (1.5) make $-l(x,\cdot)$ and $-l_r(x,\cdot)$ the convex functions conjugate to $\varphi(x,\cdot)$ and $\varphi_r(x,\cdot)$, we also have the reciprocal formulas

$$\varphi(x,u) = \sup_y \{\, l(x,y) + y\cdot u\}, \qquad \varphi_r(x,u) \;=\; \sup_y \{\, l_r(x,y) + y\cdot u\}. \tag{1.10}$$

In replacing φ in (P) by φ_r to get an augmented problem (P_r), the minimization still comes out to be that in (1.1). Only the scheme of perturbations changes, and it changes very little: in minimizing $\varphi_r(x,u)$ in x for some $u \neq 0$, the objective is the same as in mimizing $\varphi(x,u)$ in x except for the addition of a positive constant. But the effect on the properties of Lagrange multipliers is profound.

Both $l(x,y)$ and $l_r(x,y)$ are always concave in y, but $l_r(x,y)$ is also differentiable in y, and $\nabla_y l_r(x,y)$ is Lipschitz continuous in y with modulus r^{-1}. That's a consequence of $-l_r(x,\cdot)$ being conjugate to $\varphi_r(x,\cdot)$, which is r-strongly convex in u, but it can also be seen in connection with the properies of g^r and g_r^* in (1.7) that have been mentioned earlier, under which

$$\text{the augmented Lagrangian } l_r(x,y) \text{ is a } \mathcal{C}^1 \text{ function on } I\!R^n \times I\!R^n. \tag{1.11}$$

It would not be $\mathcal{C}^2$ under the stronger assumption that the functions f_i are $\mathcal{C}^2$, because g^r and g_r^* are generally not $\mathcal{C}^2$, only $\mathcal{C}^{1+}$. However, that lesser property does imply that

$$\text{the augmented Lagrangian } l_r(x,y) \text{ is } \mathcal{C}^{1+} \text{ when every } f_i \text{ is } \mathcal{C}^2. \tag{1.12}$$

Characterizations of optimality in our problem (P) begin with first-order conditions expressed by subgradients of the function φ in the sense of variational analysis [23]. The description of the subgradients is simplified here by the fact that our assumptions make φ be everywhere an *amenable* function (*fully* amenable if every $f_i \in \mathcal{C}^2$) [23, 10.23]. For an amenable function f, all subgradients are *regular* subgradients, i.e.,

$$z \in \partial f(w) \quad\Longleftrightarrow\quad f(w') \geq f(w) + z\cdot(w'-w) + o(w'-w). \tag{1.13}$$

Moreover f is *subdifferentially continuous*, meaning that $f(w') \to f(w)$ as $(w',z') \to (w,z)$ in the graph of the subgradient mapping ∂f, that being the closed set $\mathrm{gph}\,\partial f = \{\,(w,z)\,|\, z \in \partial f(w)\}$. All this accrues from the definition of amenability by way of the basic chain rule in [23, 10.6].

That same chain rule, when specifically applied to the structure of φ in (P), yields:

$$\begin{aligned}(v,y) \in \partial\varphi(x,u) \quad&\Longleftrightarrow\quad y \in \partial g(F(x)+u),\;\; v = \nabla_x L(x,y)\\ &\Longleftrightarrow\quad v \in \partial_x l(x,y),\;\; u \in \partial_y[-l](x,y),\end{aligned} \tag{1.14}$$

where the fact is used that $y \in \partial g(u) \Leftrightarrow u \in \partial g^*(y)$. Likewise,

$$\begin{gathered}(v,y) \in \partial\varphi_r(x,u) \quad\Longleftrightarrow\quad v = \nabla_x l_r(x,y),\;\; u = -\nabla_y l_r(x,y) \quad\Longleftrightarrow\\ v = \nabla_x L(x,\eta) \;\text{ and }\; u = r^{-1}(\eta - y) \;\text{ for }\; \eta = \nabla g^r(F(x) + r^{-1}y).\end{gathered} \tag{1.15}$$

Furthermore, in (1.14) and (1.15) the conditions on y correspond to attainment in the minimization formulas (1.5) and (1.7):

$$\begin{array}{lcl} u \in \partial_y[-l](x,y) & \iff & l(x,y) = \varphi(x,u) - y{\cdot}u, \\ u = -\nabla_y l_r(x,y) & \iff & l_r(x,y) = \varphi_r(x,u) - y{\cdot}u. \end{array} \tag{1.16}$$

First-order optimality through subgradients. *The first-order condition for the local optimality of $\bar{x}$ in (P), which is necessary under various constraint qualifications, is the existence of $\bar{y}$ such that*

$$(0,\bar{y}) \in \partial\varphi(\bar{x},0). \tag{1.17}$$

This can be expressed equivalently through (1.17) as

$$0 \in \partial_x l(\bar{x},\bar{y}) \text{ and } 0 \in \partial_y[-l](\bar{x},\bar{y}), \quad \text{or } \nabla_x L(\bar{x},\bar{y}) = 0 \text{ with } \bar{y} \in \partial g(F(\bar{x})), \tag{1.18}$$

or for that matter as $(0,\bar{y}) \in \partial\varphi_r(\bar{x},0)$, and then via (1.16) as

$$\nabla_x l_r(\bar{x},\bar{y}) = 0 \text{ and } \nabla_y l_r(\bar{x},\bar{y}) = 0, \text{ or } \nabla_x L(\bar{x},\bar{y}) = 0 \text{ with } \nabla g^r(F(\bar{x}) + r^{-1}\bar{y}) = \bar{y}. \tag{1.19}$$

A particular constraint qualification under which the first-order condition is necessary is

$$y \in N_{\operatorname{dom} g}(F(\bar{x})), \;\; \nabla F(\bar{x})^* y = 0 \implies y = 0, \tag{1.20}$$

where $\nabla F(x)$ denotes the Jacobian of F, $\nabla F(x)^*$ is its transpose, and $N_{\operatorname{dom} g}$ is the normal cone mapping associated with the convex effective domain of g, [23, 10.12].

In classical nonlinear programming with $g = \delta_K$ for the cone K in (1.2), the subgradient condition $\bar{y} \in \partial g(F(\bar{x}))$ in the first-order condition (1.18) specializes to the normal cone condition $\bar{y} \in N_K(F(\bar{x}))$. That expresses complementary slackness and turns (1.18) into the familiar Karush-Kuhn-Tucker system of relationships which a numerical method might try to solve. The equivalent statement in (1.19) translates that task into something perhaps more agreeable: determining a zero of a gradient mapping ∇l_r. Beyond that classical specialization, (1.19) likewise translates (1.18) into something more advantageous. From that viewpoint, augmentation can be thought of as a kind of *problem regularization* toward improved computations. Indeed, the quadratic term involved the passage from $\varphi(x,u)$ to $\varphi_r(x,u)$ in (1.9) can directly be interpreted as a regularization.

In the *convex case* of (P), where $\varphi(x,u)$ is convex as a function of (x,u), not just u, the first-order condition (1.17) is sufficient for *global* optimality in (P). The convexity of $\varphi(x,u)$ on $I\!R^n \times I\!R^n$ corresponds to the convexity of the Lagrangian $l(x,y)$ as a function of $x \in I\!R^n$ for each $y \in I\!R^m$, cf. [23, 11.48]. We already know that $l(x,y)$ is concave in y, so in that case the Lagrangian version (1.18) of the first-order optimality condition says equivalently that $(\bar{x},\bar{y})$ is a *global saddle point* of $l(x,y)$ with respect to minimizing in x and maximizing in y. That is the setting also for global duality theory, where $\bar{x}$ solves (P) and $\bar{y}$ solves an associated maximization problem (D), with the minimum value of the objective in (P) equaling the maximum value of the objective in (D).

For understanding where we are headed in this paper, it will be instructive now to observe, as a conceptual bridge, that the first-order condition is also sufficient for global optimality even if φ itself isn't convex, as long as φ_r is convex for r sufficiently high. Then it corresponds to a global saddle point of the augmented Lagrangian l_r instead of l. This can be thought of as *eliciting* global convexity and its benefits through the parameter r. The convexity of φ, or the elicitation of convexity of φ_r, can be interpreted as a *second-order* condition which, when combined with the first-order condition, is sufficient for global optimality. *The key idea here is to replace global convexity of φ_r by a local property,* *called* variational convexity [20]*, which will be shown to make the first-order condition correspond to* $(\bar{x}, \bar{y})$ *being a* local *saddle point of* convex-concave *type of the augmented Lagrangian* $l_r(x, y)$*.*

More will be explained about variational convexity shortly. The essential point, for putting our efforts into perspective, is its equivalence in the case of φ_r with local maximal monotonicity of the subgradient mapping $\partial\varphi_r$ relative to $(\bar{x}, 0)$ and $(0, \bar{y})$, which carries over also to the mapping $(x, y) \mapsto \{\, (v, u) \,|\, (v, -u) \in \partial l_r(x, y)\}$. That's important because local maximal monotonicity validates local actions of the proximal point algorithm [16], [17], which is the engine that runs augmented Lagrange multiplier methods of all sorts, as well as the progressive decompling algorithm of [21].

Our most fundamental results will be established already in Section 2. Follow-up will come in Sections 3 and 4 in showing how, in the $\mathcal{C}^2$ case of (P), the stronger version of those results can be coordinated with properties of second derivatives of various generalized kinds. That effort will also clarify relationships with other known sufficient conditions for local optimality in instances of (P).

2 Variational sufficiency and local duality

Variational convexity of a lower semicontinuous (lsc) function $f : I\!R^n \to (-\infty, \infty]$ is a condition relative to a pair $(\bar{w}, \bar{z})$ in the graph of the subgradient mapping ∂f, according to which the subgradients and associated function values of f behave exactly as if f were convex. In general, the definition requires localization in function values as well as subgradients, but in the presence of subdifferential continuity, as in the applications to be made here, localization in function values is superfluous. Variational convexity of f with respect to a pair $(\bar{w}, \bar{z}) \in \operatorname{gph} \partial f$ refers then, more simply, to having open convex neighborhoods $\mathcal{W}$ of $\bar{w}$ and $\mathcal{Z}$ of $\bar{z}$ such that

$$\begin{array}{c} \text{there exists a proper lsc convex function } h \leq f \text{ on } \mathcal{W} \text{ such that} \\ [\mathcal{W} \times \mathcal{Z}] \cap \operatorname{gph} \partial h \;=\; [\mathcal{W} \times \mathcal{Z}] \cap \operatorname{gph} \partial f \\ \text{and, for } (w, z) \text{ belonging to this common set, also } h(w) = f(w). \end{array} \tag{2.1}$$

Variational *strong* convexity has h *strongly* convex on $\mathcal{W}$.

Motivation for this property comes from attempts to understand localized monotonicity aspects of subgradient mappings. It was proved long ago by Poliquin [12] that the mapping ∂f is maximal monotone as a whole if and only if f is a convex function, but what about maximal monotonicity of ∂f relative to a neighborhood $W \times Z$ of $(\bar{w}, \bar{z})$? Variational convexity implies that and was shown in [20] to be equivalent to it when $\bar{z}$ is a regular subgradient at $\bar{w}$, as in (1.13). Earlier, in [13], local strong monotonicity of ∂f was tied to the condition of variational strong convexity (without it yet having that name), but only under the additional assumption of "prox-regularity" on f. It was identified there as characterizing the tilt stability of a local minimum.

The consequences of variational convexity for computing a local minimum were first examined in [20], where the localized version of the proximal point algorithm in [11] was translated into iterations of local minimization. Then, in [21], it emerged as essential for guaranteeing the convergence of an approach to problem decomposition with numerous specializations. That is where the second-order condition for local optimality, which we are about to work with, was introduced.

Definition (variational approach to second-order sufficiency)**.** *The variational sufficient condition for local optimality in* (P) *holds with respect to* $\bar{x}$ *and* $\bar{y}$ *satisfying the first-order condition if there exists* $r > 0$ *such that* φ_r *is variationally convex with respect to the pair* $((\bar{x}, 0), (0, \bar{y}))$ *in* $\operatorname{gph} \partial\varphi_r$*. The strong variational sufficient condition holds if* φ_r *is variationally strongly convex.*

Specifically in translating (2.1) to φ_r, the variational convexity in this definition refers to having open convex neighborhoods $\mathcal{W}$ of $(\bar{x}, 0)$ and $\mathcal{Z}$ of $(0, \bar{y})$ such that

$$\begin{array}{l} \text{there exists a proper lsc convex function } \psi \leq \varphi_r \text{ on } \mathcal{W} \text{ such that} \\ \qquad [\mathcal{W} \times \mathcal{Z}) \cap \operatorname{gph} \partial\psi \;=\; [\mathcal{W} \times \mathcal{Z}) \cap \operatorname{gph} \partial\varphi_r \\ \text{and, for } (x, u; v, y) \text{ belonging to this common set, } \psi(x, u) = \varphi_r(x, u). \end{array} \tag{2.2}$$

Variational strong convexity has ψ strongly convex on $\mathcal{W}$. Note that the first-order condition on $\bar{x}$ and $\bar{y}$ in (P) in (1.17), as captured equally by $(0, \bar{y}) \in \partial\varphi_r(\bar{x}, 0)$, is reduced by (2.2) to $(0, \bar{y}) \in \partial\psi(\bar{x}, 0)$, which through convexity guarantees local optimality in minimizing $\psi(x, u)$ subject to $x \in \mathcal{X}$ and $u = 0$. Local optimality in (P) then follows, because $\psi \leq \varphi_r$ and $\psi(\bar{x}, 0) = \varphi_r(\bar{x}, 0)$.

It may seem far-fetched that the kind of property in (2.2) could actually be useful, but it will be revealed in what follows to take on a basic role in the theory of optimality. Far from being obscure and inaccessible, it will be seen in later sections of this paper to build on second-order conditions that have already served well for special cases of problem (P), and moreover to be characterizable in the $\mathcal{C}^2$ case by second-derivative conditions.

In the strong version of variational sufficiency, strong convexity of ψ on $\mathcal{W}$ with modulus $s > 0$ corresponds to the convexity on $\mathcal{W}$ of $\psi(x, u) - \frac{s}{2}|(x.u)|^2$, which can be described equivalently as

$$\begin{array}{l} \psi(x', u') \geq \psi(x, u) + (v, y) \cdot [(x', u') - (x, u)] + \frac{s}{2}|(x', u') - (x, u)|^2 \\ \text{for } \ (x', u') \in \mathcal{W} \ \text{ when } \ (v, y) \in \partial\psi(x, u). \end{array} \tag{2.3}$$

That translates in (2.2) to φ_r having the property that

$$\begin{array}{c} \varphi_r(x', u') \geq \varphi_r(x, u) + (v, y) \cdot [(x', u') - (x, u)] + \frac{s}{2}|(x', u') - (x, u)|^2 \\ \text{for } \ (x', u') \in \mathcal{W} \ \text{ when } \ (v, y) \in \mathcal{Z} \cap \partial\varphi_r(x, u). \end{array} \tag{2.4}$$

From that angle, strong variational sufficiency can be seen as *a parametrically extended form of the "quadratic growth condition"* for a locally optimal solution $\bar{x}$ to (P), namely that $\varphi(x', 0) \geq \varphi(\bar{x}, 0) + \frac{s}{2}|x' - \bar{x}|$ for x' in a neighborhood of $\bar{x}$. That condition is the particular case of (2.4) for $(x, u) = (\bar{x}, 0)$ and $(v, y) = (0, \bar{y})$, with $u' = 0$.

Another insight into the significance of strong variational sufficiency comes from its effect of "subgradient regularization" in replacing $\partial\varphi$ by $\partial\varphi_r$. Variational strong convexity of φ_r with respect to $(0, \bar{y}) \in \partial\varphi_r(\bar{x}, 0)$ corresponds to the set-valued inverse mapping $\partial\varphi_r^{-1}$ having a single-valued locally Lipschitz localization at $(0, \bar{y})$ for $(\bar{x}, 0)$ with modulus s^{-1} [20, Theorem 3]. This means that the mapping $\partial\varphi_r$ is *strongly metrically regular* at $(\bar{x}, 0)$ for $(0, \bar{y})$ in the terminology of [4]. Again, strong variational sufficiency is highlighted as the elicitation of an advantageous property in variational analysis through augmentation with r high enough.

More than being just a gimmick for getting local optimality, variational sufficiency also amounts in (2.1) to a sort of local reduction of everything about (P) to properties in a problem of *convex* optimization for ψ. Although that convex problem may seem only "implicit," it comes fully to life through the augmented Lagrangian $l_r(x, y)$, as we now establish.

Theorem 1 (augmented Lagrangian characterization of variational sufficiency)**.** *With respect to $\bar{x}$ and $\bar{y}$ satisfying the first-order optimality condition in (P), the variational sufficient condition for local optimality holds if and only if, for $r > 0$ sufficiently large, there is a closed convex neighborhood $\mathcal{X} \times \mathcal{Y}$ of $(\bar{x}, \bar{y})$ such that $l_r(x, y)$ is convex in $x \in \mathcal{X}$ when $y \in \mathcal{Y}$ as well as concave in $y \in \mathcal{Y}$ when $x \in \mathcal{X}$. Then $(\bar{x}, \bar{y})$ is a saddle point of $l_r(x, y)$ with respect to minimizing in $x \in \mathcal{X}$ and maximizing in $y \in \mathcal{Y}$.*

Proof. Suppose first that variational sufficiency holds, and choose within $\mathcal{W}$ and $\mathcal{Z}$ open convex neighborhoods of the form $\mathcal{X}\times\mathcal{U}$ and $\mathcal{V}\times\mathcal{Y}$, so that through (1.15) and (1.16):

$$\begin{aligned}&(x,u,v,y)\in[(\mathcal{X}\times\mathcal{U})\times[\mathcal{V}\times\mathcal{Y})]\cap\operatorname{gph}\partial\varphi_r \iff\\ &\qquad (x,y)\in\mathcal{X}\times Y,\;\; \nabla_x l_r(x,y)=v\in\mathcal{V},\;\; -\nabla_y l_r(x,y)=u\in\mathcal{U},\\ &, \qquad\qquad \text{and then moreover}\;\; l_r(x,y)=\varphi_r(x,u)-y{\cdot}u.\end{aligned}\tag{2.5}$$

Consider ψ in (2.2) to be closed proper convex as a function on all of $I\!R^n\times I\!R^m$ (as can harmlesly be achieved by extension), and let λ on $I\!R^n\times I\!R^m$ be the Lagrangian associated with ψ:

$$\lambda(x,y)=\inf_u\{\,\psi(x,u)-y{\cdot}u\}.\tag{2.6}$$

By the rules of convex analysis, $\lambda(x,y)$ is convex in x as well as concave in y, and

$$(v,y)\in\partial\psi(x,u)\iff v\in\partial_x\lambda(x,y),\;\; u\in\partial_y[-\lambda](x,y),\;\; \text{and then}\;\; \lambda(x,y)=\psi(x,u)-y{\cdot}u\tag{2.7}$$

[23, 11.48]. In combining this with the properties in (2.2), we see that

$$\begin{aligned}&\text{for}\;\; (x,y)\in\mathcal{X}\times\mathcal{Y}\;\; \text{and}\;\; (v,u)\in\mathcal{V}\times\mathcal{U},\\ &\quad (\nabla_x l_r(x,y),-\nabla_y l_r(x,y))=(v,u))\iff (v,u)\in(\partial_x\lambda(x,y),\partial_y[-\lambda](x,y)),\\ &\qquad \text{and then moreover}\;\; l_r(x,y)=\lambda(x,y).\end{aligned}\tag{2.8}$$

Because $(\nabla_x l_r(x,y),-\nabla_y l_r(x,y))$ depends continuously on (x,y) and equals (0,0) for $(\bar{x},\bar{y})$, we can ensure that it belongs to $\mathcal{V}\times\mathcal{U}$ by taking the neighborhood $\mathcal{X}\times\mathcal{Y}$ of $(\bar{x},\bar{y})$ somewhat smaller, if necessary. Then, according to (2.8), $l_r(x,y)=\lambda(x,y)$ when $(x,y)\in\mathcal{X}\times\mathcal{Y}$, so l_r is convex-concave on $\mathcal{X}\times\mathcal{Y}$. The same then holds in passing to the closure of $\mathcal{X}\times\mathcal{Y}$, inasmuch as convexity in x and concavity in y are preserved when taking limits. Thus, $\mathcal{X}$ and $\mathcal{Y}$ can be replaced by their closures.

Conversely, suppose l_r is convex-concave on a closed convex neighborhood $\mathcal{X}\times\mathcal{Y}$ of $(\bar{x},\bar{y})$. Define $\lambda(x,y)$ by

$$\lambda(x,y)=\begin{cases} l_r(x,y)\;\;\text{when}\;\;(x,y)\in\mathcal{X}\times\mathcal{Y},\\ -\infty\;\;\text{when}\;\; x\in\mathcal{X}\;\;\text{but}\;\; y\notin\mathcal{Y},\\ \infty\;\;\text{when}\;\; x\notin\mathcal{X},\end{cases}\tag{2.9}$$

and define $\psi(x,u)$ by

$$\psi(x,u)=\sup_y\{\,\lambda(x,y)+y{\cdot}u\}.\tag{2.10}$$

Then ψ is a closed proper convex function on $I\!R^n\times I\!R^m$ having λ as its convex-concave Lagrangian as in (2.6):

$$\lambda(x,y)=\inf_u\{\,\psi(x,u)-y{\cdot}u\}.\tag{2.11}$$

In particular from (2.9)–(2.10),

$$\psi(x,u)=\sup_{y\in\mathcal{Y}}\{\,l_r(x,y)+y{\cdot}u\}\;\;\text{when}\;\; x\in\mathcal{X},$$

so that $\psi\le\varphi_r$ on $\mathcal{X}\times I\!R^m$ by (1.10). In taking $\mathcal{V}\times\mathcal{U}=I\!R^n\times I\!R^m$, we trivially have both (2.5) and (2.7). That leads directly to (2.2) and the claimed variational sufficiency.

Finally, we look at the first-order condition in its equivalent expression in (1.19) as the combination of $\nabla_x l_r(\bar{x},\bar{y})=0$ and $\nabla_y l_r(\bar{x},\bar{y})=0$. The first of these equations says, through the convexity in x, that $l_r(x,\bar{y})$ has a minimum at $\bar{x}$ over the neighborhood $\mathcal{X}$, whereas the second says, through the concavity in y, that $l_r(\bar{x},y)$ has a maximum at $\bar{y}$ over the neighborhood $\mathcal{Y}$. Thus, $(\bar{x},\bar{y})$ furnishes a saddle point of l_r over $\mathcal{X}\times\mathcal{Y}$, as claimed. □

In managing to identify the variational sufficient condition with a convex-concave-type saddle point property, Theorem 1 opens the door to interpreting the local optimality associated with that condition in the duality format of convex analysis [14], [15], [23, Chapter 11].

Local primal and dual problems. *In the saddle point circumstances of Theorem 1, let*

$$\begin{aligned}\Psi(x,u) &= \begin{cases} \sup_{y\in\mathcal{Y}}\{\, l_r(x,y) + y{\cdot}u\} & \textit{when } x\in\mathcal{X}, \\ \infty \ \textit{when } x\notin\mathcal{X}, \end{cases} \\ \Theta(v,y) &= \begin{cases} \inf_{x\in\mathcal{X}}\{\, l_r(x,y) - v{\cdot}x\} & \textit{when } y\in\mathcal{Y}, \\ -\infty \ \textit{when } y\notin\mathcal{Y}, \end{cases}\end{aligned} \tag{2.12}$$

The local primal problem is

$(P^r_{\mathcal{X}\times\mathcal{Y}})$ *minimize* $\Psi(x,u)$ *subject to* $u=0$,

whereas the local dual problem is

$(D^r_{\mathcal{X}\times\mathcal{Y}})$ *maximize* $\Theta(v,y)$ *subject to* $v=0$.

The incorporation of v as a perturbation variable in the dual problem, to match that role for u in the primal problem, is essential for obtaining a full picture of the symmetric interconnections between these variables and x and y. Dual problems with their own perturbations have been a theme since the early days of convex analysis.

The formulas for Ψ and Θ in (2.12) can be written in terms of the function λ defined in (2.9). For ψ this is (2.10), as utilized in the proof of Theorem 1 for the fact that then, through conjugacy, λ can be recovered by (2.11). In parallel, $\Theta(v,y) = \inf_x \{\, \lambda(x,y) - v{\cdot}x\}$. A comparison of that with (2.11) reveals another conjugacy:

$$\Theta(v,y) = \inf_{x,u}\{\, \Psi(x,u) - v{\cdot}x - y{\cdot}u\} = -\Psi^*(v,y). \tag{2.13}$$

The augmented Lagrangian property in Theorem 1 is equivalent to the "strong duality" assertion that

$$\bar{x} \text{ solves } (P^r_{\mathcal{X}\times\mathcal{Y}}),\ \ \bar{y} \text{ solves } (D^r_{\mathcal{X}\times\mathcal{Y}}), \text{ and } \ \min(P^r_{\mathcal{X}\times\mathcal{Y}}) = \max(D^r_{\mathcal{X}\times\mathcal{Y}}). \tag{2.14}$$

This local duality has a powerful implication for any computational method that aims to solve (P) by generating sequences of primal and dual vectors x^k and y^k with the hope that they will converge some locally optimal $\bar{x}$ and associated multiplier $\bar{y}$. Suppose that the variational sufficient condition holds for these elements, and that the method can be articulated as operating locally around $(\bar{x},\bar{y})$ with utilization only of the local properties of the augmented Lagrangian l_r. *Then everything about the method and its convergence reduces to its characteristics when applied in convex optimization.*

A prominent class of algorithms in this category is comprised of the augmented Lagrangian-based *multiplier methods* (ALM), which follow the pattern:

$$\begin{array}{l}\text{get } x^{k+1} \text{ by locally minimizing } \ l_r(x,y^k) \ \text{ in } x \text{ for } r \text{ sufficiently high,} \\ \text{and afterward update } y^k \text{ to } y^{k+1} \text{ by a formula } \ y^{k+1} = U_r(x^{k+1},y^k).\end{array} \tag{2.15}$$

Some variants add to $l_r(x,y^k)$ a localizing term $\rho(x,x^k)$, which is designed to keep x^{k+1} from straying too far from x^k. Some replace the fixed r by an increasing sequence of values r_k.

Applications to such algorithms must be set aside for presentation outside of this paper [22]. For now, the message is just that the parametric behavior with respect to y in the minimization of $l_r(x,y)$ in x is an important ingredient in such computations. This is where the *strong* variational sufficiency condition will contribute key assistance.

The strong form of the variational sufficient condition for local optimality has, by definition, strong convexity of the function ψ in the variational convexity of φ_r in (2.2). What extra property of the augmented Lagrangian, beyond that in Theorem 1, should this correspond to?

We have seen in the proof of Theorem 1 that in this situation the augmented Lagrangian has a local expression in terms of ψ, namely

$$l_r(x,y) = \min_u \{\, \psi(x,u) - y{\cdot}u\} \ \text{ for } \ (x,y) \in \mathcal{X} \times \mathcal{Y}. \tag{2.16}$$

It looks from this that the strong convexity of φ might carry over to strong convexity of $l_r(x,y)$ in $x \in \mathcal{X}$ when $y \in \mathcal{Y}$. That turns out to be correct, but it's not the whole story. To get to the full picture, tilt stability of a minimum should be considered.

Recall that, with respect to $\bar{x}$ furnishing the minimum over $\mathcal{X}$ of the convex function $l_r(\cdot,\bar{y})$, the additional feature associated with strong convexity is *tilt stability* in the sense that the mapping

$$v \mapsto \operatorname*{argmin}_{x \in \mathcal{X}} \{\, l_r(x,\bar{y}) - v{\cdot}x\} \tag{2.17}$$

is single-valued and Lipschitz continuous for v in some neighborhood of $v = 0$. What will come up here is an enhancement of that property which replaces fixed $\bar{y}$ by variable y.

Definition (augmented tilt stability)**.** *In the context of the local convexity-concavity of $l_r(x,y)$ in Theorem 1, the property of augmented tilt stability will be said to hold if there is a neighborhood $\mathcal{V}$ of 0 such that the mapping*

$$(v,y) \mapsto \operatorname*{argmin}_{x \in \mathcal{X}} \{\, l_r(x,y) - v{\cdot}x\} \ \text{ for } \ (v,y) \in \mathcal{V} \times \mathcal{Y} \tag{2.18}$$

is single-valued and Lipschitz continuous.

Obviously, this stability property is deeply suited to the analysis of algorithms of the ALM category described in (2.15).

Theorem 2 (augmented Lagrangian characterization of strong variational sufficiency)**.** *The strong version of the variational sufficient condition for local optimality corresponds to strengthening the characterization of variational sufficiency in Theorem 1 to include augmented tilt stability. It corresponds equally to having the functions $l_r(\cdot,y)$ on $\mathcal{X}$ for $y \in \mathcal{Y}$ be strongly convex, all with the same modulus of strong convexity. The modulus $s > 0$ for that strong convexity then yields, as s^{-1}, a modulus for the Lipschitz continuity in the augmented tilt stability.*

Proof. We know from the proof of Theorem 1 that the Lagrangian $\lambda(x,y)$ in (2.6) for the proper convex function ψ in the condition (2.2) of variational sufficiency (when considered in its lsc extension to $\operatorname{cl}\mathcal{W}$ and given the value ∞ outside of $\operatorname{cl}\mathcal{W}$) coincides locally around $(\bar{x},\bar{y})$ with $l_r(x,y)$. The question is how strong convexity of ψ is reflected equivalently in properties of λ. The answer will be emerge from an investigation of the convex function ψ^* conjugate to ψ,

$$\psi^*(v,y) = \sup_{x,u} \{\, v{\cdot}x + y{\cdot}u - \psi(x,u)\}, \qquad \psi(x,u) = \sup_{v,y} \{\, v{\cdot}x + y{\cdot}u - \psi^*(v,y)\}, \tag{2.19}$$

which is related to λ by the fact, seen from (2.6), that

$$\text{the convex functions } \lambda(\cdot,y) \text{ and } \psi^*(\cdot,y) \text{ are conjugate to each other.} \tag{2.20}$$

Strong convexity of ψ with modulus $s > 0$ corresponds to ψ^* being a $\mathcal{C}^1$ function with gradient mapping that is Lipschitz continuous with modulus s^{-1} [23, 12.60], given by

$$\nabla\psi^*(v,y) = (\nabla_v\psi^*(v,y), \nabla_y\psi(v,y)) = \operatorname{argmin}_{x,u} \{\, \psi(x,u) - v{\cdot}x - y{\cdot}u\}. \tag{2.21}$$

But there's another way of looking at it, through dualization of the description of that strong convexity in (2.3) (with $\mathcal{W}$ now $I\!R^n \times I\!R^m$). By taking the conjugates of the functions of (x', u') on both sides of the inequality in (2.3), and appealing to the fact that the subgradient mapping $\partial\psi^*$ is the inverse of $\partial\psi$, one obtains

$$\begin{gathered}\psi^*(v', y') \leq \psi^*(v, y) + (x, u)\cdot[(v', y') - (v, y)] + \tfrac{1}{2s}|(v', y') - (v, y)|^2 \\ \text{when } (x, u) \in \partial\psi^*(v, y), \text{ implying } (x, u) = \nabla\psi^*(v, y).\end{gathered} \tag{2.22}$$

Likewise in (2.20), strong convexity of $\lambda(\cdot, y)$ with modulus s corresponds to $\psi^*(\cdot, y)$ being $\mathcal{C}^1$ with its gradient mapping Lipschitz continuous with modulus s^{-1} and

$$\nabla_v\psi^*(v, y) = \mathrm{argmin}_x \{ \lambda(x, y) - v\cdot x \}. \tag{2.23}$$

That is characterized in turn by the corresponding partial version of (2.22),

$$\begin{gathered}\psi^*(v', y) \leq \psi^*(v, y) + x\cdot(v' - v) + \tfrac{1}{2s}|v' - v|^2 \\ \text{when } x \in \partial_v\psi^*(v, y), \text{ implying } x = \nabla_v\psi^*(v, y).\end{gathered} \tag{2.24}$$

From these observations it's evident that strong convexity of ψ, captured by the Lipschitz continuity of $\nabla\psi^*$ as the mapping in (2.21), yields the claimed property of augmented tilt stability. In particular it entails the Lipschitz continuity of the mapping in (2.23), which is equivalent to the strong convexity of $\lambda(x, y)$ with respect to x.

The task that remains is demonstrating how the strong convexity of $\lambda(x, y)$ in x implies that ψ^* is $\mathcal{C}^1$ with Lipschitz continuous gradient, or equivalently has a property of the form in (2.22). There's a slight twist, however. We proceed from here, not with ψ itself, but with its augmentation to $\psi_t(x, u) = \psi(x, u) + \frac{t}{2}|u|^2$ for a value $t > 0$, which can be arbitrarily small. It's easy to see that the variational sufficiency relationship that ψ has with φ_r in (2.17) extends to the same kind of relationship between ψ_t and φ_{r+t}, with at most a minor adjustment of neighborhoods, if necessary. It will suffice, therefore, to show that the strong convexity of $\lambda(x, y)$ in x with modulus s implies, for $t < s$, the strong convexity of ψ_t by way of implying the corresponding dual property of ψ_t^*, namely the property in (2.22) for t and ψ_t^* in place of s and ψ^*.

In terms of the augmented Lagrangian λ_t for ψ_t, which is given by $\lambda_t(x, y) = \min_u \{ \psi_t(x, u) - y\cdot u \}$, we have

$$\inf_x \{ \lambda_t(x, v) - v\cdot x \} = -\psi_t^*(v, y) \tag{2.25}$$

in parallel to (2.20). On the other hand, because $-\lambda_t(x, \cdot)$ is conjugate to the function $\psi(x, \cdot) + tj$, where $j(u) = \frac{1}{2}|u|^2$, it is given by inf-convolution between $-\lambda(x, \cdot)$ and the conjugate of tj, which is $t^{-1}j$. Thus,

$$\lambda_t(x, y) = \sup_z \Big\{ \lambda(x, z) - v\cdot x - \frac{1}{2t}|z - y|^2 \Big\}, \tag{2.26}$$

and from this the strong convexity of $\lambda(\cdot, z)$ is seen to be inherited by $\lambda_t(\cdot, y)$. The combination of (2.25) and (2.26) tells us that

$$-\psi_t^*(v, x) = \inf_x \Big\{ \sup_z \Big\{ \lambda(x, z) - v\cdot x - \frac{1}{2t}|z - y|^2 \Big\} \Big\}. \tag{2.27}$$

Denote the function of (x, z) on the right side this formula by $\Lambda_{v,y}(x, z)$. Obviously $\Lambda_{v,y}(x, z)$ is strongly concave in z with modulus t^{-1}, while by our assumption on λ it is strongly convex in x with modulus s.

In minimax theory, any convex-concave function on $I\!R^n \times I\!R^m$ has an associated primal problem and an associated dual problem, [15], [23, 11J]. The primal problem in the case of $\Lambda_{v,y}$ is, according to (2.27), the minimization in (2.25). The dual problem, corresponding to a switch between the inf and sup in (2.27), is the maximization in z of the expression

$$\inf_x \Lambda_{v,y}(x,z) = -\psi^*(v,z) - \frac{1}{2t}|z-y|^2. \tag{2.28}$$

The strong convexity in the primal and strong concavity in the dual guarantee that optimal solutions to these problems exist uniquely and form a saddle point of $\Lambda_{v,y}$ [23, 11.52 and 11.40]. In particular, the minimum in the primal equals the maximum in the dual, hence because the minimum in the primal is $-\psi^*(v,y)$, we have

$$\psi^*(v,y) = \min_z \Big\{ \psi(v,z) + \frac{1}{2t}|z-y|^2 \Big\}. \tag{2.29}$$

Denote the expression being minimized in (2.29) by $p_z(v,y)$, noting that it is a differentiable convex function of (v,y), so that

$$\text{for } z \text{ giving the minimum in (2.29)}, \;\; \nabla\psi^*(v,y) = \nabla p_z(v,y) = (\nabla_x \psi^*(v,z), t^{-1}(y-z)). \tag{2.30}$$

Recalling (2.24), and the corresponding property of the term $q_z(y) = \frac{1}{2t}|y-z|^2$ in (2.29), namely that

$$q_z(y') \geq q_z(y) + \nabla q_z(y)\cdot(y'-y) + \frac{1}{2t}|y'-y|^2,$$

we see from (2.30) that

$$\psi_t^*(v',y') \leq \psi_t^*(v,y) + \nabla\psi^* \cdot [(v',y')-(v,y)] + \frac{1}{2s}|v'-v|^2 + \frac{1}{2t}|y'-y|^2.$$

The same inequality holds then if s is replaced by some lower value, so under the assumption that $t < s$, it holds with s replaced by t. We then have the condition on ψ_t^* that is dual to ψ_t being strongly convex with modulus t, and the proof is finished. □

3 Second-derivative criteria for strong variational sufficiency

How is variational sufficiency related to other sufficient conditions for local optimality that have been developed from problem (P), at least in special cases including classical nonlinear programming? Such conditions all rely on second-order differentiation of some variety. Although so far we have only needed the functions $f_0, f_1, \ldots, f_m$ in (P) to be $\mathcal{C}^1$, we can try to see what more may come to light when they are $\mathcal{C}^2$ — which we are calling the $\mathcal{C}^2$ case of (P). Under that stronger assumption, there is a prospect that the strong convexity of the augmented Lagrangian in Theorem 2 can be characterized by some second-derivative property of the f_i's with respect to the pair $(\bar{x},\bar{y})$ in the first-order condition.

A simple idea to keep in mind is that, *if $l_r(x,y)$ happens to be $\mathcal{C}^2$ around $(\bar{x},\bar{y})$, the strong convexity in Theorem 2 will hold if and only if the partial Hessian $\nabla^2_{xx} l_r(\bar{x},\bar{y})$ is positive-definite.* This can be exploited directly in special circumstances — see Example 1 below. In general, though, we can't count on the augmented Lagrangian being $\mathcal{C}^2$ around $(\bar{x},\bar{y})$, because the functions g^r and g_r^* in (1.6)–(1.7) are only $\mathcal{C}^{1+}$. What we nevertheless do have in the $\mathcal{C}^2$ case of (P) is that the augmented Lagrangian l_r is $\mathcal{C}^{1+}$, as noted in (1.12). In other words, if the functions f_i are $\mathcal{C}^2$, then l_r is differentiable and its gradient $\nabla l_r(x,y)$ is locally Lipschitz continuous with respect to (x,y). That's apparent from the formulas for $\nabla l_r(x,y)$ in (1.15) and the Lipschitz continuity of ∇g^r and ∇g_r^*.

The "second-order" aspects of a function being $\mathcal{C}^{1+}$ deserve some review in a context beyond just l_r, because they will also apply to g^r and g_r^*. For a $\mathcal{C}^{1+}$ function f, the local Lipschitz continuity of ∇f causes that gradient mapping to be differentiable almost everywhere; this is according to Rademacher's Theorem, cf. [23, 9.60]. *The Jacobian $\nabla[\nabla f](w)$ at w, a point where ∇f is differentiable, is the Hessian $\nabla^2 f(w)$ of f at w in the extended sense* of [23, 13.1]. It furnishes a quadratic expansion of f,

$$f(w') = f(w) + \nabla f(w)\cdot(w' - w) + \tfrac{1}{2}(w' - w)\cdot\nabla^2 f(w)(w' - w) + o(|w' - w|^2),$$

by [23, 13.2]. When f is *prox-regular* at w [23, 13.27], which will be seen to be true of the functions we'll specifically be working with in what follows, the existence of such an expansion at w is in fact *equivalent* to the twice differentiability of f at w in the extended sense, and it guarantees that the matrix $\nabla^2 f(w)$ is symmetric [23, 13.42].

Convexity and strong convexity of f can be characterized locally through this. *Having the Hessians $\nabla^2 f(w)$ all be positive-semidefinite for w in some open convex set W (wherever they exist, or almost everywhere that they exist) is equivalent to f being convex on W.* That's because, for almost all $w \in W$, the function $\theta(\tau) = f(w + \tau\omega)$ will be $\mathcal{C}^{1+}$ on the τ-line segment where $w + \tau\omega \in W$, moreover with $\theta'(\tau)$ locally Lipshitz in τ and $\theta''(\tau) = \omega\cdot\nabla^2 f(w + \tau\omega)\omega$. Since a Lipshitz continuous function is the integral of its derivative (existing almost everywhere and essentially bounded locally), positive-semidefiniteness of the Hessians corresponds to the function $\theta'(\tau)$ being nondecreasing and therefore to θ being convex. Since f is continuous, its convexity on segments in W with only negligible exceptions means its convexity on W. Similarly, *having the Hessians $\nabla^2 f(w)$ all be positive-definite for w in some open convex set W (wherever they exist, or almost everwhere that they exist), and uniformly so (i.e., $\omega\cdot\nabla^2 f(w)\omega \geq \alpha|\omega|^2$ for some $\alpha > 0$), is equivalent to f being strongly convex on W.*

But this characterization of local strong convexity of a $\mathcal{C}^{1+}$ prox-regular function f can go farther. The norms of the matrices $\nabla^2 f(w)$ are bounded by any local Lipschitz constant for ∇f, so we can take limits and define

$$\overline{\nabla}^2 f(\bar{w}) = \{\, H \mid \exists w_k \to \bar{w} \ \text{ with } \ \nabla^2 f(w_k) \to H \} \tag{3.1}$$

to obtain a compact set of symmetric matrices, all with norms bounded by that Lipschitz constant; this is the *Hessian bundle* for f at $\bar{w}$. From the compactness of $\overline{\nabla}^2 f(\bar{w})$, we then get that

$$f \text{ is strongly convex around } \bar{w} \text{ if and only if every } H \in \overline{\nabla}^2 f(\bar{w}) \text{ is positive-definite.} \tag{3.2}$$

The convex hull of $\overline{\nabla}^2 f(\bar{w})$ would be the Clarke generalized Jacobian of the gradient mapping ∇f at $\bar{w}$, but taking convex hulls is unnecessary for our purposes and would only get in the way.

Another comparison can be made with the second-order subdifferential $\partial^2 f(\bar{w})$, a set-valued mapping introduced by Mordukhovich [8], [9], which for continuously differentiable f is the coderivative $D^*(\nabla f)(\bar{w})$ of ∇f at $\bar{w}$. We employed this in characterizing strong convexity in [13]. Our result there, specialized to the function f being $\mathcal{C}^{1+}$ as well as prox-regular, says that

$$f \text{ is strongly convex around } \bar{w} \iff \omega\cdot\zeta > 0 \text{ for all } \zeta \in \partial^2 f(\bar{w})(\omega) \text{ when } \omega \neq 0. \tag{3.3}$$

This comes out actually to be the same as the criterion in (3.2), because

$$\min\{\, \omega\cdot\zeta \mid \zeta \in \partial^2 f(\bar{w})\omega \} = \min\{\, \omega\cdot H\omega \mid H \in \overline{\nabla}^2 f(\bar{w}) \} \tag{3.4}$$

by [23, 9.62]. The disadvantage of (3.3) over (3.2), however, is that $\partial^2 f(\bar{w})$ is more troublesome to compute and can involve extraneous elements which cause asymmetry without contributing any additional information about "positive-definiteness."

This can be seen even in very elementary examples like the strongly convex function f on $I\!R$ having $f'(w) = a^+ w$ for $w \geq 0$ and $f'(w) = a^- w$ for $w \leq 0$, with $a^+ > a^- > 0$. In this one-dimensional setting, of course, Hessians are just numbers, and $\overline{\nabla}^2 f(0)$ simply reduces for this f to $\{a^-, a^+\}$. The set-valued mapping $\omega \mapsto \partial^2 f(0)(\omega)$, on the other hand, calculates out to give $\partial^2 f(0)(\omega) = \{\, a^-\omega, a^+\omega\}$ when $\omega \leq 0$, but $\partial^2 f(0)(\omega) = \{\, a\omega \,|\, a \in [a^-, a^+]\}$ when $\omega > 0$. In the specialization of (3.4) to this f, the symmetry-destroying elements $a\omega$ when $\omega > 0$ have no effect on the minimum.

When dealing with the $\mathcal{C}^{1+}$ functions g^r and g_r^* in (1.6), we have prox-regularity directly from convexity and aren't concerned with strong convexity, but we do need to understand $\overline{\nabla}^2 g^r(u)$ and $\overline{\nabla}^2 g_r^*(y)$ for the sake of the role they will soon come to play. In consequence of the definitions, the gradient mappings ∇g^r and ∇g_r^* are Lipschitz continous with global Lipschitz constants r and r^{-1}, respectively. Therefore, the Hessian bundle $\overline{\nabla}^2 g^r(u)$ is a compact set of symmetric positive-semidefinite matrices G with norms $\leq r$, and likewise for $\overline{\nabla}^2 g_r^*(y)$ except with r replaced by r^{-1}.

Theorem 3 (Hessian bundle criterion for strong variational sufficiency). *In the $\mathcal{C}^2$ case of (P), the augmented Lagrangian l_r is a $\mathcal{C}^{1+}$ prox-regular function with Hessian bundles $\overline{\nabla}^2 l_r(x,y)$ consisting of symmetric matrices H that can be partitioned into H_{xx}, H_{xy}, H_{yx}, H_{yy}, relative to the x and y arguments. Let $\overline{\nabla}^2_{xx} l_r(x,y)$ denote the set of submatrices H_{xx} for $H \in \overline{\nabla}^2 l_r(x,y)$. Then the properties of l_r in Theorem 2 around $(\bar{x},\bar{y})$ that correspond to strong variational sufficiency are equivalent to the following:*

$$\text{every } \; H_{xx} \in \overline{\nabla}^2_{xx} l_r(\bar{x},\bar{y}) \; \text{ is positive-definite.} \tag{3.5}$$

Moreover the matrices $H_{xx} \in \overline{\nabla}^2_{xx} l_r(\bar{x},\bar{y})$ are the matrices of the form

$$\nabla^2_{xx} L(\bar{x},\bar{y}) + \nabla F(\bar{x})^* G \nabla F(\bar{x}) \;\text{ for some }\; G \in \overline{\nabla}^2 g^r(F(\bar{x}) + r^{-1}\bar{y}). \tag{3.6}$$

Proof. To verify that $l_r(x,y)$ is prox-regular, we show that it is strongly amenable, i.e., representable as the composition of a $\mathcal{C}^2$ mapping with a convex function under an appropriate constraint qualification, since that is known to yield prox-regularity [23, 13.22]. The formula $l_r(x,y) = f_0(x) + g^r(F(x) + \frac{1}{r}y) - \frac{1}{2r}|y|^2$ can be interpreted as coming from the composition of the $\mathcal{C}^2$ mapping $C : (x,y) \mapsto (f_0(x), F(x) + \frac{1}{r}y, \frac{1}{2r}|y|^2)$ with the convex function $c(\omega,u,\eta) = \omega + g^r(u) - \eta$. The constraint qualification in question, involving $\operatorname{dom} c$, is trivially satisfied because c is finite everywhere.

Next we determine the Hessians $\nabla^2 l_r(x,y)$ at the points where the gradient mapping ∇l_r is differentiable by appealing to the formula in (1.15), which we consolidate here as

$$(\nabla_x l_r(x,y), \nabla_y l_r(x,y)) = (\nabla_x L(x,\eta), r^{-1}(\eta - y)), \;\text{ where }\; \eta = \nabla g^r(F(x) + r^{-1}y), \tag{3.7}$$

with the recollection that the first-order condition on $(\bar{x},\bar{y})$ has $(\nabla_x l_r(\bar{x},\bar{y}), \nabla_y l_r(\bar{x},\bar{y})) = (0,0)$, in which case $\eta = \bar{y}$. If g^r were $\mathcal{C}^2$, we could rely on the standard chain rule for differentiating (3.7), but in the circumstances faced here something more delicate is needed. Here's the fact that will help: if $M(w) = M_1(M_2(w))$ for a locally Lipshitz mapping M_1 and a $\mathcal{C}^1$ mapping M_2 with Jacobians of full rank, then M is almost everywhere differentiable with $\nabla M(w) = \nabla M_1(M_2(w))\nabla M_2(w)$. The full rank assumption guarantees that, for almost every w, M_1 is differentiable at $M_2(w)$. The first-order expansion of M_1 at $M_2(w)$ combines then with the first-order expansion of M_2 at w to provide a first-order expansion of M at w with the indicated Jacobian $\nabla M(w)$, and that establishes the claim.

We apply this version of a chain rule to $M(x,y) = \nabla g^r(F(x) + r^{-1}y)$ with $M_1 = \nabla g^r$ and $M_2(x,y) = F(x) + r^{-1}y$, for which $\nabla_x M_2(x,y) = \nabla F(x)$ and $\nabla_y M_2(x,y) = r^{-1}I$. This yields $\nabla_x M(x,y) = G\nabla F(x)$ and $\nabla_y M(x,y) = r^{-1}G$ for $G = \nabla^2 g^r(F(x)+r^{-1}y)$ at points of differentiability.

For the rest of the differentiation of (3.7), ordinary rules can be invoked. The result of the calculation at points of differentiability is that, with η as in (3.7),

$$\begin{aligned}\nabla^2_{xx} l_r(x,y) &= \nabla^2_{xx} L(x,\eta) + \nabla F(x)^* G \nabla F(x) \;\text{ for }\; G = \nabla^2 g^r(F(x)+r^{-1}y), \;\text{ while}\\ \nabla^2_{xy} l_r(x,y) &= \nabla F(x)^* G, \quad \nabla^2_{yx} l_r(x,y) = G\nabla F(x), \quad \nabla^2_{yy} l_r(x,y) = r^{-2}[G - rI].\end{aligned} \tag{3.8}$$

Then, in passing to the bundle $\overline{\nabla}^2 l_r(x,y)$, we get the same expressions but with $G \in \overline{\nabla}^2 g^r(F(x)+r^{-1}y)$. The special case of $(\bar{x},\bar{y})$ is covered by this as well, with the feature that η becomes $\bar{y}$. In particular, the description of $\overline{\nabla}^2_{xx} l_r(\bar{x},\bar{y})$ in (3.6) is thereby confirmed.

The positive-definiteness of the matrices in $\overline{\nabla}^2_{xx} l_r(\bar{x},\bar{y})$ corresponds to the existence of a neighborhood of $(\bar{x},\bar{y})$ in which every (x,y) has all the matrices in $\overline{\nabla}^2_{xx} l_r(x,y)$ positive definite, and uniformly so. Then, locally in y near $\bar{y}$, $l_r(x,y)$ is uniformly strongly convex in x near $\bar{x}$, as was to be proved. $\square$

There is work left to do, because we still have to see more clearly the conditions on first derivatives and generalized second-derivatives that correspond in problem (P) to the existence of $r > 0$ such that the property in Theorem 3 holds for l_r at $(\bar{x},\bar{y})$. For that, more will have to be uncovered about the Hessian properties of g^r. Right away, though, we can draw some conclusions that are readily at hand, ahead of the coming investigation of the ability of Hessian properties of g^r to reflect geometric "curvatures" that can influence local optimality.

Example 1 (strong variational sufficiency in nonlinear programming)**.** *In the $\mathcal{C}^2$ case of classical nonlinear programming, where $g = \delta_K$ for the cone K in (1.2), let $\bar{x}$ and $\bar{y}$ satisfy the first-order conditions, and let $I(\bar{x},\bar{y})$ be the set of indices i corresponding to the equality constraints and the active inequality constraints having Lagrange multipliers $\bar{y}_i > 0$. Then the strong variational sufficient condition holds if and only if*

$$\begin{aligned}&\nabla^2_{xx} L(\bar{x},\bar{y}) \textit{ is positive-definite relative to the subspace}\\ &S(\bar{x},\bar{y}) = \{\, \xi \mid \nabla f_i(\bar{x})\cdot\xi = 0 \;\textit{ for every } f_i \textit{ with } i \in I(\bar{x},\bar{y})\}.\end{aligned} \tag{3.9}$$

The tighter form of this standard condition in which all the active inequality constraints are required to have multipliers $\bar{y}_i > 0$ is equivalent to the augmented Lagrangian l_r being $\mathcal{C}^2$ around $(\bar{x},\bar{y})$ for r sufficiently high, with $\nabla^2_{xx} l_r(\bar{x},\bar{y})$ positive-definite.

Detail. For this choice of g the function g^r is $\frac{r}{2}d_K^2$ for $d_K(u)$ giving the distance of u from the indicated cone K. The augmented Lagrangian thus has the formula

$$l_r(x,y) = f_0(x) + \frac{r}{2}\left[\sum_{i=1}^{s} \max\{0, f_i(x) + r^{-1}y_i\}^2 + \sum_{i=s+1}^{m} (f_i(x) + r^{-1}y_i)^2\right] - \frac{1}{2r}\sum_{i=1}^{m} y_i^2.$$

Let $I_+(x,y)$ denote the set of indices $\in [1,s]$ such that $f_i(x) + r^{-1}y_i > 0$, and similarly $I_0(x,y)$ for $= 0$ and $I_-(x,y)$ for < 0. Around (x,y) having $I_0(x,y) = \emptyset$, the terms for $i \in I_-(x,y)$ vanish, and l_r is $\mathcal{C}^2$ with x-Hessian

$$\nabla^2_{xx} l_r(x,y) = \nabla^2_{xx} f_0(x) + \sum_{i \in I_+(x,y)\cup[s+1,m]} \Big[(y_i + rf_i(x))\nabla^2 f_i(x) + r\nabla f_i(x)^*\nabla f_i(x)\Big]. \tag{3.10}$$

In the first-order condition on $(\bar{x},\bar{y})$, we have for the inequality constraints that either $f_i(\bar{x}) = 0$ with $\bar{y}_i \geq 0$ or $f_i(\bar{x}) < 0$ with $\bar{y}_i = 0$. Then, in taking the possible limits of the Hessians in (3.10) as $(x,y) \to (\bar{x},\bar{y})$, the only difference that emerges is for indices $i \in I_0(\bar{x},\bar{y})$, which have $f_i(\bar{x}) = 0$ and

$\bar{y}_i = 0$, since for them y_i can approach $\bar{y}_i$ either positively or negatively. The matrices in the Hessian limit $\overline{\nabla}^2_{xx} l_r(\bar{x},\bar{y})$ are therefore exactly the ones of the form

$$\nabla^2_{xx} L(\bar{x},\bar{y}) + r \sum_{i \in I_+(\bar{x},\bar{y}) \cup [s+1,m] \cup J} \nabla f_i(\bar{x})^* \nabla f_i(\bar{x}) \text{ for some } J \subset I_0(\bar{x},\bar{y}). \tag{3.11}$$

Such a matrix is positive-definite for high-enough r as long as $\nabla^2_{xx} L(\bar{x},\bar{y})$ is positive definite relative to the subspace consisting of the vectors ξ such that $\nabla f_i(\bar{x}) \cdot \xi = 0$ for all $i \in I_+(\bar{x},\bar{y}) \cup [s+1,m] \cup J$. Since J can be any subset of $I_0(\bar{x},\bar{y})$, having this for every matrix in (3.11) comes down to having it for $J = \emptyset$, which corresponds to the biggest subspace. This yields the criterion in (3.9). When $I_0(\bar{x},\bar{y}) = \emptyset$, there is only one matrix in (3.11), and we get the $\mathcal{C}^2$ case of l_r as described. □

Example 1 confirms that the strong variational sufficient condition for local optimality reduces, in nonlinear programming, to the strong second-order sufficient condition that is commonly invoked in numerical methodology. That condition therefore does more than provide a convenient guarantee of local optimality. It corresponds, through Theorems 1 and 2, to a *local reduction of the nonlinear programming problem to a convex optimization problem in a framework of local duality.* At the same time, Example 1 demonstrates that strong variational sufficiency is a *fundamental and far-reaching generalization* of this classical sufficient condition.

The characterization of strong variational sufficiency in Example 1 in terms of positive-definiteness relative to a critical subspace extends in the theorem coming next to a larger class of problems, but some more insight into the gradient mapping ∇g^r and its potential differentiability is needed in providing support for that result. The important observation is that

$$\operatorname{gph} \nabla g^r = \{ (u + r^{-1}y, y) \mid (u,y) \in \operatorname{gph} \partial g \}, \tag{3.12}$$

which has the interpretation that

$$\operatorname{gph} \nabla g^r = A_r(\operatorname{gph} \partial g) \text{ for the linear transformation } A_r : (u,y) \mapsto (u + r^{-1}y, y). \tag{3.13}$$

This follows from the conjugacy in (1.6) and the fact that the subdifferential mappings of a convex function and its conjugate are inverse to each other:

$$\begin{aligned} y \in \partial g(u) \quad &\Longleftrightarrow \quad u \in \partial g^*(y) \quad \Longleftrightarrow \quad u + r^{-1}y \in \partial(g^* + r^{-1}j)(y) \text{ for } j(y) = \tfrac{1}{2}|y|^2 \\ &\Longleftrightarrow \quad y \in \partial(g^* + r^{-1}j)^*(u + r^{-1}y), \text{ where } (g^* + r^{-1}j)^* = g^r \text{ by (1.6).} \end{aligned}$$

The same linear transformation A_r in (3.12), which is invertible and thus preserving of tangent structure, can be utilized to understand how the difference quotient mappings

$$[\Delta_\tau \nabla g^r(w)](\omega) = \tau^{-1}[\nabla g^r(w + \tau\omega) - \nabla g^r(w)] \text{ at points } w \in I\!R^m \tag{3.14}$$

are connected with the set-valued difference quotient mappings

$$\begin{aligned} &[\Delta_\tau \partial g(u \,|\, y)](\omega) = \tau^{-1}[\partial g(u + \tau\omega) - y], \text{ for which} \\ &\operatorname{gph} [\Delta_\tau \partial g(u \,|\, y)] = \{ (\omega,\eta) \mid (u,y) + \tau(\omega,\eta) \in \operatorname{gph} \partial g \}. \end{aligned} \tag{3.15}$$

We get from (3.13) that

$$A_r^{-1}(\operatorname{gph}[\Delta_\tau \partial g(u \,|\, y)) = \Big\{ (\omega + r^{-1}\eta, \eta) \,\Big|\, (u + r^{-1}y, y) + \tau(\omega + r^{-1}\eta, \eta) \in \operatorname{gph} \nabla g^r \Big\},$$

or in other words that

$$\operatorname{gph}[\Delta_\tau \nabla g^r(u+r^{-1}y)] = \{\, (\omega + r^{-1}\eta, \eta) \,|\, (\omega,\eta) \in \operatorname{gph}[\Delta_\tau \partial g(u\,|\,y)\}. \tag{3.16}$$

In ascertaining the differentiability of ∇g^r at a point $u + r^{-1}y$, we need to see what happens to the mappings $\Delta_\tau \nabla g^r(u + r^{-1}y)$ as $\tau \searrow 0$. Because those mappings are Lipschitz continuous with modulus r, a property they inherit from ∇g^r, a pointwise limit is automatically uniform on bounded sets and corresponds moreover to set convergence of their graphs [23, 5.45]. But through (3.16) that can be identified with set-convergence of the graphs of the mappings $\Delta_\tau \partial g(u\,|\,y)$, which in fact are the subgradient mappings of the second-order different quotient functions associated with g:

$$\Delta_\tau \partial g(u\,|\,y) = \partial[\tfrac{1}{2}\Delta_\tau^2 g(u\,|\,y)], \text{ where } \tfrac{1}{2}\Delta_\tau^2 g(u\,|\,y)(\omega) = \tau^{-2}[g(u+\tau\omega) - g(u) - \tau y\cdot\omega)]. \tag{3.17}$$

Set convergence of the graphs of subdifferential mappings of convex functions is tied to set convergence of the epigraphs of those functions by Attouch's Theorem [23, p. 552]. Moreover there is the duality that

$$\tfrac{1}{2}\Delta_\tau^2 g(u\,|\,y) \text{ and } \tfrac{1}{2}\Delta_\tau^2 g^*(y\,|\,u) \text{ are conjugate convex functions.} \tag{3.18}$$

For now, we will focus on just a special case of this difference quotient framework, but later, in Section 4, it will be exploited in full. The special case concerns functions g that are polyhedral convex, i.e., for which the epigraph $\operatorname{epi} g$ is a polyhedral convex set, this being equivalent to the same property for the conjugate function g^*. These can also be described as the functions g such that $\operatorname{dom} g$ is a polyhedral convex set on which g is piecewise linear (affine), see [14]. Their one-sided directional derivatives

$$g'(u;\omega) = \lim_{\tau \searrow 0} \tau^{-1}[g(u+\tau\omega) - g(u)]$$

have the distinguishing feature from piecewise linearity that actually

$$\forall \rho > 0, \exists \varepsilon > 0 \text{ such that } g'(u;\omega) = \tau^{-1}[g(u+\tau\omega) - g(u)] \text{ when } |\omega| < \rho,\ \tau < \varepsilon. \tag{3.19}$$

Theorem 4 (strong variational sufficiency in polyhedral and piecewise linear modeling)**.** *Suppose in the $\mathcal{C}^2$ case of (P) that the modeling function g is polyhedral convex. With respect to $y \in \partial g(u)$, or equivalently $u \in \partial g^*(y)$, let*

$$T_g(u\,|\,y) = \{\, \omega \,|\, g'(u;\omega) = y\cdot\omega\}, \qquad T_{g^*}(y\,|\,u) = \{\, \eta \,|\, g^{*\prime}(y;\eta) = u\cdot\eta\}, \tag{3.20}$$

these being polyhedral convex cones polar to each other. Then for $(\bar{x},\bar{y})$ satisfying the first-order conditions in (P), expressed as having $\nabla_x L(\bar{x},\bar{y}) = 0$ with $\bar{y} \in \partial g(F(\bar{x}))$, the strong variational sufficient condition holds if and only if

$$\begin{array}{l} \nabla^2_{xx} L(\bar{x},\bar{y}) \textit{ is positive-definite relative to the subspace} \\ S(\bar{x},\bar{y}) \;=\; \Big\{\, \xi \;\Big|\; \nabla F(\bar{x})\xi \in T_g(F(\bar{x})\,|\,\bar{y}) - T_g(F(\bar{x})\,|\,\bar{y}) \,\Big\}. \end{array} \tag{3.21}$$

That subspace can be described equivalently in the notation $\eta F = \eta_1 f_1 + \cdots + \eta_m f_m$ by

$$S(\bar{x},\bar{y}) \;=\; \Big\{\, \xi \;\Big|\; \nabla[\eta F](\bar{x})\cdot\xi = 0 \textit{ for all } \eta \in T_{g^*}(\bar{y}\,|\,F(\bar{x})) \cap [-T_{g^*}(\bar{y}\,|\,F(\bar{x}))] \Big\}. \tag{3.22}$$

Proof. The polyhedral convexity of the cones in (3.20) is evident from that of g and g^*, which makes the directional derivative functions $g'(u;\cdot)$ and $g^{*\prime}(y;\cdot)$ be piecewise linear and furnishes the supplementary tangent-cone-type descriptions that

$$\begin{array}{l} T_g(u\,|\,y) \;= \{\, \omega \,|\, y \in \partial g(u+t\omega) \text{ for small } t > 0\} = T_{\partial g^*(y)}(u), \\ T_{g^*}(y\,|\,u) = \{\, \eta \,|\, u \in \partial g^*(y+t\eta) \text{ for small } t > 0\} = T_{\partial g(u)}(y). \end{array} \tag{3.23}$$

The polarity relationship will be confirmed in the calculations below. That polarity leads to the equivalence of the descriptions of $S(\bar{x},\bar{y})$ in (3.21) and (3.22), inasmuch as it makes the subspaces $T_g(F(\bar{x})\,|\,\bar{y}) - T_g(F(\bar{x})\,|\,\bar{y})$ and $T_{g^*}(\bar{y}\,|\,F(\bar{x})) \cap [-T_{g^*}(\bar{y}\,|\,F(\bar{x}))]$ be orthogonally complementary. The first is the smallest subspace containing $T_g(F(\bar{x})\,|\,\bar{y})$, while the second is the largest subspace contained within $T_{g^*}(\bar{y}\,|\,F(\bar{x}))$. Having $\nabla F(\bar{x})\xi$ lie in a subspace M is equivalent to having $0 = \eta{\cdot}\nabla F(\bar{x})\xi = \nabla[\eta F](\bar{x}){\cdot}\xi$ for all $\eta \in M^\perp$.

We get from (3.19) that the second-order difference quotient in (3.17) has the reduced form

$$\forall\rho > 0, \exists\varepsilon > 0 \ \text{ such that } \ \tfrac{1}{2}\Delta_\tau^2 g(u\,|\,y)(\omega) = \tau^{-1}[g'(u;\omega) - y{\cdot}\omega] \ \text{ when } \ |\omega| < \rho,\ \tau < \varepsilon. \tag{3.24}$$

Since $g'(u;\omega) - y{\cdot}\omega \geq 0$ in general, with equality meaning that $\omega \in T_g(u\,|\,y)$, we see that

$$\tfrac{1}{2}\Delta_\tau^2 g(u\,|\,y)(\omega) \to \delta_{T_g(u\,|\,y)}(\omega) \ \text{ as } \ \tau\searrow 0. \tag{3.25}$$

In parallel argument,

$$\tfrac{1}{2}\Delta_\tau^2 g^*(y\,|\,u)(\eta) \to \delta_{T_{g^*}(u\,|\,y)}(\eta) \ \text{ as } \ \tau\searrow 0. \tag{3.26}$$

The pointwise convergence in (3.25) of convex functions of ω in the circumstances of (3.24) is equivalent to epi-convergence of those functions, i.e., set convergence of their epigraphs by [23, 7.2]. Conjugacy is preserved by epi-convergence according to Wijsman's Theorem [23, 11.34]. The conjugacy of the difference quotients in (3.18) thus implies the conjugacy of the indicator functions in (3.25) and (3.26), which says that the cones in question are polar to each other. The epi-convergence of the functions $\frac{1}{2}\Delta_\tau^2 g(u\,|\,y)$ also corresponds by Attouch's Theorem [23, p. 552] to set convergence of the graphs of their subdifferential mappings (3.15), the limit being the subdifferential mapping for the indicator $\delta_{T_g(y\,|\,u)}$, which is the normal cone mapping $N_{T_g(y\,|\,u)}$. Because ∂g is a piecewise polyhedral mapping, this convergence of $\Delta_\tau\partial g(u\,|\,y)$ to $N_{T_g(y\,|\,u)}$, brings the existence of a neighborhood $\mathcal{U}\times\mathcal{Y}$ such that

$$[\mathcal{U}\times\mathcal{Y}] \cap \operatorname{gph}[\Delta_\tau\partial g(u\,|\,y)] = [\mathcal{U}\times\mathcal{Y}] \cap \operatorname{gph} N_{T_g(u\,|\,y)} \ \text{ for small } \ \tau > 0. \tag{3.27}$$

Recall that the normal cone mapping N_C for a closed convex set C is related to its projection mapping P_C by $P_C = (I + N_C)^{-1}$. This can be expressed for our purposes here as

$$(\omega,\eta) \in \operatorname{gph} N_C \quad\Longleftrightarrow\quad (\omega + r^{-1}\eta, \eta) \in \operatorname{gph}[r(I - P_C)]$$

and then applied to $C = T_g(u\,|\,y)$ to obtain, from (3.16) and the graphical convergence of $\Delta_\tau\partial g(u\,|\,y)$ to $N_{T_g(u\,|\,y)}$, that

$$\Delta_\tau\nabla g^r(u + r^{-1}y) \ \text{ converges pointwise locally uniformly to } \ r\Big(I - P_{T_g(u\,|\,y)}\Big) \ \text{ as } \ \tau\searrow 0. \tag{3.28}$$

Differentiability of ∇g^r at $u + r^{-1}y$ is the case of (3.28) where the limit is a linear mapping. Because P_C is a linear mapping if and only if C is a subspace, that corresponds precisely to the cone $T_g(u\,|\,y)$ being a subspace, or equivalently through polarity, to the cone $T_{g^*}(y\,|\,u)$ being a subspace. That can be identified through the supplementary formulas (3.23) with the relative interior condition

$$u \in \operatorname{ri}\partial g^*(y), \ \text{ or equivalently } \ y \in \operatorname{ri}\partial g(u). \tag{3.29}$$

This guides us to the question of the possible circumstances in which that property can hold for (u,y) near $(\bar{u},\bar{y}))$ with $\bar{u} = F(\bar{x})$.

Note first that, because $T_g(\bar{u}\,|\,\bar{y})$ is the tangent cone to $\partial g^*(\bar{y})$ at $\bar{u}$, the tangent cone to $\partial g^*(\bar{y})$ at any $u \in \operatorname{ri} \partial g^*(\bar{y})$ is the subspace generated by $T_g(\bar{u}\,|\,\bar{y})$. Thus,

$$T_g(u\,|\,\bar{y}) = T_g(\bar{u}\,|\,\bar{y}) - T_g(\bar{u}\,|\,\bar{y}) \ \text{ at points } u \text{ arbitrarily close to } \bar{u}. \tag{3.30}$$

Observe next from the case of (3.27) for $(\bar{u}, \bar{y})$ that the set $\operatorname{gph} \partial g - (\bar{u}, \bar{y})$ coincides locally around the origin with $\operatorname{gph} N_{T_g(\bar{u}\,|\,\bar{y})}$. Hence, for (u, y) in some neighborhood of $(\bar{u}, \bar{y})$, having $y \in \partial g(u)$, or equivalently $u \in \partial g^*(y)$, entails $u - \bar{u} \in T_g(\bar{u}\,|\,\bar{y})$. Then the tangent cone to $\partial g^*(y)$ at u, which is $T_g(u\,|\,y)$, must lie in the subspace $T_g(\bar{u}\,|\,\bar{y}) - T_g(\bar{u}\,|\,\bar{y})$. It follows that this subspace, generated as in (3.30), includes all others that might be generated from pairs (u, y) approaching $(\bar{u}, \bar{y})$ in the graph of ∂g. Hence the condition in (3.21) is equivalent in these circumstances to the Hessian bundle criterion in Theorem 3 for strong variational sufficiency. □

To illustrate this result, we look at several cases which, primally or dually, involve faces of convex sets. Recall that the *face* of a polyhedral convex set K with respect to a vector y is the polyhedral convex set

$$\operatorname{face}(K\,|\,y) = \operatorname{argmax}\{\, y{\cdot}u \mid u \in K \,\} = \{\, u \in K \mid y \in N_K(u) \,\}. \tag{3.31}$$

Example 2 (illustrations of the criterion in Theorem 4)**.**

(a) *If g is the indicator δ_K of a polyhedral convex set K, then g^* is the support function σ_K for K, while $T_g(F(\bar{x})\,|\,\bar{y})$ is the tangent cone to* $\operatorname{face}(K\,|\,\bar{y})$ *at $F(\bar{x})$, which is the same as the critical cone consisting of the elements of $T_K(F(\bar{x}))$ that are $\perp \bar{y}$. In this case, (3.21) has*

$$T_g(F(\bar{x})\,|\,\bar{y}) - T_g(F(\bar{x})\,|\,\bar{y}) = \textit{ the subspace of } I\!\!R^m \textit{ parallel to the affine hull of } \operatorname{face}(K\,|\,\bar{y}). \tag{3.32}$$

(b) *If g is the support function σ_Y of a polyhedral convex set Y, then g^* is the indicator δ_Y, and $T_{g^*}(\bar{y}\,|\,F(\bar{x}))$ is the tangent cone to* $\operatorname{face}(Y\,|\,F(\bar{x}))$ *at $\bar{y}$, so that in (3.22)*

$$T_{g^*}(\bar{y}\,|\,F(\bar{x}) \cap [-T_{g^*}(\bar{y}\,|\,F(\bar{x})] = \{\, \eta \in I\!\!R^m \mid \bar{y} \pm \tau\eta \in \operatorname{face}(Y\,|\,F(\bar{x})) \ \textit{ for some } \ \tau > 0 \}. \tag{3.33}$$

(c) *Cases (a) and (b) unite when K is a polyhedral convex cone and Y is the polar cone K^*. In particular, when K is the nonlinear programming constraint cone (1.2), the subspace $S(\bar{x}, \bar{y})$ in Theorem 4 becomes the subspace $S(\bar{x}, \bar{y})$ in Example 1.*

(d) *If g is the vecmax function appearing in (1.3), then $g = \sigma_Y$ and $g^* = \delta_Y$ for the unit simplex $Y = \{\, y \mid y_i \geq 0,\ y_1 + \cdots + y_m = 1 \}$. This is an instance of (b) in which a face of Y is a subsimplex:*

$$\operatorname{face}(Y\,|\,F(\bar{x})) = \{\, y \in Y \mid y_i = 0 \ \textit{ for inactive } \ i \},$$

where the indices i called active have $f_i(\bar{x}) = \operatorname{vecmax}(F(\bar{x})) = \max\{f_1(\bar{x}), \ldots, f_m(\bar{x})\}$. Then

$$S(\bar{x}, \bar{y}) = \{\, \xi \mid \nabla f_i(\bar{x}){\cdot}\xi \ \textit{ has the same value for every active } i \textit{ with } \ \bar{y}_i > 0 \}.$$

(e) *If g is the norm $||\cdot||_\infty$, then $g^* = \delta_Y$ for $Y = \{\, y = (y_1, \ldots, y_m) \mid |y_1| + \cdots + |y_m| \leq 1 \} = I\!\!B_1$, the unit ball for the norm $||\cdot||_1$. This is an instance of (b) similar to (d) in which the active indices i have $|f_i(\bar{x})| = ||F(\bar{x})||_\infty$ and*

$$\operatorname{face}(I\!\!B_1\,|\,F(\bar{x})) \ \textit{ restricts } y \in I\!\!B_1 \textit{ to } \begin{cases} y_i = 0 \ \textit{ for inactive } i, \\ y_i \geq 0 \ \textit{ for active } i \textit{ having } \ f_i(\bar{x}) > 0, \\ y_i \leq 0 \ \textit{ for active } i \textit{ having } \ f_i(\bar{x}) < 0. \end{cases}$$

Then $S(\bar{x}, \bar{y}) = \{\, \xi \mid \nabla f_i(\bar{x}){\cdot}\xi \ $ has the same value for all active i with $\ \bar{y}_i \neq 0 \}$.

(f) *If* $g = ||\cdot||_1$*, then* $g^* = \delta_Y$ *for* $Y = [-1,1]^m = I\!B_\infty$*, the unit ball for* $||\cdot||_\infty$*. This is an instance of (b) in which*

$$y \in \mathrm{face}(Y \,|\, F(\bar{x})) \quad \Longleftrightarrow \quad \begin{cases} y_i = 1 \ \ \text{if} \ \ f_i(\bar{x}) > 0, \\ y_i = -1 \ \ \text{if} \ \ f_i(\bar{x}) < 0, \\ y_i \in [-1,1] \ \ \text{if} \ \ f_i(\bar{x}) = 0. \end{cases}$$

Then $S(\bar{x},\bar{y}) = \{\, \xi \,|\, \nabla f_i(\bar{x}){\cdot}\xi = 0 \ \text{ for all } i \text{ such that } \ f_i(\bar{x}) = 0 \ \text{ and } \ -1 < \bar{y}_i < 1 \}$*.*

Detail. (a) When $g = \delta_K$, having $y \in \partial g(u)$, or equivalently $u \in \partial g^*(y)$, reduces to having $y \in N_K(u)$, which is the same as u belonging to $\mathrm{face}(K \,|\, y)$. Hence $\partial g^*(y) = \mathrm{face}(K \,|\, y)$. Since $T_g(u \,|\, y)$ is by (3.23) the tangent cone to $\partial g^*(y)$ at u, it is the cone $\mathrm{face}(K \,|\, y) - u$. Therefore, the subspace $T_g(u \,|\, y) - T_g(u \,|\, y)$ generated by $T_g(u \,|\,)$ is the same as the subspace generated by $\mathrm{face}(K \,|\, y)$. In particular, the subspace $T_g(F(\bar{x}) \,|\, \bar{y}) - T_g(F(\bar{x}) \,|\, \bar{y})$ entering into (3.21) is the one claimed.

(b) This is the dual version of (a) with $T_{g^*}(\bar{y} \,|\, F(\bar{x}))$ being the tangent cone at $\bar{y}$ to $\partial g(F(\bar{x}))$, which here is $\mathrm{face}\,(Y \,|\, F(\bar{x}))$. Thus, $T_{g^*}(\bar{y} \,|\, F(\bar{x})) = \{\, \eta \,|\, \bar{y} + t\eta \in \mathrm{face}\,(Y \,|\, F(\bar{x})) \ \text{ for some } \ t > 0 \}$.

(c) For the polar pair of cones K and Y, we have

$$\begin{aligned} \bar{y} \in N_K(F(\bar{x})) \quad &\Longleftrightarrow \quad F(\bar{x}) \in N_Y(\bar{y}) \quad \Longleftrightarrow \quad F(\bar{x}) \in K, \ \ \bar{y} \in Y, \ \ F(\bar{x}) \perp \bar{y}, \\ &\Longleftrightarrow \quad F(\bar{x}) \in \mathrm{face}(K \,|\, \bar{y}) \quad \Longleftrightarrow \quad \bar{y} \in \mathrm{face}\,(Y \,|\, F(\bar{x})), \end{aligned} \tag{3.34}$$

so that both versions (3.21) and (3.22) of the criterion in Theorem 4 can be looked at. For the nonlinear programming constraint cone (1.2) as an example, $\mathrm{face}(K \,|\, \bar{y})$ is the cone consisting of $u = (u_1, \ldots, u_m)$ such that $u_i \leq 0$ for $i \in [1,s]$ with $\bar{y}_i = 0$, but $u_i = 0$ otherwise. The affine hull of this is already a subspace, obtained by dropping the nonpositivity requirement. That results in $S(\bar{x},\bar{y})$ in (3.21) being the subspace in (3.9). From the other side, $\mathrm{face}\,(Y \,|\, F(\bar{x}))$ is the cone consisting of $y = (y_1, \ldots, y_m)$ such that $y_i \geq 0$ for $i \in [1,s]$ having $f_i(\bar{x}) = 0$ but $y_i = 0$ for $i \in [1,s]$ having $f_i(\bar{x}) < 0$. The vectors η such that $\bar{y} \pm t\eta$ belongs to this for some $\tau > 0$ are those for which $\eta_i = 0$ except for $i \in [1,s]$ such that $f_i(\bar{x}) < 0$, the inactive constraints. The condition in (3.22) then has $S(\bar{x},\bar{y})$ consisting of ξ such that $\xi \perp \sum_{i=1}^m \eta_i \nabla f_i(\bar{x})$ for all such η, and that again is the subspace in (3.9).

(d) This continues (b). For $\bar{y}$ in the indicated subsimplex, the vectors η such that $\bar{y} \pm \tau\eta$ belongs to that subsimplex for some $\tau > 0$ are the vectors $\eta = (\eta_1, \ldots, \eta_m)$ such that $\eta_1 + \cdots \eta_m = 0$ with $\eta_i = 0$ unless $\bar{y}_i > 0$. Then in (3.22) we are looking at ξ such that $\sum_{i=1}^m \eta_i [\nabla f_i(\bar{x}){\cdot}\xi] = 0$ for all such η. That corresponds to the values of $\nabla f_i(\bar{x}){\cdot}\xi$ agreeing for indices i having $\bar{y}_i > 0$.

(e) The argument of (d) needs only minor adjustment to cover this as well.

(f) The face in this instance allows little wiggle room. It contains $\bar{y} \pm \tau\eta$ for some $\tau > 0$ if and only if $\eta_i = 0$ unless $-1 < \bar{y}_i < 1$. That means in applying (b) that the resulting version of (3.22) comes down to $S(\bar{x},\bar{y})$ consisting of ξ such that $\nabla f_i(\bar{x}){\cdot}\xi = 0$ for i such that $-1 < \bar{y}_i < 1$. □

4 Criteria using generalized quadratic forms

In moving beyond polyhedral modeling, we lose piecewise linearity and have to take "curvature" into account. Realizations of the criterion for strong variational sufficiency in Theorem 3 have to go beyond the positive definiteness of $\nabla^2_{xx} L(\bar{x},\bar{y})$ on some subspace $S(\bar{x},\bar{y})$ as in Theorem 4.

Much of the foundation has already been laid for the steps to be taken next. The key is the relationship in (3.16) between the graph of the difference quotient mapping $\Delta_\tau \nabla g^r(u + r^{-1}y)$ in (3.14) and the graph of the difference quotient mapping $\Delta_\tau \partial g(u \,|\, y)$ in (3.15). The way forward in utilizing

Theorem 3 requires us to identify when the mappings $\Delta_\tau \nabla g^r(u+r^{-1}y)$ converge locally uniformly as $\tau \searrow 0$ to a linear mapping — the meaning of ∇g^r being differentiable at $u+r^{-1}y$ — and furthermore to understand how that linear mapping in the limit relates to the function g. On the basis of (3.16), this translates into figuring out when the difference quotient mappings $\Delta_\tau \partial g(u \,|\, y)$ converge graphically at $\tau \searrow 0$ to some mapping that is *generalized linear* in the sense that its graph is an m-dimensional *subspace* of $\mathbb{R}^m \times \mathbb{R}^m$. However, we also know that graphical convergence of the mappings $\Delta_\tau \partial g(u \,|\, y)$ is tied to epigraphical convergence of the second-order difference quotients in (3.17) through Attouch's Theorem [23, p. 552].

In the second-order theory of variational analysis developed in [23, Chapter 13], g is called *twice epi-differentiable* at u for a subgradient $y \in \partial g(u)$ if $\Delta_\tau^2 g(u \,|\, y)$ converges epigraphically to some limit as $\tau \searrow 0$. The limit function, called the *second-order subderivative*, is denoted by $d^2 g(u \,|\, y)$. Like $\Delta_\tau^2 g(u \,|\, y)$, it is convex and positively homogeneous of degree 2, i.e., has $d^2 g(u \,|\, y)(\tau\omega) = \tau^2 d^2 g(u \,|\, y)(\omega)$ when $\tau > 0$. The mappings $\Delta_\tau g(u \,|\, y)$ then converge accordingly in graph to the subgradient mapping $\partial[\frac{1}{2} d^2 g(u \,|\, y)]$. Putting this together with the facts already mentioned, we get that

$$
\begin{array}{l}
\Delta_\tau \nabla g^r(u+r^{-1}y) \text{ converges locally uniformly as } \tau \searrow 0 \text{ to a mapping } M \\
\iff \quad \left[\begin{array}{l} g \text{ is twice epi-differentiable at } u \text{ for } y \text{ and } M = \nabla q^r(u+r^{-1}y), \text{ where} \\ q^r(\omega) = \min_{\omega'}\{\, q(\omega') + \frac{r}{2}|\omega' - \omega|^2 \} \;\text{ in the case of }\; q(\omega) = \frac{1}{2} d^2 g(u \,|\, y)(\omega), \end{array} \right.
\end{array}
\tag{4.1}
$$

and furthermore

$$
M \text{ in (4.1) is linear} \quad \iff \quad \partial[\tfrac{1}{2} d^2 g(u \,|\, y)] \text{ is generalized linear.} \tag{4.2}
$$

This brings in the following concepts, where we build on the idea that the differentiable functions q on $\mathbb{R}^m$ for which the gradient mapping is a linear transformation are the quadratic functions without a linear or constant term. We will call them *quadratic forms* for simplicity.

Definition (generalized quadratic forms and generalized second-order differentiability)**.** *By a generalized quadratic form on $\mathbb{R}^m$ will be meant a function $q : \mathbb{R}^m \to (-\infty, \infty]$ with $q(0) = 0$ for which the subgradient mapping $\partial q : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is generalized linear. A function g on $\mathbb{R}^m$ will be called generalized twice differentiable at u for a subgradient y if it is twice epi-differentiable at u for y with the second-order subderivative $d^2 g(u \,|\, y)$ being a generalized quadratic form q.*

In broad terms, a generalized linear mapping A, say from $\mathbb{R}^n$ to $\mathbb{R}^m$, is a set-valued mapping for which $\operatorname{gph} A$ is a subspace of $\mathbb{R}^n \times \mathbb{R}^m$. This comes down to meaning that $\operatorname{dom} A$ is a subspace S of $\mathbb{R}^n$ and $A(0)$ is a subspace S' of $\mathbb{R}^m$, and there is an ordinary linear mapping $A_0 : S \to \mathbb{R}^m$ such that $A(x) = A_0(x) + S'$ for $x \in S$. In applying that description to the subgradient mapping ∂q for a function q on $\mathbb{R}^m$ with $q(0) = 0$, the subspace $S = \operatorname{dom} \partial q$ has to be $\operatorname{dom} q$. The subspace $S' = \partial q(0)$ is then $S^\perp$. This makes clear that generalized quadratic forms q are the functions representable as

$$
q(\omega) = \tfrac{1}{2}\omega \cdot Q\omega + \delta_S(\omega) \;\text{ for a subspace } S \text{ and a symmetric matrix } Q \text{ with } P_S Q P_S = Q, \tag{4.3}
$$

where P_S gives the projection onto S. The condition $P_S Q P_S = Q$ causes $\omega \cdot Q\omega$ to depend only on $P_S\omega$ and makes Q be uniquely determined by q and S as $\operatorname{dom} q$. Then the subgradient mapping, given by

$$
\partial q(\omega) = \begin{cases} Q\omega + S^\perp \;\text{ when }\; \omega \in S, \\ \emptyset \;\text{ otherwise,} \end{cases} \tag{4.4}
$$

is the corresponding generalized linear mapping. Note that the conjugate q^* of a convex generalized quadratic form q is another generalized quadratic form, and therefore from (3.18),

$$
\begin{array}{l}
g \text{ is generalized twice differentiable at } u \text{ for } y \text{ with } \frac{1}{2} d^2 g(u \,|\, y) = q \\
\quad \iff \quad g^* \text{ is generalized twice differentiable at } y \text{ for } u \text{ with } \frac{1}{2} d^2 g^*(y \,|\, u) = q^*.
\end{array}
\tag{4.5}
$$

There are different ways of thinking about how a finite function g that is only differentiable almost everywhere might nontheless be deemed twice differentiable at a point u. One can look to a second-order expansion of g, or a first-order expansion of ∂g, or a limit of a second-order difference quotients, and so forth. But when g is convex, they all come together as meaning that g is twice epi-differentiable at u for y being $\nabla g(u)$, and $d^2 g(u \,|\, y)$ being a quadratic form [23, 13.42]. This is the natural background for our definition of generalized twice differentiability.

To understand the generalization better through an easily visualized example, consider the finite convex function

$$g(u_1, u_2) = \max\Big\{ \tfrac{1}{2}u_1^2, \tfrac{1}{2}u_2^2 \Big\} \text{ for } (u_1, u_2) \in I\!R^2, \tag{4.6}$$

which is $\mathcal{C}^2$ except at points (u_1, u_2) having $|u_1| = |u_2|$. At the origin, g is merely differentiable with $\nabla g(0,0) = (0,0)$. At points $(u_1, u_2) = (t,t)$ other than the origin, $\partial g(t,t)$ is the line segment between $(t,0)$ and $(0,t)$. Hence, for any $\lambda \in (0,1)$, the graph of ∂g around (u,y) for $u = (t,t)$ and $y = (1-\lambda)(t,0) + \lambda(0,t)$ consists of the pairs (u', y') having $u' = u + (\tau,\tau)$ and $y' = y + (\sigma, -\sigma)$ for τ and σ sufficiently small. The vectors $(\tau,\tau,\sigma,-\sigma)$ in $I\!R^4$, regardless of size, comprise the graph of a generalized linear mapping from $I\!R^2$ to $I\!R^2$. This mapping is the subdifferential of the convex function q on $I\!R^2$ given by

$$q(\omega_1, \omega_2) = \begin{cases} \tfrac{1}{2}\tau^2 & \text{when } (\omega_1,\omega_2) = (\tau,\tau) \\ \infty & \text{otherwise} \end{cases}, \tag{4.7}$$

which is a generalized quadratic form. This tells us that g is generalized twice differentiable at u for y with this q as $\frac{1}{2}d^2 g(u \,|\, y)$. Note that generalized twice differentiability would not be obtained if the subgradient $y = (1-\lambda)(t,0) + \lambda(0,t)$ were taken with $\lambda = 0$ or $\lambda = 1$, instead of $\lambda \in (0,1)$. The analysis of generalized twice differentiability is much the same at points $u = (t,-t)$.

In the terminology of generalized quadratic forms and generalized twice differentiability, we can recast (4.1) and (4.2) in the following way:

$$\begin{array}{l} g^r \text{ is twice-continuously differentiable at } u + r^{-1}y \text{ if and only if } g \text{ is generalized} \\ \text{twice differentiable at } u \text{ for } y\text{, and then } \nabla^2 g^r(u + r^{-1}y) \text{ is the constant Hessian} \\ G \text{ associated with the quadratic form } q^r \text{ generated from } q = \tfrac{1}{2}d^2 g(u \,|\, y) \text{ by} \\ \qquad\qquad q^r(\omega) = \min_{\omega'}\Big\{ q(\omega') + \tfrac{r}{2}|\omega' - \omega|^2 \Big\}. \end{array} \tag{4.8}$$

However, for the sake of applying Theorem 3, we require not only the Hessians of g^r at various points $u + r^{-1}y$, but also their possible limits for sequences of points $u_k + r^{-1}y_k$ approaching $F(\bar{x}) + r^{-1}\bar{y}$, which form the set $\overline{\nabla}^2 g^r(F(\bar{x}) + r^{-1}\bar{y})$. This leads us to define the *quadratic bundle* of g at u for y by

$$\operatorname{quad} g(u \,|\, y) = \left[\begin{array}{l} \text{the collection of generalized quadratic forms } q \text{ for which} \\ \exists\, (u_k, y_k) \to (u,y) \text{ with } g \text{ generalized twice differentiable} \\ \text{at } u_k \text{ for } y_k \text{ and such that the generalized quadratic} \\ \text{forms } \; q_k = \tfrac{1}{2}d^2 g(u_k \,|\, y_k) \; \text{ converge epigraphically to } q, \end{array} \right. \tag{4.9}$$

where the duality in (4.5) implies

$$q \in \operatorname{quad} g(u \,|\, y) \iff q^* \in \operatorname{quad} g^*(y \,|\, u). \tag{4.10}$$

Theorem 5 (strong variational sufficiency through generalized quadratic forms)**.** *In the $\mathcal{C}^2$ case of (P) with $\bar{x}$ and $\bar{y}$ satisfying the first-order conditions, the strong variational sufficient condition for local optimality holds if and only if*

$$\textit{every } \; q \in \operatorname{quad} g(F(\bar{x}) \,|\, \bar{y}) \; \textit{ has } \; \tfrac{1}{2}\xi \cdot \nabla^2_{xx} L(\bar{x}, \bar{y})\xi + q(\nabla F(\bar{x})\xi) > 0 \; \textit{ when } \; \xi \neq 0. \tag{4.11}$$

Proof. We know now that the matrices $G \in \overline{\nabla}^2 g^r(F(\bar{x}) + r^{-1}\bar{y})$ in Theorem 3 correspond to the generalized quadratic forms $q \in \operatorname{quad} g(F(\bar{x})\,|\,\bar{y})$ in having $\frac{1}{2}\omega{\cdot}G\omega = q^r(\omega)$, the function defined in (4.8). The criterion for strong variational sufficiency in Theorem 3 can be restated then as the existence of high-enough r such that

$$\text{every } \ q \in \operatorname{quad} g(F(\bar{x})\,|\,\bar{y}) \ \text{ has } \ \tfrac{1}{2}\xi{\cdot}\nabla^2_{xx}L(\bar{x},\bar{y})\xi + q^r\Big(\nabla F(\bar{x})\xi\Big) > 0 \ \text{ if } \ \xi \neq 0. \tag{4.12}$$

For each q, let $\mu_r(q)$ denote the minimum attained by the quadratic form in (4.12) with respect to ξ lying on the unit sphere of $I\!R^m$. The desired property in (4.12) is that $\mu_r(q) > 0$. The nonnegative convex quadratic forms q^r increase pointwise as r increases, and in that way these functions converge both pointwise and epigraphically to q itself [23, 7.4(d)]. Then, through the positive homogeneity of degree 2, the functions obtained by adding the indicator of the unit sphere to the quadratic forms in (4.12) likewise epi-converge to the sum of that indicator and the generalized quadratic form in (4.11). The values $\mu_r(q)$ converge in that case to $\mu(q)$, the corresponding minimum for the generalized quadratic form in (4.11), because epi-convergence of lsc functions with uniformly bounded level sets makes their minimum values converge [23, 7.33]. Thus, the property in (4.12) holds for a particular q and high-enough r if and only if $\mu(q) > 0$.

In fact, $\mu_r(q)$ is also continuous with respect to epi-convergence of the generalized quadratic forms q [23, 7.33], and the bundle $\operatorname{quad} g(F(\bar{x})\,|\,\bar{y})$ of such forms is compact with respect to that topology. The bundle's compactness is based on the correspondence of its elements with those of the compact set $\overline{\nabla}^2 g^r(F(\bar{x}) + r^{-1}\bar{y})$ under a transformation that pairs epi-convergence on one side with matrix convergence on the other. Hence the desired existence of r such that $\mu_r(q) > 0$ for all q in that set corresponds to having $\mu(q) > 0$ for all q in that bundle, which is the property formulated in (4.11). □

In the criterion of Theorem 5, it is only necessary to test the forms q that are *minimal* in $\operatorname{quad} g(F(\bar{x})\,|\,\bar{y})$, in the sense that there is no form $q' \neq q$ in $\operatorname{quad} g(F(\bar{x})\,|\,\bar{y})$ with $q' \leq q$, $q' \neq q$. Every element of $\operatorname{quad} g(F(\bar{x})\,|\,\bar{y})$ has a minimul q below it, as can be seen by considering totally ordered subsets of the quadratic bundle and drawing on compactness with respect to epi-convergence.

Question 1. *Does the criterion in Theorem 5 really need sometimes to invoke more than one of the generalized quadratic forms $q \in \operatorname{quad} g(F(\bar{x})\,|\,\bar{y})$. In other words, could more than one such q be minimal, and could that make a difference?* **Answer.** *Yes.*

Detail. An example is furnished by the function g on $I\!R^2$ in (4.6) at $(0,0)$, where it is differentiable with gradient $(0,0)$. We examine how the condition (4.12) operates with that g when $F(\bar{x}) = (0,0)$ and $\bar{y} = (0,0)$. For this we must determine the bundle $\operatorname{quad} g(0,0\,|\,0,0)$ by way of the definition (4.9). The first step is identifying when g is generalized twice-differentiable at u for y.

In the region where $|u_1| > |u_2|$, g is $\mathcal{C}^2$ with the quadratic form being $q_1(\omega_1,\omega_2) = \frac{1}{2}\omega_1^2$, while the region where $|u_1] < |u_2|$ similarly yields $q_2(\omega_1,\omega_2) = \frac{1}{2}\omega_2^2$. Along the line where $\omega_1 = \omega_2$, which is a one-dimensional subspace S, we get, away from the origin, the form in (4.7). We can denote it here by q_3 and regard it as the restriction to S of q_1 or equivalently q_2, since the two agree on S. Likewise, along the line where $\omega_1 = -\omega_2$, which is the subspace $S^\perp$, we get q_4 the restriction to that subspace of q_1 or q_2. These are the only possibilities for generalized twice differentiability. Taking limits here is trivial, so $\operatorname{quad} g(0,0\,|\,0,0) = \{q_1,q_2,q_3,q_4\}$. But both q_3 and q_4 are $\geq$ both q_1 and q_2, so that the positive-definiteness condition in (3.48) needs only to be checked for q_1 and q_2.

The example is concerned with $F(\bar{x}) = (f_1(\bar{x}), f_2(\bar{x})) = (0,0)$ and $\bar{y} = (0,0)$, hence $\nabla^2_{xx}L(\bar{x},\bar{y}) = \nabla^2 f_0(\bar{x})$. Thus, in the case of q_1 the positive-definiteness condition requires

$$\xi{\cdot}\nabla^2 f_0(\bar{x})\xi + |\nabla f_1(\bar{x}){\cdot}\xi|^2 > 0 \ \text{ when } \ \xi \neq 0, \tag{4.13}$$

whereas for q_2 it requires

$$\xi{\cdot}\nabla^2 f_0(\bar{x})\xi + |\nabla f_2(\bar{x}){\cdot}\xi|^2 > 0 \ \text{ when } \ \xi \neq 0. \tag{4.14}$$

In general, neither of these conditions is can subsume the other. They are needed in tandem. □

Question 2. *When g, although not generalized twice differentiable, is twice epi-differentiable at $F(\bar{x})$ for $\bar{y}$ with second-order subderivative $d^2 g(F(\bar{x})\,|\,\bar{y})$, is the criterion in Theorem 5 equivalent perhaps to the single condition*

$$\xi{\cdot}\nabla^2_{xx} L(\bar{x},\bar{y})\xi + d^2 g(F(\bar{x})\,|\,\bar{y})\Big(\nabla F(\bar{x})\xi\Big) > 0 \ \text{ if } \xi \neq 0, \tag{4.15}$$

despite $d^2 g(F(\bar{x})\,|\,\bar{y})$ not being a generalized quadratic form? **Answer.** *No, the single condition (4.15) can be weaker than (4.11) and not enough to trigger strong variational sufficiency.*

Detail. This can be established by a continuation of the same example as in the preceding question, but specializing to $x \in I\!R^2$, $f_1(x_1,x_2) = x_1$, $f_2(x_1,x_2) = x_2$, and

$$f_0(x_1,x_2) = \frac{1}{4}(1+a)(x_1+x_2)^2 + \frac{1}{4}(1-a)(x_1+x_2)^2, \ \text{ with } \ \nabla^2 f_0(x_1,x_2) \equiv \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}. \tag{4.16}$$

Then $\begin{bmatrix} 2 & a \\ a & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & a \\ a & 2 \end{bmatrix}$ are the Hessians for the quadratic forms in (4.13) and (4.14), so these are positive-definite as long as their determinant values, both equal to $2-a^2$, are positive, i.e., as long as $a < \sqrt{2}$. Therefore, that condition on a corresponds to the criterion in (4.11) being fulfilled.

The criterion in (4.15) has the simplification here that $\frac{1}{2}d^2 g(0,0\,|\,0,0) = g$, because the g in question is itself positively homogeneous of degree 2 with $\nabla g(0,0) = (0,0)$. Other simplifications are that $\nabla F(\bar{x})\xi = (\xi_1,\xi_2)$ and $\frac{1}{2}\xi{\cdot}\nabla^2_{xx} L(\bar{x},\bar{y})\xi = f_0(\xi_1,\xi_2)$. The issue then in (4.15) is whether

$$(\xi_1,\xi_2) \neq (0,0) \quad \Longrightarrow \quad 0 < f_0(\xi_1,\xi_2) + g(\xi_1,\xi_2) = \max\{f_0(\xi_1,\xi_2) + \tfrac{1}{2}\xi_1^2,\, f_0(\xi_1,\xi_2) + \tfrac{1}{2}\xi_2^2\}.$$

For that to be true, we have to have, away from the $(\xi_1,\xi_2) = (0,0)$, that $f_0(\xi_1,\xi_2) + \frac{1}{2}x_1^2 > 0$ in the "east-west" quadrants where $|x_1| \geq |x_2|$, as well as $f_0(\xi_1,\xi_2) + \frac{1}{2}x_2^2 > 0$ in the "north-south" quadrants where $|x_2| \geq |x_1|$. Both come down to behavior along the line $\xi_1 = -\xi_2$, where the two functions agree and, in terms $(\xi_1,\xi_2) = (t,-t)$ reduce according to (4.16) to $\frac{1}{4}(1-a)(2t)^2 + \frac{1}{2}t^2 = (\frac{3}{2}-a)t^2$. Thus, the criterion in (4.15) will be satisfied if and only if $a < \frac{3}{2}$. It follows that, for a between $\sqrt{2}$ and $\frac{3}{2}$, (4.15) holds without (4.11). In other words, (4.11) demands more than (4.15). □

Second-order optimality conditions of the kind in (4.15) have recently been investigated very deeply by Mohammadi et al. [7] in the case of $g = \delta_K$ for a closed convex set K. The negative answer to Question 2 helps to delineate the difference between the results there and ours here, focused together on augmented Lagrangians. The results are complementary. The theme in [7] is second-derivative criteria corresponding to the standard quadratic growth condition on the objective function at optimality, whereas our theme is second-derivative criteria corresponding to strong variational sufficiency — which has been indentified in Section 1 with the parametrically extended form of the quadratic growth condition in (2.4). The results here also cover a wider range of problems, with g not just of the "geometric" form δ_K.

The example of g in (4.6) that has helped in resolving Questions 1 and 2 fits into the larger class, beyond g being polyhedral convex, or equivalently piecewise linear, in which g is *piecewise linear-quadratic.* This term means that $\operatorname{dom} g$ is the union of finitely many polyhedral convex sets, on each of which g can be expressed as a polynomial function of degree at most 2. Such functions on $I\!R^m$ are characterized by having the graph of ∂g be the union of finitely many polyhedral convex sets of dimension m in $I\!R^m \times I\!R^m$, [23, 12.30]. That property generalizes piecewise linearity of ∂g and corresponds to having the Lipschitz continuous gradient mapping ∇g^r be piecewise linear. Then there are only finitely many generalized quadratic forms q in $\operatorname{quad} g(u\,|\,y)$, one for the interior of each m-dimensional "cell" in $\operatorname{gph} \partial g$ to which (u, y) belongs.

Of course, $\operatorname{quad} g(u\,|\,y)$ can consist of finitely many forms q, or even just one, without g having to be piecewise linear-quadratic. Here is an illustration.

Example 3 (indicators of level sets and the Lorenz cone/second-order cone)**.**

(a) *Suppose $g = \delta_K$, and let $\mathcal{U}$ be an open set where K has a local representation as a level set:*

$$K \cap \mathcal{U} = \{\, u \in \mathcal{U} \mid h(u) \leq 0 \} \;\; \textit{for a } \mathcal{C}^2 \textit{ function } h \textit{ on } \mathcal{U}. \tag{4.17}$$

Let $u \in K \cap \mathcal{U}$ be a point where $h(u) = 0$ and $\nabla h(u) \neq 0$, so that $\partial g(u) = N_K(u) = \{\, \lambda \nabla h(u) \mid \lambda \geq 0 \}$. Then g is generalized twice differentiable at u for any nonzero $y \in \partial g(u)$, and the quadratic bundle $\operatorname{quad} g(u\,|\,y)$ *consists solely of the generalized quadratic form*

$$q = \tfrac{1}{2} d^2 g(u\,|\,y) \;\; \textit{with} \;\; q(\omega) = \begin{cases} \tfrac{1}{2}\omega \cdot \nabla^2 h(u)\omega & \textit{if } \nabla h(u)\omega = 0, \\ \infty & \textit{otherwise.} \end{cases} \tag{4.18}$$

(b) *The Lorenz cone/second-order cone in $I\!R^m$ is the closed convex cone*

$$\begin{aligned} K &= \{\, (u_1, u_2, \ldots, u_m) \mid u_1 \geq |(u_2, \ldots, u_m)\} \\ &= \{\, u \mid h(u) \leq 0 \} \;\; \textit{for} \;\; h(u_1, u_2, \ldots, u_m) = -u_1 + |(u_2, \ldots, u_m)|. \end{aligned} \tag{4.19}$$

This fits the prescription in (a) at boundary points away from the origin, thus offering a specific instance of a nonpolyhedral cone of importance in numerical optimization in which $\operatorname{quad} g(u\,|\,y)$ *typically consists of just a single q.*

(c) *For problem (P) with g as in (a) or (b) serving to model the constraint $F(x) \in K$, the condition in (4.11) reduces, when $h(F(\bar{x})) = 0$ and $\nabla h(F(\bar{x})) \neq 0$, to invoking a single q of form (4.18), the sole element of* $\operatorname{quad} g(F(\bar{x})\,|\,\bar{y})$*. That sufficient condition for local optimality corresponds to strong variational sufficiency, ensuring the augmented tilt stability property in Theorem 2.*

Detail. (a) This is a highly specialized case of a broader rule in [23, 13.17] for sets defined by constraint systems. That rule applies here to δ_K by interpreting it as the composite function $\delta_{I\!R_-} \circ h$. (b) This extends the particular case of (a) by elementary calculation. (c) This just reviews the consequences from Theorem 5 and Theorem 2. □

In further analysis of the Lorenz cone, looking at its apex at $u = 0$, the vectors $y \in \partial g(0) = N_K(0)$ have $y_1 = -1$ and $|(y_2, \ldots, y_m)| \leq 1$. For such y having the inequality strict, there is again generalized twice differentiability of δ_K and again a unique q which is δ_S for the subspace S consisting of the vectors $(\omega_1, \omega_2, \ldots, \omega_m)$ such that $\omega_1 = 0$. On the other hand for y with $y_1 = -1$ and $|(y_2, \ldots, y_m)| = 1$, both that q and the earlier one comprise the bundle $\operatorname{quad} g(0\,|\,y)$.

Problems (P) in which $g = \delta_K$ for the Lorenz cone K in (4.19) are called *second-order cone programming*. They are rich in applications and have been the subject of much research, cf. [1], [3], [10]. The formula for the second-order subderivative for this case in Example 3(b) is not itself new, nor is the sufficiency of the resulting optimality condition in (4.11) — see [5, Prop. 2.1]. What is new is the placement of this formula and optimality condition in the framework of quadratic bundles and the criterion for strong variational sufficiency in Theorem 5, as brought out in Example 3(c). In contrast, recent work in [6] only derives from that special case of (4.11) a standard type of quadratic growth property, not the parametrically extended quadratic growth property in (2.4) that corresponds to strong variational sufficiency.

Beyond situations where the quadratic bundle $\operatorname{quad} g(F(\bar{x}) \,|\, \bar{y})$ in Theorem 5 may only consist of a single q, as illustrated in Example 3, there are others where $\operatorname{quad} g(F(\bar{x}) \,|\, \bar{y})$ can be comprised of a multiplicity of generalized quadratic forms, but a unique one of them is minimal. Then (4.11) would only need to be tested for that one q. The next example demonstrates this possibility in building on Example 2(b) and utilizing the face concept in (3.31).

Example 4 (a case of quadratic bundles having a single minimal element)**.** *Suppose for a polyhedral convex set Y and a $\mathcal{C}^2$ convex function k on $\mathbb{R}^m$ that g is the convex function conjugate to $g^* = \delta_Y + k$, so that $g(u) = \sup_{y \in Y} \{ y \cdot u - k(y) \}$ and*

$$y \in \partial g(u) \iff u - \nabla k(y) \in N_Y(y) \iff y \in \operatorname{face}(Y \,|\, u - \nabla k(y)), \tag{4.20}$$

Then g is twice epi-differentiable at u for every $y \in \partial g(u)$, with $\frac{1}{2} d^2 g(u \,|\, y)$ being the function $h^*_{y,u}$ that is conjugate to the convex function

$$h_{y,u}(\eta) = \delta_{T(y \,|\, u)}(\eta) + \tfrac{1}{2} \eta \cdot \nabla^2 k(y) \eta, \tag{4.21}$$

where

$$T(y \,|\, u) \text{ is the tangent cone at } y \text{ to } \operatorname{face}(Y \,|\, u - \nabla k(y)). \tag{4.22}$$

In other words,

$$\tfrac{1}{2} d^2 g(u \,|\, y)(\omega) = \sup_{\eta \in T(y \,|\, u)} \{ \omega \cdot \eta - \tfrac{1}{2} \eta \cdot \nabla^2 k(y) \eta \}. \tag{4.23}$$

Generalized twice differentiability of g corresponds in this to the polyhedral convex cone $T(y \,|\, u)$ being a subpace, which holds if and only if $u - \nabla k(y) \in \operatorname{ri} N_Y(y)$.

Because Y has only finitely many faces, there are only finitely many generalized quadratic forms q arising as conjugates $h^*_{y,u}$ with $T(y \,|\, u)$ being a subspace. Any $(\bar{u}, \bar{y}) \in \operatorname{gph} \partial g$ has a neighborhood such that the bundle $\operatorname{quad} g(\bar{u} \,|\, \bar{y})$ is comprised of such q coming from (u, y) in that neighborhood. But $\operatorname{quad} g(\bar{u} \,|\, \bar{y})$ contains a single minimal element q among them, namely the generalized quadratic form $h^*_{\bar{y},u}$ coming from the points u for which $u - \nabla k(\bar{y}) \in \operatorname{ri} N_Y(\bar{y})$.

Thus, in the $\mathcal{C}^2$ case of (P) with the modeling function g being of the category described here, the criterion for strong variational sufficiency in Theorem 5 only requires checking for that single q.

Detail. In fact, all the assertions made about g translate to assertions about g^*, inasmuch as generalized twice differentiability dualizes as in (4.5), and the same for twice epi-differentiability, again due to (3.18), [23, 13.21]. It is easy to calculate directly from convergence of second-order difference quotients that g^* is twice epi-differentiable at y for any $u \in \partial g^*(y)$ with $\frac{1}{2} d^2 g^*(y \,|\, u)$ being the function $h_{y,u}$ in (4.21). Polyhedrality of Y leads to having only finitely many generalized quadratic forms arise as such functions. Then $\operatorname{quad} g^*(\bar{y} \,|\, \bar{u})$ is comprised of the ones coming from $(y, u) \in \operatorname{gph} \partial g^*$ in a small enough neighborhood of $(\bar{y}, \bar{u})$. That translates through (4.10) into the corresponding claim about $\operatorname{quad} g(\bar{u} \,|\, \bar{y})$. Because conjugacy reverses dominance of one convex function over another, the existence of a single minimal element in $\operatorname{quad} g(\bar{u} \,|\, \bar{y})$ corresponds to the existence of a single maximal element in $\operatorname{quad} g^*(\bar{y} \,|\, \bar{u})$. The latter existence is established through the fact that having $u - \nabla k(\bar{y}) \in \operatorname{ri} N_Y(y)$ is equivalent to having $\bar{y} \in \operatorname{ri} \operatorname{face}(Y \,|\, u - \nabla k(\bar{y}))$, with $\operatorname{face}(Y \,|\, u - \nabla k(\bar{y}))$ then being the smallest face of Y containing $\bar{y}$. Then $T(\bar{y}, u)$ is a subspace included in all the other nearby subspaces $T(y, u)$. $\square$

The case of Example 4 in which $k(y) \equiv 0$ can be seen as coinciding with Example 2(b). That previous example relied on Theorem 4, but here an alternative argument via Theorem 5 is provided.

Problem (P) with the modeling function g being of the kind in Example 4 is *extended nonlinear programming* in the terminology of [18], [23, p. 506]. However, those problem models generally also allow for a term $\delta_X(x)$ in the objective, which corresponds to adding a constraint $x \in X$ to (1.1). The Lagrangian $l(x, y)$ comes out then nicely as the restriction of $f_0(x) + y_1 f_1(x) + \ldots + y_m f_m(x) - k(y)$ to $X \times Y$.

Concluding remark. There is little doubt that the results in this paper could be broadened to incorporate the extra feature in (P) of a geometric constraint $x \in X$, drawing ultimately on curvature aspects of X captured by second-order epi-derivatives of δ_X. That was not carried out for reasons of exposition. In an attempt to accommodate a $\delta_X(x)$ term, everything would get complicated for many readers all too soon — to the point where the basic message about this topic being a natural generalization of nonlinear programming theory could get lost. Such an extension was therefore relegated to a potential future effort.

References

[1] Alizadeh,f., Goldfarb, D., "Second-order cone programming," *Mathematical Programming* **95** (2003), 3-51.

[2] Bertsekas, D. P., *Constrained Optimization and Lagrange Multiplier Methods*, Aademic Press, 1982.

[3] Bonnans, J.F., Ramirez, H.C., "Perturbation analysis of second-order cone programming problems," *Mathematical Programming* **104** (2005), 205–227.

[4] Dontchev, A.D., Rockafellar, R.T., *Implicit Functions and Solution Mappings: a View From Variational Analysis*, Springer-Verlag, 2009; second edition 2014.

[5] Hang, N.V.T., Mordukhovich, B.S., Sarabi, M.E., "Second-order variational analysis in second-order cone programming," *Mathematical Programming* **180** (2020), 930–946.

[6] Hang, N.V.T., Mordukhovich, B.S., Sarabi, M.E., "Augmented Lagrangian method for second-order cone programs under second-order sufficiency," *J. Global Optim.* (2021), doi.org/10.1007/s10898-021-0168-1.

[7] Mohammadi, A., Mordukhovich, B.S., Sarabi, M.E., "Parabolic regularity in geometric variational analysis," *Transactions Amer. Math. Soc.* **374** (2021), 1711–1763.

[8] Mordukhovich, B.S., "Sensitivity analysis in nonsmooth optimization." in *Theoretical Aspects of Industrial Design*, D.A. Field and V. Komkov, eds., SIAM Volumes in Applied Mathematics, no. 58, 1992, 32–46.

[9] Mordukhovich, B.S., "Lipschitzian stability of constraint systems and generalized equations," *Nonliner Analysis: Theory, Methods and Applications* **22** (1994), 173–206.

[10] Mordukhovich, B.S., Outrata, J.V., Sarabi, M.E., "Full stability of locally optimal solutions in second-rder cone programming," *SIAM J. Optimization* **24** (2014), 1581–1613.

[11] Pennanen, T., "Local convergence of the proximal point algorithm and multiplier methods without monotonicity," *Mathematics of Operations Research* **27** (2002), 170–191.

[12] Poliquin, R. A., "Subgradient monotonicity and convex functions," *Nonlinear Analysis: Theory, Methods and Applications* **14** (1990), 385–398

[13] Poliquin, R. A., Rockafellar, R. T., "Tilt stability of a local minimum," *SIAM J. Optimization* **8** (1998), 287–289.

[14] Rockafellar, R. T., *Convex Analysis*, Princeton University Press, 1970.

[15] Rockafellar, R. T., *Conjugate Duality and Optimization*, No. 16 in Conference Board of MathṠciences Series, SIAM Publications, 1974.

[16] Rockafellar, R. T., "Monotone operators and the proximal point algorithm," *SIAM J. Control Opt.* **14** (1976), 877–898.

[17] Rockafellar, R. T., "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," *Math. of Operations Research* **1** (1976), 97–116.

[18] Rockafellar, R. T., "Extended nonlinear programming," in *Nonlinear Optimization and Related Topics* (G. di Pillo and F. Giannessi, eds.), Kluwer, 1999, 381–399.

[19] Rockafellar, R. T., "Lagrange multipliers and optimality," *SIAM Review* **35** (1993), 183–238.

[20] Rockafellar, R. T., "Variational convexity and local monotonicity of subgradient mappings," *Vietnam J. Math.* **47** (2019), 547–561: dx.doi.org/10.1007/s10013-019-00339-5.

[21] Rockafellar, R. T., "Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity," *Set-Valued and Variational Analysis* **27** (2019), 863–893: dx.doi.org/10.1007/s1128-018-0496-1.

[22] Rockafellar, R. T., "Convergence of augmented Lagrangian methods in extensions beyond nonlinear programming," forthcoming.

[23] Rockafellar, R. T., Wets, R. J-B, *Variational Analysis*, No. 317 in the series *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, 1998.