Set-Valued Anal (2006) 14: 249–261
DOI 10.1007/s11228-006-0022-8

Supercalm Multifunctions For Convergence Analysis

Adam B. Levy

Received: 10 February 2005 / Accepted: 30 March 2006

Abstract Calmness of multifunctions is a well-studied concept of generalized continuity in which single-valued selections from the image sets of the multifunction exhibit a restricted type of local Lipschitz continuity where the base point is fixed as one point of comparison. Generalized continuity properties of multifunctions like calmness can be applied to convergence analysis when the multifunction appropriately represents the iterates generated by some algorithm. Since it involves an essentially linear relationship between input and output, calmness gives essentially linear convergence results when it is applied directly to convergence analysis. We introduce a new continuity concept called 'supercalmness' where arbitrarily small calmness constants can be obtained near the base point, which leads to essentially superlinear convergence results. We also explore partial supercalmness and use a well-known generalized derivative to characterize both when a multifunction is supercalm and when it is partially supercalm. To illustrate the value of such characterizations, we explore in detail a new example of a general primal sequential quadratic programming method for nonlinear programming and obtain verifiable conditions to ensure convergence at a superlinear rate.

Key words calmness · multifunctions · generalized continuity · variational analysis · convergence analysis · sequential quadratic programming.

AMS 2000 Subject Classification Primary: 49J53 · Secondary: 90C30.

A. B. Levy (✉)
Department of Mathematics, Bowdoin College,
Brunswick, ME 04011, USA
e-mail: alevy@bowdoin.edu

1. Introduction

A (single-valued) mapping $x\colon \mathcal{P} \to \mathcal{X}$ between normed linear spaces $\mathcal{P}$ and $\mathcal{X}$ is called *calm at* $\bar{p}$ if there is a constant $L > 0$ and a neighborhood $P \subseteq \mathcal{P}$ of $\bar{p}$ such that the function satisfies

$$\|x(p) - x(\bar{p})\| \leqslant L\|p - \bar{p}\| \quad \forall\, p \in P. \tag{1}$$

This property is evidently a restricted kind of local Lipschitz continuity where the base point $\bar{p}$ is fixed as one point of comparison. The term 'calmness' was originally used in [2] to describe this property for the optimal value function associated with an optimization problem, and it has since become a relatively well-established terminology to describe this property for more general mappings (see [19] in the case when the normed linear spaces are finite-dimensional). We say that a multifunction (i.e. set-valued mapping) $S\colon \mathcal{P} \rightrightarrows \mathcal{X}$ is *calm at* $\bar{p}$ *for* $\bar{x}$ if there exist neighborhoods $P \subseteq \mathcal{P}$ of $\bar{p}$ and $X \subseteq \mathcal{X}$ of $\bar{x}$ together with a constant $L > 0$ such that any local selection $x(p) \in S(p) \cap X$ satisfies

$$\|x(p) - \bar{x}\| \leqslant L\|p - \bar{p}\| \quad \forall\, p \in P. \tag{2}$$

A global version of this property was first introduced for multifunctions in [17] where it was called 'Lipschitz continuity at $\bar{p}$'. Notice that one consequence of calmness at $\bar{p}$ for $\bar{x}$ is that $\bar{x}$ is a locally isolated element of the image set $S(\bar{p})$ since $x(\bar{p}) = \bar{x}$ for all local selections. A more general concept without this local isolation was introduced by Robinson [16] and called the 'upper Lipschitzian' property (the same property was later labeled 'calmness at $\bar{p}$' in [19]). The exact same terminology as ours was used in [5] to denote almost the same property but without the local isolation inherent in our version, and our version essentially appears in that paper as the "strong metric subregularity of S^{-1} at $\bar{x}$ for $\bar{p}$". Moreover, it was shown in [5] that without local isolation, this property is not as useful for stability analysis, which consequently gives motivation for our definition of calmness. Our version of calmness was also studied in [13] where it was called 'selection calmness' as well as in other places under different labels: In [12] it was called 'local upper Lipschitz continuity'; in [9] it was called 'local Lipschitz upper semicontinuity'; in [3] it was called 'upper Lipschitz continuity at a point'; in [1] the calmness property was called 'semistability' in the context of variational inequalities and in [7, 14], and [15] it was called 'stability' when it was present with the additional property of nonempty local image sets $S(p) \cap X$.

Like most generalized continuity properties for multifunctions, calmness is useful for convergence analysis. For example, if the sequence of iterates $x_k \to \bar{x}$ generated by an algorithm can be represented via the inclusions $x_{k+1} \in S(x_k)$ in terms of a multifunction $S\colon \mathcal{X} \rightrightarrows \mathcal{X}$ that is calm at $\bar{x}$ for $\bar{x}$, then eventually we have the bound $\|x_{k+1} - \bar{x}\| \leqslant L\|x_k - \bar{x}\|$. Evidently, this implies the bound

$$R := \lim_{k\to\infty} \frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} \leqslant L \tag{3}$$

indicating a linear convergence rate of R as long as $L \in [0, 1)$. A superlinear convergence rate corresponds to $R = 0$, which can be ensured in the same way by the calmness of S if the calmness constant L can be assumed to be arbitrarily small by shrinking the neighborhood of $\bar{x}$ if necessary. Accordingly, we define a (single-valued) mapping $x\colon \mathcal{P} \to \mathcal{X}$ between normed linear spaces $\mathcal{P}$ and $\mathcal{X}$ to be *supercalm at* $\bar{p}$ if for every constant $L > 0$ there is a neighborhood $P \subseteq \mathcal{P}$ of $\bar{p}$ such that the function satisfies (1), and just as in the case of calmness, we extend this notion to multifunctions: We define a multifunction $S\colon \mathcal{P} \rightrightarrows \mathcal{X}$ to be *supercalm at* $\bar{p}$ *for* $\bar{x}$ if for every constant $L > 0$ there exist neighborhoods $P \subseteq \mathcal{P}$ of $\bar{p}$ and $X \subseteq \mathcal{X}$ of $\bar{x}$ such that any local selection $x(p) \in S(p) \cap X$ satisfies (2). Evidently, supercalmness gives the arbitrarily small constants L in (3) which ensures a superlinear convergence rate for the algorithm defined by $x_{k+1} \in S(x_k)$.

Another related concept that we develop here is 'partial supercalmness' for multifunctions whose domain variables can be split into two categories. A multifunction $S\colon \mathcal{P} \times \mathcal{Q} \rightrightarrows \mathcal{X}$ is said to be *partially supercalm with respect to* p *at* $(\bar{p}, \bar{q})$ *for* $\bar{x}$ if for every constant $L_p > 0$ there exists a constant $L_q > 0$ and neighborhoods $P \subseteq \mathcal{P}$ of $\bar{p}$, $Q \subseteq \mathcal{Q}$ of $\bar{q}$, and $X \subseteq \mathcal{X}$ of $\bar{x}$ such that any local selection $x(p, q) \in S(p, q) \cap X$ satisfies

$$\|x(p,q) - \bar{x}\| \leqslant L_p \|p - \bar{p}\| + L_q \|q - \bar{q}\| \quad \forall (p, q) \in P \times Q. \tag{4}$$

Notice that partial supercalmness with respect to p essentially entails supercalmness with respect to p and calmness with respect to q. The connection between partial supercalmness and convergence analysis is not as direct as in the cases of calmness and supercalmness, so we devote Section 2 to showing that partial supercalmness is actually the key property for ensuring a superlinear convergence rate for a new example of a general primal sequential quadratic programming method for nonlinear programming.

Having established that the supercalmness and partial supercalmness properties are useful for convergence analysis, we turn our attention in Section 3 to developing verifiable conditions for these properties. We need to do this because both of these properties might be difficult to verify directly since, among other issues, there might be an infinite number of local selections to check. This problem of direct verification is typical with generalized continuity properties and so alternate characterizations of other generalized continuity properties have already been developed. For instance, a characterization of multifunction calmness was given in [12] using a computable generalized derivative. In Section 3 we develop and prove similar characterizations for both multifunction supercalmness and partial supercalmness, and we apply our characterization of partial supercalmness to our primal sequential quadratic programming method to obtain verifiable conditions ensuring convergence at a superlinear rate.

2. Convergence of a Primal SQP Method via Partial Supercalmness

To illustrate how the partial supercalmness property is useful for convergence analysis, we present a generalization of the famous SQP method for solving nonlinear programs of the form

$$\min\ f_0(x) \text{ over } x \in X := \left\{x \in \mathbb{R}^n \text{ with } f_i(x) \leqslant 0 \text{ for } i \in \mathcal{I}, \text{ and } f_i(x) = 0 \text{ for } i \in \mathcal{E}\right\}$$

for smooth functions f_i: $\mathbb{R}^n \to \mathbb{R}$ and index sets $\mathcal{I}$ and $\mathcal{E}$ together containing m indices. Notice that if we define the set $K \subseteq \mathbb{R}^m$ by

$$K := \{z \in \mathbb{R}^m \text{ with } z_i \leqslant 0 \text{ for } i \in \mathcal{I} \text{ and } z_i = 0 \text{ for } i \in \mathcal{E}\},$$

then we can rewrite the nonlinear program in the equivalent (unconstrained) form

$$\min\ f_0(x) + \delta_K(g(x)) \text{ over } x \in \mathbb{R}^n$$

in terms of the 'indicator function'

$$\delta_K(z) := \begin{cases} 0 & \text{if } z \in K \\ \infty & \text{otherwise} \end{cases}$$

and the constraint mapping

$$g(x) := \big(f_1(x),\ f_2(x), \ldots,\ f_m(x)\big).$$

One key component of our generalization of the SQP method is a family of smooth functions g_u: $\mathbb{R}^n \to \mathbb{R}$ satisfying the following assumption: The mappings $(u, x) \mapsto g_u(x)$, $(u, x) \mapsto \nabla g_u(x)$, and $(u, x) \mapsto \nabla^2 g_u(x)$ are all continuous on some neighborhood $U \times X \subseteq \mathbb{R}^m \times \mathbb{R}^n$ of a target parameter–primal pair $(\bar{u}, \bar{x})$.

The Primal SQP Method: Given a current iterate x_k and parameter u_k, choose the next iterate $x_{k+1} = x'$ by solving the (unconstrained) approximation problem

$$\min\ f_0(x_k)+\nabla f_0(x_k)[x'-x_k]+1/2[x'-x_k]^T \cdot \nabla^2\big(f_0 + g_u\big)(x_k)[x'-x_k] \text{ over } x' \in L(x_k) \tag{5}$$

for the linearized constraint set

$$L(x) := \{x' \in \mathbb{R}^n | \nabla g(x)[x' - x] + g(x) \in K\}.$$

The standard choice for the functions g_u is

$$g_u = u \cdot g \text{ with } \big(u \cdot g\big)(x) := u^T \cdot g(x) \tag{6}$$

where the parameters u act as multipliers that are updated by adding the multiplier associated with the solution x' to the approximation problem. Many other interesting possibilities for the g_u are possible in our formulation. For instance, we could instead use functions of the form $g_u(x) := u\,\|g(x)\|^2$ for $u \in [0, \infty)$ which act as penalty functions in the case when all the constraints are equations (i.e. $\mathcal{I} = \emptyset$). Our primal SQP method not only allows more general functions g_u than is standard, but also allows more general parameter updating. The latter flexibility motivates the *primal* in the label for our method, since only an update on the primal iterates x is specified.

Our convergence theorem for the primal SQP method relies on the famous Mangasarian–Fromovitz constraint qualification at $\bar{x}$:

$$\text{The set of multipliers } \widetilde{Y} := \big\{\tilde{y} \in N_K\big(g(\bar{x})\big) \text{ with } \nabla f_0(\bar{x}) + \nabla g(\bar{x})^T \cdot \tilde{y} = 0\big\} \text{ is bounded,} \tag{7}$$

where the notation $N_K\big(g(\bar{x})\big)$ indicates the cone of normal vectors to the set K at $g(\bar{x})$. In the statement of our convergence theorem, we will also use the following notation to denote the set of vectors obtained by products of the transposed Jacobian $\nabla g(x)^T$ with vectors from the normal cone:

$$\nabla g(x)^T \cdot N_K\big(\nabla g(x)[x' - x] + g(x)\big) := \big\{\nabla g(x)^T \cdot y \,\big|\, y \in N_K\big(\nabla g(x)[x' - x] + g(x)\big)\big\}.$$

PROPOSITION 2.1 (superlinear convergence of the primal SQP method). *Under the Mangasarian–Fromovitz constraint qualification at $\bar{x}$, if the multifunction*

$$S(x, q) := \left\{ x' \in \mathbb{R}^n \,\middle|\, q \in \begin{array}{c} \nabla^2\big(f_0 + g_{\bar{u}}\big)(x)[x' - \bar{x}] + \nabla g_{\bar{u}}(\bar{x}) - \nabla g_{\bar{u}}(x) \\ +\nabla g(x)^T \cdot N_K\big(\nabla g(x)[x' - x] + g(x)\big) \end{array} \right\} \tag{8}$$

is partially supercalm with respect to x at $(\bar{x}, \bar{q})$ for $\bar{x}$ (where $\bar{q} := -\nabla f_0(\bar{x})$), then there exists a neighborhood $X \subseteq \mathbb{R}^n$ of $\bar{x}$ such that any sequence of iterates $\{x_k\} \subseteq X$ generated by the primal SQP method for a parameter sequence $u_k \to \bar{u}$ converges to $\bar{x}$ at a superlinear rate.

Proof. We choose any $L \in [0, 1)$, set $L_x := L/8$, and apply the partial supercalmness assumption to obtain $L_q > 0$. We recall the neighborhood $U \times X$ of $(\bar{u}, \bar{x})$ from the continuity assumptions on the g_u and shrink X if necessary to be contained in the two neighborhoods of $\bar{x}$ assured in this case by the partial supercalmness assumption (one neighborhood from the domain and one from the range). We shrink X further if necessary to ensure that the images of $-\nabla f_0(x)$ for all $x \in X$ are contained in the neighborhood Q of $\bar{q}$ assured by partial supercalmness.

Throughout the proof, we use x' to denote a typical new iterate $x_{k+1} \in X$ generated by the primal SQP method from the current parameter–primal pair denoted by $(u, x) \in U \times X$. Any such x' must satisfy the following necessary condition for optimality associated with the minimization problem (5) with (u, x) in place of (u_k, x_k):

$$-\nabla f_0(x) - \nabla^2 f_0(x)[x' - x] \in \partial n_{u,x}(x') \tag{9}$$

in terms of the set of subgradients $\partial n_{u,x}(x')$ at x' associated with the function $n_{u,x}$ defined by

$$n_{u,x}(x') := 1/2[x' - x]^T \cdot \nabla^2 g_u(x)[x' - x] + \delta_K\big(\nabla g(x)[x' - x] + g(x)\big)$$

From [19, Example 10.8], we know that under the Mangasarian–Fromovitz constraint qualification the subgradients of $n_{u,x}$ are given by

$$\partial n_{u,x}(x') = \nabla^2 g_u(x)[x' - x] + \nabla g(x)^T \cdot N_K\big(\nabla g(x)[x' - x] + g(x)\big)$$

in which case the necessary condition (9) becomes

$$-\nabla f_0(x) - \nabla^2 f_0(x)[x' - x] \in \nabla^2 g_u(x)[x' - x] + \nabla g(x)^T \cdot N_K\big(\nabla g(x)[x' - x] + g(x)\big),$$

which is the same as

$$-\nabla f_0(x)-\nabla^2\big(f_0+g_u\big)(x)[\bar{x}-x]+\nabla^2\big(g_{\bar{u}}-g_u\big)(x)[x'-\bar{x}]+\nabla g_{\bar{u}}(\bar{x})-\nabla g_{\bar{u}}(x)\in S_x(x') \tag{10}$$

for the (x-parameterized) family of multifunctions S_x: $\mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by

$$S_x(x'):=\nabla^2\big(f_0+g_{\bar{u}}\big)(x)[x'-\bar{x}]+\nabla g_{\bar{u}}(\bar{x})-\nabla g_{\bar{u}}(x)+\nabla g(x)^T\cdot N_K\big(\nabla g(x)[x'-x]+g(x)\big)$$

Since the multifunction S defined in (8) satisfies

$$S(x,q)=\left\{x'\in\mathbb{R}^n \,\middle|\, q\in S_x(x')\right\},$$

its assumed partial supercalmness and the inclusion (10) implies the following series of bounds:

$$\begin{aligned}
\|x'-\bar{x}\| \leqslant\ & L_x\,\|x-\bar{x}\| + L_q\,\|\nabla f_0(\bar{x})-\nabla f_0(x)-\nabla^2 f_0(x)[\bar{x}-x]\| \\
& +L_q\,\|-\nabla^2 g_u(x)[\bar{x}-x]+\nabla^2\big(g_{\bar{u}}-g_u\big)(x)[x'-\bar{x}]+\nabla g_{\bar{u}}(\bar{x})-\nabla g_{\bar{u}}(x)\| \\
\leqslant\ & L_x\,\|x-\bar{x}\| + L_q\,\|\nabla f_0(\bar{x})-\nabla f_0(x)-\nabla^2 f_0(x)[\bar{x}-x]\| \\
& +L_q\,\|\nabla g_u(\bar{x})-\nabla g_u(x)-\nabla^2 g_u(x)[\bar{x}-x]\| \\
& +L_q\,\|\nabla^2\big(g_{\bar{u}}-g_u\big)(x)\|\,\|[x'-\bar{x}]\| \\
& +L_q\,\|\nabla\big(g_{\bar{u}}-g_u\big)(\bar{x})-\nabla\big(g_{\bar{u}}-g_u\big)(x)\|
\end{aligned} \tag{11}$$

From the continuity assumption on $(u,x)\mapsto\nabla^2 g_u(x)$ we know that we can shrink $U\times X$ if necessary to have

$$L_q\,\|\nabla^2\big(g_{\bar{u}}-g_u\big)(x)\| \leqslant 1/2 \text{ for all } u\in U \text{ and } x\in X \tag{12}$$

so that the bounds (11) imply the bound

$$\begin{aligned}
\|x'-\bar{x}\| \leqslant\ & L/4\,\|x-\bar{x}\| + 2\,L_q\,\|\nabla f_0(\bar{x})-\nabla f_0(x)-\nabla^2 f_0(x)[\bar{x}-x]\| \\
& +2\,L_q\,\|\nabla g_u(\bar{x})-\nabla g_u(x)-\nabla^2 g_u(x)[\bar{x}-x]\| \\
& +2\,L_q\,\|\nabla\big(g_{\bar{u}}-g_u\big)(\bar{x})-\nabla\big(g_{\bar{u}}-g_u\big)(x)\|
\end{aligned} \tag{13}$$

where we have also applied the identity $L_x := L/8$.

Since f_0 is smooth, we know the linear Taylor approximation of ∇f_0 at x satisfies

$$\|\nabla f_0(\bar{x})-\nabla f_0(x)-\nabla^2 f_0(x)[\bar{x}-x]\| \leqslant \epsilon_1\|x-\bar{x}\|^2 \text{ for all } u\in U \text{ and } x\in X \tag{14}$$

for some constant $\epsilon_1>0$. The continuity assumptions on $(u,x)\mapsto\nabla g_u(x)$ and $(u,x)\mapsto\nabla^2 g_u(x)$ ensure uniform linear Taylor approximations of ∇g_u at x for some constant $\epsilon_2>0$:

$$\|\nabla g_u(\bar{x})-\nabla g_u(x)-\nabla^2 g_u(x)[\bar{x}-x]\| \leqslant \epsilon_2\|x-\bar{x}\|^2 \text{ for all } u\in U \text{ and } x\in X \tag{15}$$

as well as the uniform calmness of $\nabla\big(g_{\bar{u}}-g_u\big)$ at $\bar{x}$:

$$\|\nabla\big(g_{\bar{u}}-g_u\big)(x)-\nabla\big(g_{\bar{u}}-g_u\big)(\bar{x})\| \leqslant L/4\,\|x-\bar{x}\| \text{ for all } u\in U \text{ and } x\in X \tag{16}$$

where the constant $L/4$ can be assumed by shrinking U if necessary. Applying the combination of the bounds (13)–(16) with $\beta := 2\,L_q\,(\epsilon_1 + \epsilon_2)$, we get the first inequality in the series

$$\begin{aligned}\|x_{k+1} - \bar{x}\| &\leqslant L/2\,\|x_k - \bar{x}\| + \beta\,\|x_k - \bar{x}\|^2 \\ &= (L/2 + \beta\,\|x_k - \bar{x}\|)\;\|x_k - \bar{x}\| \\ &\leqslant L\,\|x_k - \bar{x}\| \end{aligned} \tag{17}$$

where the last inequality follows since $\beta\,\|x_k - \bar{x}\|$ can be made less than $L/2$ by shrinking X if necessary. The resulting X is the neighborhood stipulated in the statement of the theorem.

Since the sequence of parameters u_k is assumed to converge to $\bar{u}$, we know that eventually they satisfy $u_k \in U$ so the final inequality in (17) eventually applies to our sequence of iterates $\{x_k\}$. Since L was chosen in $[0, 1)$, this implies that $x_k \to \bar{x}$ as claimed. To prove the superlinear convergence rate, we notice that any $\widetilde{L} \in [0, 1)$ produces, by the same argument above, a corresponding neighborhood $\widetilde{U} \times \widetilde{X} \subseteq \mathcal{U} \times \mathcal{X}$ of $(\bar{u}, \bar{x})$ for which the analog of (17) holds. As a result we have the bound

$$\frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} \leqslant \widetilde{L} \tag{18}$$

as long as $u_k \in \widetilde{U}$ and $x_{k+1}, x_k \in \widetilde{X}$. Since $u_k \to \bar{u}$ and we have already established that $x_k \to \bar{x}$, we conclude that the bound (18) eventually holds for our sequence of iterates $\{x_k\}$. The superlinear convergence rate follows since $\widetilde{L} \in [0, 1)$ is arbitrary. □

Remarks. At first glance, the result of Proposition 2.1 might seem underwhelming since it is well known that the usual SQP method exhibits quadratic convergence. However, there is an important distinction here between the usual SQP method and the *primal* SQP method defined by (5) and covered by Proposition 2.1. For one, the usual SQP method always uses the standard choice of functions $g_u = u \cdot g$ in (6). Moreover, the usual SQP method involves simultaneous iteration of the parameter–primal pairs (u, x), whereas the primal SQP method iterates the primal variables x only, while allowing completely general updating of the parameters u. Of course, one option for updating the parameters in the primal SQP method would be to follow the update proscribed by the parameter–primal pair iteration process from the usual SQP method, which would essentially reproduce the usual SQP method.

One advantage of the extra structure demanded in the usual SQP method is that there are optimality conditions of a much simpler form than (9). The analysis of these conditions leads to the well known quadratic convergence result, but this result relates only to the distances $\|(u, x) - (\bar{u}, \bar{x})\|$ between parameter–primal pairs (u, x) and a parameter-primal target pair $(\bar{u}, \bar{x})$, and not, for example, to the distance $\|x - \bar{x}\|$ between primal iterates alone x and the target $\bar{x}$. In fact even for the usual SQP method, superlinear convergence is all that can typically be guaranteed for the primal iterates alone (see e.g. [6, Inequality 12.4.16]), so Proposition 2.1 reproduces the usual *primal* convergence result for the usual SQP method.

A generalized Newton method in [4] was applied to the optimality condition associated with a generalized SQP method to show quadratic convergence under some circumstances. However, the generalized Newton method from [4] only covers optimality conditions whose set-valued components depend only on x' (like those that result from the usual SQP method). Therefore, the results in [4] do not apply to our primal SQP method since its optimality condition (9) includes a set-valued component

$$\nabla g(x)^T \cdot N_K\big(\nabla g(x)[x' - x] + g(x)\big)$$

that clearly depends on both x and x'.

3. Derivative Characterizations of Supercalmness

In [12], a characterization of multifunction calmness was given in terms of the 'outer graphical derivative' associated with the multifunction. For any multifunction S: $\mathcal{P} \rightrightarrows \mathcal{X}$, the *outer graphical derivative of S at* $\bar{p}$ *for* $\bar{x}$ is the multifunction $DS(\bar{p}|\bar{x})$: $\mathcal{P} \rightrightarrows \mathcal{X}$ defined by

$$DS(\bar{p}|\bar{x})(p) := \{x\colon \exists p_k \to p,\ t_k \downarrow 0,\ \text{and}\ x_k \to x \text{ with } \bar{p} + t_k\, p_k \in S(\bar{x} + t_k\, x_k)\}\,.$$

THEOREM 3.1 [12, Proposition 4.1]. *For any multifunction* S: $\mathcal{P} \rightrightarrows \mathbb{R}^n$ *with* $\mathcal{P}$ *a normed linear space, the following are equivalent:*

(a) *The inclusion* $x \in DS(\bar{p}|\bar{x})(0)$ *implies that* $x = 0$.
(b) *The multifunction S is calm at* $\bar{p}$ *for* $\bar{x}$.

Remarks. This result was first proved in [18, Theorem 4.1] under an assumption of 'proto-differentiability' of S. The same result was then proved without the proto-differentiability assumption in [12, Proposition 4.1] (using the proof of the implication (a) ⇒ (b) from [8, Proposition 2.1]) where the stipulation that $\bar{x} \in S(\bar{p})$ was assumed throughout since outer graphical derivatives there are defined (as they are usually, see [19]) only at $\bar{p}$ for $\bar{x} \in S(\bar{p})$. Notice that if $\bar{x} \notin S(\bar{p})$ and the graph of S is closed near $(\bar{p}, \bar{x})$, then the outer graphical derivative image-set $DS(\bar{p}|\bar{x})(p)$ is empty for any p, and S is trivially calm at $\bar{p}$ for $\bar{x}$.

We can characterize our new property of multifunction supercalmness in a similar way with the outer graphical derivative, but only for multifunctions between finite-dimensional spaces.

THEOREM 3.2 (supercalmness characterization). *For any multifunction* $S: \mathbb{R}^d \rightrightarrows \mathbb{R}^n$, *the following are equivalent:*

(a) *For all* $p \in \mathbb{R}^d$, *the inclusion* $x \in DS(\bar{p}|\bar{x})(p)$ *implies that* $x = 0$.
(b) *The multifunction S is supercalm at* $\bar{p}$ *for* $\bar{x}$.

Proof. To show (a)⇒(b), we assume that (b) does not hold. This means that there exists a constant $L > 0$ and sequences $\tilde{x}_k \to \bar{x}$ and $\tilde{p}_k \to \bar{p}$ satisfying $\tilde{x}_k \in S(\tilde{p}_k)$ and

$$\|\tilde{x}_k - \bar{x}\| > L\|\tilde{p}_k - \bar{p}\|.$$

From this, we immediately conclude the bound

$$\frac{1}{L} > \frac{\|\tilde{p}_k - \bar{p}\|}{t_k} \tag{19}$$

in terms of $t_k := \|\tilde{x}_k - \bar{x}\|$. If we further define

$$p_k := \frac{\tilde{p}_k - \bar{p}}{t_k} \quad \text{and} \quad x_k := \frac{\tilde{x}_k - \bar{x}}{t_k},$$

then the fact that $\tilde{x}_k \in S(\tilde{p}_k)$ translates into

$$\bar{x} + t_k x_k \in S(\bar{p} + t_k p_k).$$

From the definition of t_k and the bound (19), we conclude that (passing to subsequences if necessary) $x_k \to x$ with $\|x\| = 1$ and that $p_k \to p$ for some $p \in \mathbb{R}^d$. It follows from the definition of the outer graphical derivative that $x \in DS(\bar{p}|\bar{x})(0)$ with $x \neq 0$, which contradicts (a).

Assuming (b), we consider any $x \in DS(\bar{p}|\bar{x})(p)$. This means there are sequences $x_k \to x$, $p_k \to p$, and $t_k \downarrow 0$ satisfying $\bar{x} + t_k x_k \in S(\bar{p} + t_k p_k)$. By the supercalmness assumption, we know that for any $L > 0$, we eventually have the bound $\|x_k\| \leqslant L\|p_k\|$, which in the limit implies that $\|x\| \leqslant L\|p\|$. Since this bound holds for any $L > 0$, we conclude that $x = 0$. □

Remark. Notice that condition (a) of Theorem 3.2 is clearly stronger than condition (a) of the calmness characterization Theorem 3.1 where the same outer graphical derivative multifunction is only presumed to be zero at $p = 0$.

Notice too that this result could be extended to multifunctions from a more general normed linear space $\mathcal{P}$ if the outer graphical derivative was modified to use all sequences p_k weak* converging to p.

We can also characterize the partial supercalmness property with a condition on an outer graphical derivative.

THEOREM 3.3 (partial supercalmness characterization). *For any multifunction* $S\colon \mathbb{R}^d \times \mathcal{Q} \rightrightarrows \mathbb{R}^n$ *with* $\mathcal{Q}$ *a normed linear space, the following are equivalent:*

(a) *For all* $p \in \mathbb{R}^d$*, the inclusion* $x \in DS(\bar{p}, \bar{q}|\bar{x})(p, 0)$ *implies that* $x = 0$.
(b) *The multifunction* S *is partially supercalm with respect to* p *at* $(\bar{p}, \bar{q})$ *for* $\bar{x}$.

Proof. To show (a)⇒(b), we assume that (b) does not hold. This means that there exists a constant $L_p > 0$ and sequences $L_{q,k} \to \infty$, $\tilde{x}_k \to \bar{x}$, $\tilde{p}_k \to \bar{p}$, and $\tilde{q}_k \to \bar{q}$ satisfying $\tilde{x}_k \in S(\tilde{p}_k, \tilde{q}_k)$ and

$$\|\tilde{x}_k - \bar{x}\| > L_p\|\tilde{p}_k - \bar{p}\| + L_{q,k}\|\tilde{q}_k - \bar{q}\|.$$

From this, we immediately conclude the pair of inequalities

$$\|\tilde{x}_k - \bar{x}\| > L_p \|\tilde{p}_k - \bar{p}\| \quad \text{and} \quad \|\tilde{x}_k - \bar{x}\| > L_{q,k} \|\tilde{q}_k - \bar{q}\|$$

which, with $t_k := \|\tilde{x}_k - \bar{x}\|$, implies the corresponding pair of inequalities

$$\frac{1}{L_p} > \frac{\|\tilde{p}_k - \bar{p}\|}{t_k} \quad \text{and} \quad \frac{1}{L_{q,k}} > \frac{\|\tilde{q}_k - \bar{q}\|}{t_k} \tag{20}$$

If we define

$$p_k := \frac{\tilde{p}_k - \bar{p}}{t_k} \quad q_k := \frac{\tilde{q}_k - \bar{q}}{t_k} \quad \text{and} \quad x_k := \frac{\tilde{x}_k - \bar{x}}{t_k}$$

then the inclusion $\tilde{x}_k \in S(\tilde{p}_k, \tilde{q}_k)$ translates into

$$\bar{x} + t_k x_k \in S(\bar{p} + t_k p_k, \bar{q} + t_k q_k).$$

Moreover, from the definition of t_k and the bounds (20), we conclude that (passing to subsequences if necessary) $x_k \to x$ with $\|x\| = 1$, $p_k \to p$ for some $p \in \mathbb{R}^d$, and $q_k \to 0$. It follows from the definition of the outer graphical derivative that $x \in DS(\bar{p}, \bar{q}|\bar{x})(p, 0)$ with $x \neq 0$, which contradicts (a).

Assuming (b), we consider any $x \in DS(\bar{p}, \bar{q}|\bar{x})(p, 0)$. This means there are sequences $x_k \to x$, $p_k \to p$, and $q_k \to 0$ satisfying $\bar{x} + t_k x_k \in S(\bar{p} + t_k p_k, \bar{q} + t_k q_k)$. From the partial supercalmness assumption, we know that for any $L_p > 0$ there exists an $L_q > 0$ such that we eventually have the bound $\|x_k\| \leqslant L_p \|p_k\| + L_q \|q_k\|$, which in the limit implies that $\|x\| \leqslant L_p \|p\|$. Since $L_p > 0$ is arbitrary, we conclude that $x = 0$. $\square$

Remark. Note that Theorem 3.3 essentially covers both Theorems 3.2 and 3.1. For Theorem 3.2, we just extend the multifunction to have trivial dependence on $q \in \mathcal{Q}$, and for Theorem 3.1 we consider the multifunction as depending on $q \in \mathcal{Q}$ in place of $p \in \mathcal{P}$.

3.1. Convergence of the Primal SQP Method Revisited

The key assumption in Proposition 2.1 giving superlinear convergence of the primal SQP method was the partial supercalmness at $(\bar{x}, \bar{q})$ for $\bar{x}$ of the multifunction S defined in (8) as

$$S(x, q) := \left\{ x' \in \mathbb{R}^n \;\middle|\; q \in \begin{array}{c} \nabla^2\big(f_0 + g_{\bar{u}}\big)(x)[x' - \bar{x}] + \nabla g_{\bar{u}}(\bar{x}) - \nabla g_{\bar{u}}(x) \\ +\nabla g(x)^T \cdot N_K\big(\nabla g(x)[x' - x] + g(x)\big) \end{array} \right\}.$$

With Theorem 3.3 in hand, we can establish verifiable conditions to ensure this partial supercalmness as soon as we compute the appropriate outer graphical derivative.

First, notice that the outer graphical derivative of S at $(\bar{x}, \bar{q})$ for $\bar{x}$ is empty-valued unless $\bar{x} \in S(\bar{x}, \bar{q})$, in which case we can apply [10, Theorem 4.1] to get the formula

$$DS(\bar{x}, \bar{q}|\bar{x})(x, q) = \{x' | q \in \nabla^2\big(f_0 + g_{\bar{u}}\big)(\bar{x})[x'] - \nabla^2 g_{\bar{u}}(\bar{x})[x] + DM(\bar{x}, \bar{x}|\bar{q})(x, x')\} \tag{21}$$

in terms of the multifunction M: $\mathbb{R}^{2n} \rightrightarrows \mathbb{R}^n$ defined by

$$M(x, x') := \nabla g(x)^T \cdot N_K\big(\nabla g(x)[x' - x] + g(x)\big).$$

Now we use the formula from [11, Theorem 3.2] for the outer graphical derivative of multifunctions like M, with x playing the role of w, x' playing the role of x, and the mapping G from [11] defined by $G(x, w) := \big(\nabla f_1(w)[x - w] + f_1(w), \ldots, \nabla f_m(w)[x - w] + f_m(w)\big)$. In this case, [11, Theorem 3.2] gives that the set $DM(\bar{x}, \bar{x}|\bar{q})(x, x')$ is empty unless x' is in the critical cone

$$X' := \left\{ x' \in \mathbb{R}^n \;\middle|\; \begin{array}{lll} \nabla f_i(\bar{x})[x'] \leqslant 0 & \text{for} & i \in \mathcal{I} \text{ with } f_i(\bar{x}) = 0 \\ \nabla f_i(\bar{x})[x'] = 0 & \text{for} & i \in \mathcal{E} \cup \{0\} \end{array} \right\}$$

and that if $x' \in X'$, the set $DM(\bar{x}, \bar{x}|\bar{q})(x, x')$ consists of all the points

$$\nabla^2\big(\tilde{y} \cdot g\big)(\bar{x})[x] + \nabla\big(y' \cdot g\big)(\bar{x}) + y_0' \nabla f_0(\bar{x})$$

generated by choices of $y_0' \in \mathbb{R}$, y' in the set

$$Y'(x') := \big\{ y' \in N_K\big(g(\bar{x})\big) \text{ with } y_i' \nabla f_i(\bar{x})[x'] = 0 \big\}$$

and $\tilde{y} \in Y_{\max}(x, x')$, where $Y_{\max}(x, x')$ is the set of vectors $\tilde{y} \in \mathbb{R}^m$ maximizing the function

$$\tilde{y} \mapsto \big\langle x, \nabla^2\big(\tilde{y} \cdot g\big)(\bar{x})[2x' - x]\big\rangle$$

over the set of multipliers $\widetilde{Y}$ defined in (7). Plugging all of this into the formula (21) for the outer graphical derivative of S gives

$$DS(\bar{x}, \bar{q}|\bar{x})(x, q) = \left\{ x' \in X' \;\middle|\; \begin{array}{c} \nabla^2\big(f_0 + g_{\bar{u}}\big)(\bar{x})[x'] \\ q \in + \bigcup_{\tilde{Y} \in_{\max}(x, x')} \Big(\nabla^2\big(\tilde{y} \cdot g\big)(\bar{x}) - \nabla^2 g_{\bar{u}}(\bar{x})\Big)[x] \\ + \bigcup_{y' \in Y'(x')} \nabla\big(y' \cdot g\big)(\bar{x}) + \bigcup_{y_0' \in \mathbb{R}} y_0' \nabla f_0(\bar{x}) \end{array} \right\}, \qquad (22)$$

which is the formula we need to state and prove the corollary to Theorem 3.3 that applies in this case.

COROLLARY 3.1. *Assume the following:*

- *The second-order condition*

$$0 \in \nabla^2\big(f_0 + g_{\bar{u}}\big)(\bar{x})[x'] + \bigcup_{y' \in Y'(x')} \nabla\big(y' \cdot g\big)(\bar{x}) + \bigcup_{y_0' \in \mathbb{R}} y_0' \nabla f_0(\bar{x}) \text{ for } x' \in X' \Rightarrow x' = 0. \qquad (23)$$

- *The strict Mangasarian–Fromovitz constraint qualification at* $\bar{x}$:

  The set of multipliers from (7) is a singleton $\widetilde{Y} = \{\tilde{y}\}$.

- *The function* $g_{\bar{u}}$ *satisfies*

$$\nabla^2 g_{\bar{u}}(\bar{x}) = \nabla^2\big(\tilde{y} \cdot g\big)(\bar{x}). \qquad (24)$$

Then there exists a neighborhood $X \subseteq \mathbb{R}^n$ of $\bar{x}$ such that any sequence of iterates $\{x_k\} \subseteq X$ generated by the primal SQP method for a parameter sequence $u_k \to \bar{u}$ converges to $\bar{x}$ at a superlinear rate.

Proof. Using formula (22) for the outer graphical derivative and the assumptions that $\widetilde{Y} = \{\tilde{y}\}$ (which implies that $Y_{\max}(x, x') = \{\tilde{y}\}$) and (24), we see that the image set $DS(\bar{x}, \bar{q}|\bar{x})(x, 0)$ satisfies

$$DS(\bar{x}, \bar{q}|\bar{x})(x, 0) = \left\{ x' \in X' \;\middle|\; 0 \in \nabla^2\big(f_0 + g_{\bar{u}}\big)(\bar{x})[x'] + \bigcup_{y' \in Y'(x')} \nabla\big(y' \cdot g\big)(\bar{x}) + \bigcup_{y'_0 \in \mathbb{R}} y'_0 \nabla f_0(\bar{x}) \right\},$$

so that the second-order condition (23) implies condition (a) from Theorem 3.3 in this case. The result then follows from Theorem 3.3 and Proposition 2.1. $\square$

Remark. After multiplying the inclusion defining the second-order condition (23) on both sides by x' and noting the definitions of the sets $Y'(x')$ and X', we see that (23) is implied by the condition

$$\big\langle x', \nabla^2\big(f_0 + g_{\bar{u}}\big)(\bar{x})[x']\big\rangle = 0 \quad \text{for} \quad x' \in X' \Rightarrow x' = 0$$

which is ensured by the positive-definiteness on the critical cone X' of the Hessian of the Lagrangian $x \mapsto f_0(x) + g_{\bar{u}}(x)$. For the standard choice (6) of $g_{\bar{u}} = \bar{u} \cdot g$, such positive-definiteness is a standard second-order sufficient condition for optimality.

Note that condition (24) identifies the families of functions g_u by a condition involving only the single representative $g_{\bar{u}}$ corresponding to the target parameter. Moreover, only the Hessian at $\bar{x}$ of this representative needs to match the Hessian of $\tilde{y} \cdot g$ at $\bar{x}$ in order to get convergence at a superlinear rate. In particular, the standard choice (6) of $g_{\bar{u}} = \bar{u} \cdot g$ trivially satisfies (24) with the target parameter $\bar{u}$ equal to the multiplier $\tilde{y}$. Even in this standard case, Corollary 3.1 provides new insight since general multipliers u_k are allowed and since the second-order condition (23) is weaker than the standard second-order condition.

References

1. Bonnans, J.F.: Local analysis of Newton-type methods for variational inequalities and nonlinear programming. Appl. Math. Optim. **29**, 161–186 (1994)
2. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
3. Dontchev, A.D.: Characterizations of Lipschitz stability in optimization. In: Lucchetti, R., Revalski, J. (eds.) Recent Developments in Well-Posed Problems, pp. 95–115. Kluwer, Dordrecht, The Netherlands (1995)
4. Dontchev, A.D.: Lipschitzian stability of Newton's method for variational inclusions. In: System Modeling and Optimization (Cambridge, 1999), pp. 119–147. Kluwer, Boston (2000)
5. Dontchev, A.D., Rockafellar, R.T.: Regularity and conditioning of solution mappings in variational analysis. Set-Valued Anal. **12**, 79–109 (2004)
6. Fletcher, R.: Practical Methods of Optimization, 2nd edn. Wiley, New York (1987)
7. Gowda, M.S., Pang, J.-S.: Stability analysis of variational inequalities and nonlinear complementarity problems, via the mixed linear complementarity problem and degree theory. Math. Oper. Res. **19**, 831–879 (1994)
8. King, A.J., Rockafellar, R.T.: Sensitivity analysis for nonsmooth generalized equations. Math. Programming **55**, 193–212 (1992)

9. Klatte, D., Kummer, B.: Contingent derivatives of implicit (multi-) functions and stationary points in optimization with data perturbations, II. Ann. Oper. Res. **101**, 313–331 (2001)
10. Levy, A.B., Rockafellar, R.T.: Sensitivity analysis of generalized equations. Trans. Amer. Math. Soc. **345**, 661–671 (1994)
11. Levy, A.B., Rockafellar, R.T.: Sensitivity of solutions in nonlinear programming problems with nonunique multipliers. In: Du, D.-Z., Qi, L., Womersley, R.S. (eds.) Recent Advances in Nonsmooth Optimization, pp. 215–223. World Scientific, River Edge, New Jersey (1995)
12. Levy, A.B.: Implicit multifunction theorems for the sensitivity analysis of variational conditions. Math. Programming **74**, 333–350 (1996)
13. Levy, A.B.: Solution sensitivity from general principles. SIAM J. Control Optim. **40**, 1–38 (2001)
14. Pang, J.-S.: A degree-theoretic approach to parametric nonsmooth equations with multivalued perturbed solution sets. Math. Programming **62**, 359–383 (1993)
15. Pang, J.-S.: Necessary and sufficient conditions for solution stability of parametric nonsmooth equations. In: Du, D.-Z., Qi, L., Womersley, R.S. (eds.) Recent Advances in Nonsmooth Optimization, pp. 261–288. World Scientific, River Edge, New Jersey (1995)
16. Robinson, S.M.: Generalized equations and their solutions. Part I. Basic theory. Math. Programming Stud. **10**, 128–141 (1979)
17. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–989 (1976)
18. Rockafellar, R.T.: Proto-differentiability of set-valued mappings and its application in optimization. In: Attouch, H., Aubin, J.-P., Clarke, F.H., Ekeland, I. (eds.) Analyse Non Linéaire, pp. 449–482. Gauthier-Villars, Paris (1989)
19. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin Heidelberg New York (1998)