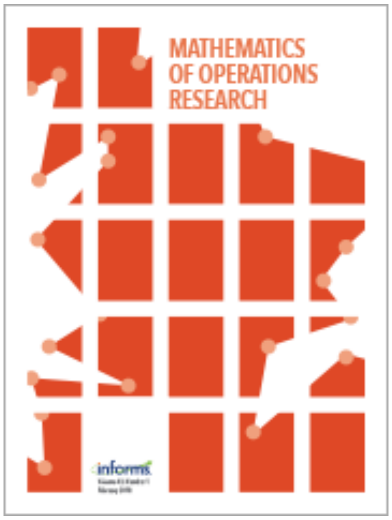
1-1-2005

Convergence of successive approximation methods with parameter target sets

A.B. Levy



Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Convergence of Successive Approximation Methods with Parameter Target Sets

Adam B. Levyhttp://academic.bowdoin.edu/faculty/A/alevy/,

M THEM TICS OF OPER TIONS RESE RCH
Vol. 30, No. 3, August 2005, pp. 765–784
ISSN 0364-765X | EISSN 1526-5471 | 05 | 3003 | 0765
DOI 10.1287/moor.1050.0153

Convergence of Successive Approximation Methods with Parameter Target Sets

Adam B. Levy
Department of Mathematics, Bowdoin College, Brunswick, Maine 04011,
alevy@bowdoin.edu, http://academic.bowdoin.edu/faculty/A/alevy/

Successive approximation methods appear throughout numerical optimization, where a solution to an optimization problem is sought as the limit of solutions to a succession of simpler approximation problems. Such methods include essentially any standard penalty method, barrier method, trust region method, augmented Lagrangian method, or sequential quadratic programming (SQP) method, as well as many other methods. The approximation problems on which a successive approximation method is based typically depend on parameters, in which case the performance of the method is related to the corresponding sequence of parameters. For many successive approximation methods, the sequence of parameters might need only approach some parameter target set for the method to have nice convergence properties. Successive approximation methods could be analyzed as examples of a generic inclusion solving method from Levy [23] because the solutions to the approximation problems satisfy necessary optimality inclusions. However, the inclusion solving method from Levy [23] was developed for single-parameter target points. In this paper, we extend the results from Levy [23] to allow parameter target sets and apply these results to the convergence analysis of successive approximation methods. We focus on two important convergence issues: (1) the rate of convergence of the iterates generated by a successive approximation method and (2) the validity of the limit as a solution to the original problem. An augmented Lagrangian method allowing quite general parameter updating is explored in detail to illustrate how the framework presented here can expose interesting new alternatives for numerical optimization.

Key words: constrained optimization; numerical optimization; penalty methods; barrier methods; trust region methods; augmented Lagrangian methods; sequential quadratic programming; convergence analysis; inclusion solving; variational analysis; generalized continuity
MSC2000 subject classi cation: Primary: 49M30; secondary: 65K05, 90C31
OR/MS subject classi cation: Primary: programming nonlinear algorithms
History: Received January 22, 2004; revised July 9, 2004, and January 6, 2005.

1. Introduction. The general constrained optimization problem we will consider is

$$\min f_0(x) \quad \text{over } x \in X \tag{1}$$

with X some subset of the normed linear space $\mathscr{X}$ and with $\mathscr{C}^1$ objective function $f_0: \mathscr{X} \to \mathbb{R}$. The *successive approximation methods* described in this paper solve the general constrained optimization problem (1) by successively solving unconstrained approximation problems whose objective functions include a quadratic term $q_{h,x}: \mathscr{X} \to \mathbb{R}$, approximating f_0 at the current iterate x. As the notation suggests, these quadratic functions are de ned by a xed parameter h from the space $\mathscr{H}$ of linear mappings from $\mathscr{X}$ to its dual $\mathscr{X}^*$, and they moreover satisfy $q_{h,x}(x) = f_0(x)$ and $\nabla q_{h,x}(x') = \nabla f_0(x) + h[x' - x]$ for all $x' \in \mathscr{X}$. The classic example of such a quadratic approximation $q_{h,x}$ for smooth objective functions f_0 is the quadratic Taylor approximation, in which case the linear mapping h is just the Hessian $\nabla^2 f_0(x)$. Other choices of h cover various "quasi-Newton" methods that could be applied in both the constrained and unconstrained cases (when the set X is the whole space $\mathscr{X}$).

Successive approximation methods. Given a current iterate x_k and parameter $u_k := (\mu_k, h_k)$, choose the next iterate $x_{k+1} = x'$ by solving the (unconstrained) approximation problem

$$\min q_{h_k, x_k}(x') + n_{\mu_k, x_k}(x') \quad \text{over all } x' \in \mathscr{X}, \tag{2}$$

where $q_{h,x}$ is a quadratic approximation to f_0 at x as outlined above, and $n_{\mu,x}: \mathscr{X} \to \mathbb{R} \cup \{\infty\}$ is some family of functions.

Successive approximation methods are quite common, including essentially any standard penalty method, barrier method, trust region method (where $n_{\mu,x}$ includes a penalty for violating the trust region), augmented Lagrangian method, or SQP method. As we will illustrate with two simple penalty examples in §2, the parameters μ_k in successive approximation methods might need only approach a parameter target set $\bar{M}$ for the method to have nice convergence properties. For instance, a typical parameter target set for an exact penalty method is a semi-in nite interval $\bar{M} := [\bar{\mu}, \infty)$, indicating that the real-valued penalty parameter need only be big enough.

The iterates generated by a successive approximation method satisfy the inclusions

$$-\nabla q_{h_k, x_k}(x') \in \partial n_{\mu_k, x_k}(x'), \tag{3}$$

encoding the basic necessary condition for optimality associated with the approximation problem (2). This inclusion is an "approximating inclusion" for the inclusion

$$-\nabla f_0(x) \in N_X(x), \tag{4}$$

encoding the basic necessary optimality condition associated with the original minimization problem (1). In Levy [23], a generic algorithm was studied for solving inclusions of the form

$$f(x) \in F(x), \tag{5}$$

which includes (4) with $f(x) := -\nabla f_0(x)$ and $F(x) := N_X(x)$ by successively solving approximating inclusions of the form

$$a(u, x, x') \in A(u, x, x'), \tag{6}$$

which includes (3) with $u := (\mu, h)$, $a(u, x, x') := -\nabla q_{h,x}(x')$, and $A(u, x, x') := \partial n_{\mu,x}(x')$. The convergence of the generic algorithm was analyzed with a variety of tools developed in Levy [23], including generalized rate functions, "appropriate" sequences of parameters, and "valid" approximating inclusions. We would like to apply the results in Levy [23] to analyze the convergence of successive approximation methods, however, the tools in Levy [23] were all developed in terms of a single parameter target $\bar{u}$, and evidently many successive approximation methods use more general parameter target sets $\bar{U}$. Therefore, in this paper, we extend many of the tools and results from Levy [23] to allow parameter target sets. To make these extensions, we develop new notions of "appropriate" sequences of parameters and of "valid" approximating inclusions, and we explore new connections between nonsingularity conditions and continuity properties of set-valued mappings. Once we develop the necessary extensions, we apply the results to analyze the convergence of successive approximation methods.

Our results for successive approximation methods cover two major convergence issues: (1) convergence rate (developed in §2) and (2) the validity as a solution to the original problem of the limit of the iterates generated by a successive approximation method (developed in §3). We say that a successive approximation method is *valid at* $\bar{x}$ *for* $\bar{U}$ if whenever the sequence of parameters $\{u_k\}$ converges to the target set $\bar{U}$ and the sequence of iterates generated by the corresponding successive approximation method converges, then those iterates converge to a solution of the basic necessary optimality condition (4) associated with the original minimization problem (1). In §3, we develop a useful suf cient condition for validity of a successive approximation method and apply this condition in the cases of the two simple penalty examples from §2, as well as in the cases of a new primal SQP method and a new primal augmented Lagrangian method. In the nal section of this paper, we apply our general convergence rate results to analyze the convergence rate for our primal augmented Lagrangian method, which illustrates how our approach can expose interesting new alternatives for numerical optimization.

There is a vast amount of literature on particular successive approximation methods and the reader can consult Boukari and Fiacco [3] for a survey of penalty, exact penalty, and multiplier methods through 1993. For basic background, we refer the reader to Nocedal and Wright [24] and Fletcher [22], which are comprehensive textbooks on numerical optimization. There are too many particular examples to cite here, but some recent contributions include (in rough categories): penalty methods DiPillo and Facchinei [8], DiPillo [7], Facchinei et al. [21], Facchinei and Lucidi [20], DiPillo et al. [12], Demyanov et al. [6], DiPillo et al. [10], Facchinei [18], Facchinei and Lucidi [19], Contaldi et al. [5], Facchinei [17], Byrd et al. [4]; barrier methods DiPillo and Facchinei [9], DiPillo et al. [14]; augmented Lagrangian methods DiPillo et al. [11], DiPillo et al. [13], Ben-Tal and Zibulevsky [1]; SQP methods Boggs and Tolle [2]; and trust region methods Sadjadi and Ponnambalam [26]. Our results here are unusual in their scope and breadth, and there are several bene ts to be gained from this. First, new insight into particular existing methods can be gained by recognizing a broader surrounding category. We illustrate this by showing that the standard parameter update scheme associated with a classical augmented Lagrangian method is just one among many new possibilities that lead to similar convergence results. Other optimization methods, where parameters are speci ed as part of an updating scheme, can similarly be expanded within our framework. Not only does our framework allow the expansion of existing methods of optimization, but also it provides a template for creating and analyzing entirely new methods. Moreover, the knowledge that any particular method ts into a more general framework deepens our understanding of that method and exposes its connections to other methods in the framework. Such connections can be exploited to translate and propagate specialized results throughout the entire framework.

2. Convergence rate. To motivate our extension here of the results in Levy [23], we now consider the simple constrained minimization problem

$$\min f_0(x) := x_1^2 + x_2^2 \quad \text{over } x \in X := \{x = (x_1, x_2) \in \mathbb{R}^2 \text{ with } x_1 + x_2 = 1\}, \tag{7}$$

whose solution is $\bar{x} = (1/2, 1/2)$. Since the objective function f_0 is quadratic itself, the quadratic approximation in this case can be $q_{h,x}(x') := f_0(x') = (x_1')^2 + (x_2')^2$. One choice for the $n_{\mu,x}$ is the exact penalty functions of the form

$$n_{\mu,x}(x') := \mu|x_1' + x_2' - 1|, \tag{8}$$

de ned by penalty parameters $\mu \in [0, \infty)$. The solution to the approximation problem (2) in this case is

$$x' = \begin{cases} (\mu_k/2, \mu_k/2) & \text{if } \mu_k \in [0,1] \\ (1/2, 1/2) & \text{if } \mu_k \in [1, \infty) \end{cases},$$

so for the successive approximation method to identify the solution $\bar{x} = (1/2, 1/2)$ to the original minimization problem (7), the parameters need to approach the parameter target set $[1, \infty)$ (such parameter target sets are typical in exact penalty methods).

Another simple choice for the $n_{\mu,x}$ in this case is inexact penalty functions like

$$n_{\mu,x}(x') := \mu(x_1' + x_2' - 1)^2, \tag{9}$$

again de ned by penalty parameters $\mu \in [0, \infty)$. The solution to the approximation problem (2) with this choice of functions $n_{\mu,x}$ is $x' = (\mu_k/(1+2\mu_k), \mu_k/(1+2\mu_k))$. Thus, for the successive approximation method to identify the solution $\bar{x} = (1/2, 1/2)$ to the original minimization problem (7), the parameters need to approach the parameter target set $\{\infty\}$ (which is the typical parameter target set for inexact penalty methods). It is clear from this example that parameter target sets (and not just nite target points) need to be used if penalty methods are to be accommodated.

2.1. Convergence rate for the generic algorithm with a parameter target set. We recall the following definition from Levy [23]: A function $\psi: \mathbb{R} \to \mathbb{R}$ is called a *convergence function* if it is nondecreasing on $\mathbb{R}^+$ with $\psi(0) = 0$. As in Levy [23], we will use families of convergence functions $\psi_{(u,u')}$ parameterized by $(u, u') \in U \times U$ for some parameter set U in the parameter space $\mathcal{U}$, as well as a notion of the "target appropriateness" of a sequence of parameters. Extending the definition from Levy [23] to parameter target sets $\bar{U} \subseteq U$, we say that a sequence of parameters $u_k \in U$ is *target set appropriate for* $(\psi_{(u,\bar{u})}, r_0)$ *with target set* $\bar{U}$ if the sequence of scalars $\{r_k\}$ defined by

$$r_0, r_k := \inf_{\bar{u} \in \bar{U}} \psi_{(u_{k-1},\bar{u})}(r_{k-1}) \quad \text{for } k = 1, 2, \ldots$$

converges to zero. The definition from Levy [23] of target appropriateness with parameter target $\bar{u}$ is covered by target set appropriateness with a singleton target set $\bar{U} := \{\bar{u}\}$.

The following theorem generalizes Levy [23, Theorem 3.2] to allow parameter target sets. This result uses the closed ball $(\bar{x}; \epsilon_x) \subseteq \mathcal{X}$ defined for some $\epsilon_x \geq 0$ by

$$(\bar{x}; \epsilon_x) := \{x \in \mathcal{X} \mid \|x - \bar{x}\| \leq \epsilon_x\},$$

as well as the concept of a family of *bivariate convergence functions* $\phi_{(u,u')}: \mathbb{R}^2 \to \mathbb{R}$ parameterized by $(u, u') \in U \times U$ and defined by the sublinearity inequality

$$\phi_{(u,u')}(r, r') \leq \psi_{(u,u')}(r) + \alpha r' \quad \text{for all } (u, u') \in U \times U \text{ and } (r, r') \in \mathbb{R}^+ \times \mathbb{R}^+$$

in terms of a family of convergence functions $\psi_{(u,u')}$ and a fixed sublinearity constant $\alpha \in (0, \infty)$.

Theorem 2.1 (Convergence Rate for the Generic Algorithm with a Parameter Target Set). *For the generic iteration mapping*

$$G(u, x) := (A(u, x, \ ) - a(u, x, \ ))^{-1}(0),$$

if there exist scalars $\alpha \in (0, \infty)$ *and* $\epsilon_x \geq 0$, *a parameter set* $U \subseteq \mathcal{U}$, *a parameter target set* $\bar{U} \subseteq U$, *and a single-valued mapping* $\sigma: \mathcal{U} \times \mathcal{U} \times \mathcal{X} \times \mathcal{X} \times \mathcal{X} \to \mathcal{Y}$ *such that the following pair of conditions hold*:

- (*Target set* σ-*approximation*) *There exists a family of bivariate convergence functions* $\phi^1_{(u,u')}$ *parameterized by* $(u, u') \in U \times U$ *and with sublinearity constant* $1/(2\alpha)$ *such that any trio* $(u, x, x') \in U \times (\bar{x}; \epsilon_x) \times G(u, x) \cap (\bar{x}; \epsilon_x)$ *satisfies*

$$\|\sigma(u, \bar{u}, x, x', \bar{x})\| \leq \phi^1_{(u,\bar{u})}(\|x - \bar{x}\|, \|x' - \bar{x}\|) \quad \forall\, \bar{u} \in \bar{U}$$

- (*Target set nonsingularity*) *There exists a family of bivariate convergence functions* $\phi^2_{(u,u')}$ *parameterized by* $(u, u') \in U \times U$ *and with sublinearity constant* α *such that any trio* $(u, x, x') \in U \times (\bar{x}; \epsilon_x) \times G(u, x) \cap (\bar{x}; \epsilon_x)$ *satisfies*

$$\|x' - \bar{x}\| \leq \inf_{\bar{u} \in \bar{U}} \phi^2_{(u,\bar{u})}(\|x - \bar{x}\|, \|\sigma(u, \bar{u}, x, x', \bar{x})\|),$$

then any sequence $\{x_k\} \subseteq (\bar{x}; \epsilon_x)$ *initiated from* x_0 *and conforming to the generic algorithm* $x_{k+1} \in G(u_k, x_k)$ *with a sequence of parameters* $\{u_k\} \subseteq U$ *satisfies*

$$\|x_{k+1} - \bar{x}\| \leq \inf_{\bar{u} \in \bar{U}} \psi_{(u_k,\bar{u})}(\|x_k - \bar{x}\|) \tag{10}$$

in terms of the convergence functions $\psi_{(u,u')}$ *defined by*

$$\psi_{(u,u')}(r) := 2\alpha\psi^1_{(u,u')}(r) + 2\psi^2_{(u,u')}(r)$$

Moreover, if the sequence of parameters $\{u_k\}$ *is target set appropriate for* $(\psi_{(u,\bar{u})}, \|x_0 - \bar{x}\|)$ *with target set* $\bar{U}$, *then the sequence* $\{x_k\}$ *converges to* $\bar{x}$.

Proof. Since $x_k \in \ (\bar{x}; \epsilon_x)$, $u_k \in U$, and $x_{k+1} \in G(u_k, x_k) \cap \ (\bar{x}; \epsilon_x)$, we can apply the target set nonsingularity bound with $(u, x, x') := (u_k, x_k, x_{k+1})$ to deduce that

$$\begin{aligned}\|x_{k+1} - \bar{x}\| &\le \inf_{\bar{u} \in \bar{U}} \phi^2_{(u_k, \bar{u})}(\|x_k - \bar{x}\|, \|\sigma(u_k, \bar{u}, x_k, x_{k+1}, \bar{x})\|)\\ &\le \inf_{\bar{u} \in \bar{U}} \psi^2_{(u_k, \bar{u})}(\|x_k - \bar{x}\|) + \alpha\|\sigma(u_k, \bar{u}, x_k, x_{k+1}, \bar{x})\|,\end{aligned}$$

where the second inequality comes from the sublinearity of ϕ^2. Applying the target set σ-approximation bound to this inequality, we get

$$\begin{aligned}\|x_{k+1} - \bar{x}\| &\le \inf_{\bar{u} \in \bar{U}} \psi^2_{(u_k, \bar{u})}(\|x_k - \bar{x}\|) + \alpha\phi^1_{(u_k, \bar{u})}(\|x_k - \bar{x}\|, \|x_{k+1} - \bar{x}\|)\\ &\le \inf_{\bar{u} \in \bar{U}} \psi^2_{(u_k, \bar{u})}(\|x_k - \bar{x}\|) + \alpha\psi^1_{(u_k, \bar{u})}(\|x_k - \bar{x}\|) + \tfrac{1}{2}\|x_{k+1} - \bar{x}\|,\end{aligned}$$

where the second inequality comes from the sublinearity of ϕ^1. After collecting the $\|x_{k+1} - \bar{x}\|$-terms and multiplying through by 2, we get the desired bound (10).

If the sequence of parameters $\{u_k\}$ is target appropriate for $(\psi_{(u, \bar{u})}, \|x_0 - \bar{x}\|)$ with target $\bar{u}$, we know that the sequence of scalars

$$r_0 := \|x_0 - \bar{x}\|, \qquad r_k := \inf_{\bar{u} \in \bar{U}} \psi_{(u_{k-1}, \bar{u})}(r_{k-1}) \quad \text{for } k = 1, 2, \ldots \tag{11}$$

converges to zero. From the bound (10) in the case when $k = 0$, we deduce that

$$\|x_1 - \bar{x}\| \le \inf_{\bar{u} \in \bar{U}} \psi_{(u_0, \bar{u})}(\|x_0 - \bar{x}\|) = \psi_{(u_0, \bar{u})}(r_0) = r_1$$

Likewise, from the bound (10) in the case when $k = 1$, we deduce that

$$\|x_2 - \bar{x}\| \le \inf_{\bar{u} \in \bar{U}} \psi_{(u_1, \bar{u})}(\|x_1 - \bar{x}\|) \le \inf_{\bar{u} \in \bar{U}} \psi_{(u_1, \bar{u})}(r_1) = r_2,$$

where the second inequality follows from the preceeding bound and fact that ψ is nondecreasing on $\mathbb{R}^+$. We can continue in this way inductively to deduce the general bound

$$\|x_k - \bar{x}\| \le r_k \quad \text{for } k = 0, 1, 2, \ldots$$

Since we know that the sequence of scalars $\{r_k\}$ converges to zero, we conclude that $x_k \to \bar{x}$.

Remark. If the convergence functions $\psi_{(u, \bar{u})}$ in (10) satisfy

$$\psi_{(u, \bar{u})}(r) \le r \quad \text{for all } u \in U, \ \ \bar{u} \in \bar{U}, \ \ \text{and} \ \ r \in [0, \bar{r}],$$

then as long as the initial point x_0 is close enough to $\bar{x}$, all iterates are at least as close to $\bar{x}$ as the initial point, in which case the requirement that $\{x_k\} \subseteq \ (\bar{x}; \epsilon_x)$ does not restrict the algorithm.

2.2. Target set nonsingularity and continuity. In Levy [23], connections were shown between a generalized continuity property and target nonsingularity (for a single parameter target $\bar{u}$), and this too can be extended to cover the more general case of parameter target sets. The continuity property we need is a parametric version of "selection calmness" for set-valued mappings (also sometimes called "local upper Lipschitz continuity" and "upper Lipschitz continuity at a point"). We say that a set-valued mapping $Z: \mathcal{X} \times \mathcal{Y} \rightrightarrows \mathcal{Z}$ has *calm selections near* $((\bar{x}, \bar{y}), \bar{z})$ if there exist scalars $\alpha \in (0, \infty)$, $\epsilon_x \ge 0$, $\epsilon_y \ge 0$, and $\epsilon_z \ge 0$ such that

$$Z(x, y) \cap \ (\bar{z}; \epsilon_z) \subseteq \ (\bar{z}; \alpha\|(x, y) - (\bar{x}, \bar{y})\|) \quad \forall (x, y) \in \ (\bar{x}; \epsilon_x) \times \ (\bar{y}; \epsilon_y)$$

To connect the target set nonsingularity bound with this continuity property, we extend the definition of selection calmness by considering parameterized families of set-valued mappings and replacing the term on the right side of calmness bound with a bivariate convergence function. A family of set-valued mappings $Z_{(u,u')}\colon \mathscr{X}\times\mathscr{Y}\rightrightarrows\mathscr{Z}$ parameterized by $(u,u')\in\mathscr{U}\times\mathscr{U}$ is said to have *ϕ-calm selections near $((\bar{x},\bar{y}),\bar{z})$ for parameter target set* $\bar{U}$ if there is a parameter set $U\subseteq\mathscr{U}$ containing $\bar{U}$, and scalars $\epsilon_x\geq 0$, $\epsilon_y\geq 0$, and $\epsilon_z\geq 0$ such that the family of bivariate convergence functions $\phi_{(u,u')}$ parameterized by $(u,u')\in U\times U$ satisfies

$$\sup_{z\in Z_{(u,\bar{u})}(x,y)\cap\ (\bar{z};\epsilon_z)}\|z-\bar{z}\|\leq\phi_{(u,\bar{u})}(\|x-\bar{x}\|,\|y-\bar{y}\|)$$
$$\forall\,(u,\bar{u},x,y)\in U\times\bar{U}\times\ (\bar{x};\epsilon_x)\times\ (\bar{y};\epsilon_y)\qquad(12)$$

When the target set is a singleton $\bar{U}:=\{\bar{u}\}$, the ϕ-calmness bound (12) is equivalent to the inclusion

$$Z_{(u,\bar{u})}(x,y)\cap\ (\bar{z};\epsilon_z)\subseteq\ (\bar{z};\phi_{(u,\bar{u})}(\|x-\bar{x}\|,\|y-\bar{y}\|))\quad\forall\,(u,x,y)\in U\times\ (\bar{x};\epsilon_x)\times\ (\bar{y};\epsilon_y),$$

which is precisely the inclusion used in Levy [23] to define selection ϕ-calmness in the case of a single parameter target $\bar{u}$. Moreover, it is easy to see that the original selection calmness property for an unparameterized set-valued mapping $Z\colon\mathbb{R}^n\times\mathbb{R}^m\rightrightarrows\mathscr{Z}$ is covered in this case by $\phi_{(u,u')}(r,r'):=\alpha\sqrt{r^2+(r')^2}$.

The next result shows a relationship between selection ϕ-calmness and target set nonsingularity of the generic iteration mapping over the restricted parameter target set

$$\bar{U}_{(u,x,x',\bar{x})}:=\bar{U}\cap\sigma(u,\ ,x,x',\bar{x})^{-1}(\ (0;\epsilon_y))\qquad(13)$$

defined in terms of the inverse image in the second component of σ of the ϵ_y-ball about the origin.

Proposition 2.1. *The following are equivalent*:

- *The family of partial inverse mappings* $X_{(u,u')}\colon\mathscr{X}\times\mathscr{Y}\rightrightarrows\mathscr{X}$ *defined by*

$$X_{(u,u')}(x,y):=G(u,x)\cap\sigma(u,u',x,\ ,\bar{x})^{-1}(y)\qquad(14)$$

has ϕ^2-calm selections near $((\bar{x},0),\bar{x})$ for parameter target set $\bar{U}$.

- *There exist scalars $\epsilon_x\geq 0$ and $\epsilon_y\geq 0$, and a parameter set $U\subseteq\mathscr{U}$ containing $\bar{U}$ such that any trio $(u,x,x')\in U\times\ (\bar{x};\epsilon_x)\times G(u,x)\cap\ (\bar{x};\epsilon_x)$ satisfies the target set nonsingularity bound in the Q-convergence condition with σ, ϕ^2, and restricted parameter target set* (13).

Proof. Since $x'\in X_{(u,\bar{u})}(x,y)$ if, and only if, $x'\in G(u,x)$ and $y:=\sigma(u,\bar{u},x,x',\bar{x})$, the selection ϕ^2-calmness assumption in this case is equivalent to

$$\sup_{x'\in G(x,y)\cap\ (\bar{x};\epsilon_x)}\|x'-\bar{x}\|\leq\phi^2_{(u,\bar{u})}(\|x-\bar{x}\|,\|\sigma(u,\bar{u},x,x',\bar{x})\|)$$

holding for all $(u,\bar{u},x)\in U\times\bar{U}\times\ (\bar{x};\epsilon_x)$ satisfying $\|\sigma(u,\bar{u},x,x',\bar{x})\|\leq\epsilon_y$. This is the same as the bound

$$\|x'-\bar{x}\|\leq\inf_{\bar{u}\in\bar{U}_{(u,x,x',\bar{x})}}\phi^2_{(u,\bar{u})}(\|x-\bar{x}\|,\|\sigma(u,\bar{u},x,x',\bar{x})\|)$$

holding for all trios $(u,x,x')\in U\times\ (\bar{x};\epsilon_x)\times G(u,x)\cap\ (\bar{x};\epsilon_x)$ as claimed.

Remark. Recall from the statement of Theorem 2.1 that what we would really like from Proposition 2.1 is the target set nonsingularity bound holding over the parameter target set $\bar{U}$ and not just its restriction (13). However, in practice, it turns out that these parameter sets can often be assumed equal by shrinking U and ϵ_x if necessary, so that the target set nonsingularity bound provided by Proposition 2.1 is enough to make exactly the same conclusions as in Theorem 2.1.

Note that Levy [23, Proposition 4.1.2] is covered by Proposition 2.1 when $\bar{U} := \{\bar{u}\}$ since then the restricted set (13) is just $\{\bar{u}\}$ whenever the image of $X_{(u,\bar{u})}(x, y)$ is nonempty.

2.3. Convergence rate for successive approximation methods. We have already noted that the convergence analysis of the generic algorithm for solving inclusions can be applied to study the convergence of successive approximation methods by using the identities

$$\begin{aligned} f(x) &:= -\nabla f_0(x) \\ F(x) &:= N_X(x) \\ u &:= (\mu, h) \\ a(u, x, x') &:= -\nabla q_{h,x}(x') \\ A(u, x, x') &:= \partial n_{\mu,x}(x') \end{aligned}$$

Theorem 2.1 will be applied directly to these to obtain the following result, which uses the (set-valued) solution mapping SolAp de ned by:

$$\operatorname{SolAp}(\mu, h, x) := \{x' \in \mathscr{X} \text{ such that } -\nabla q_{h,x}(x') \in \partial n_{\mu,x}(x')\}$$

Theorem 2.2 (Convergence of Valid Successive Approximation Methods). *Consider a successive approximation method that is valid at $\bar{x}$ for $\bar{U} := \bar{M} \times \{\bar{h}\}$. Assume the following*:

- *The family of mappings $X_{(u,u')}\colon \mathscr{X} \times \mathscr{X}^* \rightrightarrows \mathscr{X}$ de ned for $u = (\mu, h) \in \mathscr{U} := \mathscr{M} \times \mathscr{H}$ by*

$$X_{(u,u')}(x, y) := \begin{cases} \operatorname{SolAp}(\mu, h, x) & \text{if } y = \nabla f_0(\bar{x}) - \nabla f_0(x) - h[\bar{x} - x], \\ \varnothing & \text{otherwise} \end{cases} \tag{15}$$

has ϕ^2-calm selections near $((\bar{x}, 0), \bar{x})$ (with scalars $\epsilon_x \geq 0$ and $\epsilon_y \geq 0$) for parameter target set $\bar{U}$, with bivariate convergence functions $\phi^2_{(u,u')}$ parameterized by $(u, u') \in U \times U$ for $U := M \times H$ having sublinearity constant α.

- *There exists a family of convergence functions $\psi^q_{(h,h')}$ parameterized by $(h, h') \in H \times H$ such that*

$$\|\nabla f_0(\bar{x}) - \nabla f_0(x) - h[\bar{x} - x]\| \leq \psi^q_{(h,\bar{h})}(\|x - \bar{x}\|) \quad \text{for all } x \in \ \ (\bar{x}; \epsilon_x) \text{ and } h \in H \tag{16}$$

- *There exists a sequence of parameters $u_k \in U$ with $d(u_k, \bar{U}) \to 0$ that is target set appropriate for $(\psi, \|x_0 - \bar{x}\|)$ with target set $\bar{U}$ for the family of convergence functions $\psi_{(u,u')}$ de ned by*

$$\psi_{(u,u')}(r) := 2\alpha \psi^q_{(h,h')}(r) + 2\psi^2_{(u,u')}(r)$$

Then, every sequence $\{x_k\} \subseteq \ \ (\bar{x}; \epsilon_x)$ conforming to the successive approximation method for the sequence of parameters $\{u_k\}$ satis es $x_k \to \bar{x}$, where $\bar{x}$ solves the necessary conditions for optimality $-\nabla f_0(x) \in N_X(x)$ associated with the original optimization problem (1), and moreover the convergence rate is governed by

$$\|x_{k+1} - \bar{x}\| \leq \inf_{\bar{u} \in \bar{U}} \psi_{(u_k,\bar{u})}(\|x_k - \bar{x}\|) \tag{17}$$

Proof. Theorem 2.1 applies since any solution to the approximation problem (2) defining the successive approximation method is also a solution to the inclusion (3).

If we define the mapping σ by

$$\sigma(u, \bar{u}, x, x', \bar{x}) := \nabla f_0(\bar{x}) - \nabla f_0(x) - h[\bar{x} - x], \tag{18}$$

then the condition (16) gives the target set σ-approximation bound in Theorem 2.1 with (bivariate) convergence functions $\phi^1_{(u,u')}(r, r') := \psi^q_{(h,h')}(r)$.

The mappings $X_{(u,u')}$ from Proposition 2.1 work out in this case as (15), so the target set nonsingularity condition in Theorem 2.1 for the restricted parameter target set (13) follows. From the definition (18) of σ, we know that the inverse image in the expression for the restricted parameter target set (13) satisfies

$$\sigma(u, \ , x, x', \bar{x})^{-1}(\ (0; \epsilon_y)) = \begin{cases} \mathcal{U} & \text{if } \|\nabla f_0(\bar{x}) - \nabla f_0(x) - h[\bar{x} - x]\| \le \epsilon_y, \\ \varnothing & \text{otherwise} \end{cases}$$

It is clear that by shrinking H and ϵ_x if necessary, we can ensure that this inverse image is the entire parameter space $\mathcal{U}$, so the "restricted" parameter target set (13) is the same as the parameter target set $\bar{U}$. Thus, Theorem 2.1 applies in this case to give the convergence bound (17) and the convergence of $\{x_k\}$ to $\bar{x}$.

Under the assumption that the sequence of parameters $\{u_k\}$ approaches the target set $\bar{U}$, we can conclude that $\bar{x}$ solves the inclusion $-\nabla f_0(x) \in N_X(x)$ from the validity of the successive approximation method since we have already established the convergence to $(\bar{x}, \bar{x})$ of the sequence of pairs $(x, x') := (x_k, x_{k+1})$.

Remarks. Note that we could have used a parameter target set $\bar{H}$ for the matrices $h \in H$ in Theorem 2.2, however, the singleton parameter target $\bar{h}$ is sufficient in this case to cover all the important examples. For instance, in the classic case when $q_{h,x}$ is the quadratic Taylor approximation at x of a smooth objective function f_0 (so $h = \nabla^2 f_0(x)$), condition (16) is satisfied with parameter target $\bar{h} := \nabla^2 f_0(\bar{x})$, parameter set $H =$ the union of the Hessians $\nabla^2 f_0(x)$ over $x \in \ (\bar{x}; \epsilon_x)$, and convergence functions $\psi^q_{(h,h')}(r) := \epsilon r^2$ for some $\epsilon \ge 0$.

2.3.1. Example: Newton's method. One famous successive approximation method to which Theorem 2.2 applies is Newton's method for the (unconstrained) minimization of $f_0(x)$. In this case, the quadratic Taylor approximation is used for $q_{h,x}$, the approximation function $n_{\mu,x}$ is identically zero, and, consequently, the nonempty images $X_{(u,u')}(x, y)$ contain precisely the points x' for which $-\nabla f_0(x) - \nabla^2 f_0(x)[x' - x] = 0$. This, together with the fact that $\nabla f_0(\bar{x}) = 0$, generates the identity

$$X_{(u,u')}(x, y) = \begin{cases} x - \nabla^2 f_0(x)^{-1} \nabla f_0(x) & \text{if } y = -\nabla f_0(x) - \nabla^2 f_0(x)[\bar{x} - x], \\ \varnothing & \text{otherwise,} \end{cases} \tag{19}$$

where the familiar rule from Newton's method for minimization $x_{k+1} = x_k - \nabla^2 f_0(x_k)^{-1} \nabla f_0(x_k)$ appears. The traditional nonsingularity condition assuring the convergence of Newton's method for minimization is the invertibility of the Hessian $\nabla^2 f_0(\bar{x})$ at the target solution $\bar{x}$, which immediately implies that the inverse Hessians satisfy $\|\nabla^2 f_0(x)^{-1}\| \le \alpha$ for some $\alpha \in (0, \infty)$ and all $x \in \ (\bar{x}; \epsilon_x)$. Under these circumstances and according to the identity (19), any point $x' \in X_{(u,u')}(x, y)$ satisfies

$$\begin{aligned} \|x' - \bar{x}\| &= \left\|x - \nabla^2 f_0(x)^{-1} \nabla f_0(x) - \bar{x}\right\| \\ &= \left\|x + \nabla^2 f_0(x)^{-1}\left[y + \nabla^2 f_0(x)[\bar{x} - x]\right] - \bar{x}\right\| \\ &= \left\|\nabla^2 f_0(x)^{-1}[y]\right\| \\ &\le \alpha \|y\| \end{aligned}$$

This series of inequalities evidently implies that the family $X_{(u,u')}$ has ϕ^2-calm selections near $((\bar{x},0),\bar{x})$ for any parameter target set $\bar{U}$ with (bivariate) convergence functions $\phi^2_{(u,u')}(r,r') := \alpha r'$. Applying Theorem 2.2 in this case, the bound (17) translates into $\|x_{k+1}-\bar{x}\| \le 2\alpha\epsilon\|x_k-\bar{x}\|^2$, which is the familiar statement of quadratic convergence for this method.

3. Valid successive approximation methods. Recall from the Introduction that a successive approximation method is said to be *valid at* $\bar{x}$ *for* $\bar{U}$ if, whenever the sequence of parameters $\{u_k\}$ converges to the target set $\bar{U}$ and the sequence of iterates generated by the corresponding successive approximation method converges, then those iterates converge to a solution of the basic necessary optimality condition (4) associated with the original minimization problem (1). Note that validity ensures only that $\bar{x}$ solves the necessary conditions for optimality in the original minimization problem (1). To conclude that $\bar{x}$ is a minimizer, we either need more information about the data in (1) (for example, convexity of X and f_0 ensures that stationary points are minimizers) or more information about the data in the approximation problems (2) (since the iterates are minimizers of the approximating problems, their limit will be a minimizer of the original problem if the approximations are good enough).

As we will describe in detail in the following section, validity can be assured if the approximating inclusions (3) are good enough approximations of the original inclusion (4).

3.1. Valid approximating inclusions. The characterization of quality for approximating inclusions that we develop here uses the following *outer limit set*

$$\limsup_{\substack{(x,x')\to(\bar{x},\bar{x})\\ d(u,\bar{U})\to 0}} (a\cap A)(u,x,x') := \{\bar{y} \mid \exists (u,x,x',y) \text{ with } y\in(a\cap A)(u,x,x'),\ (x,x',y)\to(\bar{x},\bar{x},\bar{y}),\ \text{and } d(u,\bar{U})\to 0\}$$

of the set-valued mapping $a\cap A$ de ned by

$$(a\cap A)(u,x,x') := a(u,x,x')\cap A(u,x,x') = \begin{cases} a(u,x,x') & \text{if } a(u,x,x')\in A(u,x,x'),\\ \varnothing & \text{otherwise}\end{cases}$$

An approximating inclusion (6) is a *valid approximation of* (5) *at* $\bar{x}$ *for* $\bar{U}$ if the following two conditions hold:

- $f(\bar{x}) \in \limsup_{\substack{(x,x')\to(\bar{x},\bar{x})\\ d(u,\bar{U})\to 0}} (a\cap A)(u,x,x')$
- $\limsup_{\substack{(x,x')\to(\bar{x},\bar{x})\\ d(u,\bar{U})\to 0}} (a\cap A)(u,x,x') \subseteq F(\bar{x})$.

The notion of validity at $\bar{x}$ for $\bar{U}$ generalizes a similar concept from Levy [23] for a single parameter target point $\bar{u}$, where an approximating inclusion (6) was said to be a *valid approximation of* (5) *at* $\bar{x}$ *for* $\bar{u}$ if it satis es the criteria

$$\begin{aligned} &\text{(i) } a \text{ is continuous at } (\bar{u},\bar{x},\bar{x}) \text{ and } a(\bar{u},\bar{x},\bar{x}) = f(\bar{x})\\ &\text{(ii) } A \text{ is outer semicontinuous at } (\bar{u},\bar{x},\bar{x}) \text{ and } A(\bar{u},\bar{x},\bar{x}) \subseteq F(\bar{x}) \end{aligned} \tag{20}$$

Proposition 3.1. *If an approximating inclusion* (6) *is a valid approximation of* (5) *at* $\bar{x}$ *for* $\bar{u}$ *according to the de nition* (20) *from Levy* [23], *then it is also a valid approximation of* (5) *at* $\bar{x}$ *for* $\bar{U} := \{\bar{u}\}$ *as long as there exists a sequence* $(u,x,x')\to(\bar{u},\bar{x},\bar{x})$ *for which* $a(u,x,x')\in A(u,x,x')$.

Proof. Because of the existence of a sequence $(u,x,x')\to(\bar{u},\bar{x},\bar{x})$ for which $y := a(u,x,x')\in A(u,x,x')$, we can conclude from condition (i) of (20) that $y\to f(\bar{x})$. Since in this case, the stipulation $d(u, \bar{U}) \to 0$ is the same as convergence $u \to \bar{u}$, we conclude that $f(\bar{x})$ is an outer limit point of $a \cap A$ at $(\bar{u}, \bar{x}, \bar{x})$, which veri es the rst condition for target set validity.

To show the second criterion for target set validity, we consider any element $\bar{y}$ of the outer limit set

$$\bar{y} \in \limsup_{\substack{(x,x') \to (\bar{x},\bar{x}) \\ d(u,\bar{U}) \to 0}} (a \cap A)(u, x, x')$$

By the de nition of the outer limit, there exists a sequence of trios $\{(u, x, x')\}$ converging to $(\bar{u}, \bar{x}, \bar{x})$ (again $d(u, \bar{U}) \to 0$ is equivalent to $u \to \bar{u}$ since $\bar{U} = \{\bar{u}\}$) with a corresponding sequence $y \in a(u, x, x') \cap A(u, x, x')$ that satis es $y \to \bar{y}$. It follows from condition (i) in (20) that $\bar{y} = f(\bar{x})$, so condition (ii) in (20) implies that $\bar{y} = f(\bar{x}) \in A(\bar{u}, \bar{x}, \bar{x}) \subseteq F(\bar{x})$. Since $\bar{y}$ was an arbitrary element of the outer limit set, we conclude that second condition for validity at $\bar{x}$ for $\bar{U}$ also holds.

Remark. Notice that the assumption in Proposition 3.1 of the existence of a sequence $(u, x, x') \to (\bar{u}, \bar{x}, \bar{x})$ for which $a(u, x, x') \in A(u, x, x')$ is automatically assured when validity is used in Levy [23, Theorem 3.2] to establish that the limit $\bar{x}$ of the iterates conforming to the generic algorithm for solving inclusions is a solution to the inclusion (5).

Notice also that the new de nition of a valid approximation of (5) at $\bar{x}$ for $\bar{U}$ in the case of a singleton target $\bar{U} = \{\bar{u}\}$ is generally a weaker condition than the original de nition of validity from Levy [23]. In this sense, the new de nition is an improvement over the original even in the setting native to the original de nition.

Proposition 3.2. *If the approximating inclusion* (6) *is a valid approximation of* (5) *at $\bar{x}$ for $\bar{U}$, then $\bar{x}$ solves the original inclusion* (5).

Proof. This follows since the combination of inclusions indicated by target set validity immediately gives $f(\bar{x}) \in F(\bar{x})$.

Remark. Note that in Levy [23, Theorem 3.2], the parameter sequence $\{u_k\}$ was assumed to converge to the single parameter target $\bar{u}$ to conclude that $\bar{x}$ solves the inclusion (5), but in Proposition 3.2, no such assumption is necessary to make the same conclusion.

3.2. Suf cient condition or validity of successive approximation methods. The validity of a successive approximation method is closely connected to the question of whether the approximating inclusion

$$-\nabla q_{h,x}(x') \in \partial n_{\mu,x}(x') \tag{21}$$

is a valid approximation of

$$-\nabla f_0(x) \in N_X(x) \tag{22}$$

at $\bar{x}$ for $\bar{U} := \bar{M} \times \{\bar{h}\}$. In this case, the two conditions de ning validity of the approximating inclusions are

- $-\nabla f_0(\bar{x}) \in \limsup_{\substack{(x,x') \to (\bar{x},\bar{x}) \\ d(u,\bar{U}) \to 0}} (-\nabla q_{h,x}(x') \cap \partial n_{\mu,x}(x'))$
- $\limsup_{\substack{(x,x') \to (\bar{x},\bar{x}) \\ d(u,\bar{U}) \to 0}} (-\nabla q_{h,x}(x') \cap \partial n_{\mu,x}(x')) \subseteq N_X(\bar{x})$.

If the intersection set

$$-\nabla q_{h,x}(x') \cap \partial n_{\mu,x}(x') \tag{23}$$

is nonempty, then it evidently reduces to the singleton $\{-\nabla q_{h,x}(x')\}$. According to its de - nition, the quadratic approximation satis es $\nabla q_{h,x}(x') = \nabla f_0(x) + h[x' - x]$, which has limit point $\nabla f_0(\bar{x})$ for any sequence $\{(u, x, x')\}$ with $(x, x') \to (\bar{x}, \bar{x})$ and $d(u, \bar{U}) \to 0$. Thus, if the sequence of intersection sets (23) is nonempty for some sequence $\{(u, x, x')\}$ with

$(x, x') \to (\bar{x}, \bar{x})$ and $d(u, \bar{U}) \to 0$, it has the limit point $-\nabla f_0(\bar{x})$. Therefore, the first condition defining validity follows from the nonemptiness of the intersection sets (23) for some sequence $\{(u, x, x')\}$ with $(x, x') \to (\bar{x}, \bar{x})$ and $d(u, \bar{U}) \to 0$. Such nonemptiness is presumed in the definition of a valid successive approximation method, so only the second condition for validity will need to be verified in the following theorem.

Theorem 3.1 (Sufficient Condition for Validity of Successive Approximation Methods). *If the inclusion*

$$\limsup_{\substack{(x,x')\to(\bar{x},\bar{x})\\ d(u,\bar{U})\to 0}} (-\nabla q_{h,x}(x') \cap \partial n_{\mu,x}(x')) \subseteq N_X(\bar{x}) \tag{24}$$

holds, then the corresponding successive approximation method is valid at $\bar{x}$ for $\bar{U}$.

Proof. To verify the validity of the successive approximation method, we presume the existence of a sequence (u_k, x_k, x_{k+1}) satisfying $d(u_k, \bar{U}) \to 0$ and $(x_k, x_{k+1}) \to (\bar{x}, \bar{x})$, where (x_k, x_{k+1}) corresponds to the successive approximation method for the parameter sequence $\{u_k\}$. Since the approximating inclusion (21) encodes the necessary optimality condition for the approximation problem defining the successive approximation method, we can conclude that the intersection sets (23) corresponding to (u_k, x_k, x_{k+1}) are nonempty. Then, it follows from the above discussion and the assumed inclusion (24) that (21) is a valid approximation of (22) at $\bar{x}$ for $\bar{U}$. The validity of the successive approximation then follows from Proposition 3.2.

Remark. The proof of Theorem 3.1 shows why we have chosen the same terminology of validity to describe successive approximation methods and approximating inclusions.

3.3. Penalty examples. Recall from the beginning of §2, the optimization problem of minimizing $f_0(x) := x_1^2 + x_2^2$ over the constraint set

$$X := \{x = (x_1, x_2) \in \mathbb{R}^2 \text{ with } x_1 + x_2 = 1\} \tag{25}$$

and the corresponding quadratic approximation function

$$q_{h,x}(x') := (x_1')^2 + (x_2')^2$$

The normal cone to X at any $\bar{x}$ in this case is

$$N_X(\bar{x}) = \begin{cases} \bigcup_{r\in\mathbb{R}} \begin{bmatrix} r \\ r \end{bmatrix} & \text{if } \bar{x} \in X, \\ \varnothing & \text{otherwise} \end{cases} \tag{26}$$

and we considered two different families of penalty functions $n_{\mu,x}$ in §2.

3.3.1. Exact penalty example. The first family of functions $n_{\mu,x}$ we considered in §2 was the family of exact penalty functions defined by (8):

$$n_{\mu,x}(x') := \mu|x_1' + x_2' - 1|$$

The subgradients of the $n_{\mu,x}$ at x' in this case are given by

$$\partial n_{\mu,x}(x') = \begin{cases} \begin{bmatrix} \mu \\ \mu \end{bmatrix} & \text{if } x_1' + x_2' > 1 \\ \bigcup_{r\in[-\mu,\mu]} \begin{bmatrix} r \\ r \end{bmatrix} & \text{if } x_1' + x_2' = 1\,, \\ -\begin{bmatrix} \mu \\ \mu \end{bmatrix} & \text{if } x_1' + x_2' < 1 \end{cases} \tag{27}$$

so the intersection sets in the outer limit set on the left side of (24) are given by

$$-\nabla q_{h,x}(x') \cap \partial n_{\mu,x}(x') = \begin{cases} \begin{bmatrix} \mu \\ \mu \end{bmatrix} & \text{if } x'_1 + x'_2 > 1 \text{ and } \mu = -2x'_1 = -2x'_2 \\ -\begin{bmatrix} 1 \\ 1 \end{bmatrix} & \text{if } x'_1 = x'_2 = \frac{1}{2} \text{ and } \mu \geq 1 \\ -\begin{bmatrix} \mu \\ \mu \end{bmatrix} & \text{if } x'_1 + x'_2 < 1 \text{ and } \mu = 2x'_1 = 2x'_2, \\ \varnothing & \text{otherwise} \end{cases} \tag{28}$$

As long as the normal cone $N_X(\bar{x})$ is nonempty, it is clear that the intersection sets in (28) are contained in $N_X(\bar{x})$, so the closedness of the normal cone in this case ensures that the corresponding outer limit set in (24) is also contained in $N_X(\bar{x})$. According to Theorem 3.1 then, the successive approximation method using the exact penalty functions $n_{\mu,x}$ de ned by (8) is valid at any $\bar{x} \in X$ for $\bar{U} := \bar{M} \times \{\nabla^2 f_0(\bar{x})\}$ with $\bar{M}$ any parameter target set.

3.3.2. Inexact penalty example. The second family of functions $n_{\mu,x}$ we considered in §2 was the family of inexact penalty functions de ned by (9):

$$n_{\mu,x}(x') := \mu(x'_1 + x'_2 - 1)^2$$

There is a single subgradient associated with each function in this family, and it is given by

$$\partial n_{\mu,x}(x') = \begin{bmatrix} 2\mu(x'_1 + x'_2 - 1) \\ 2\mu(x'_1 + x'_2 - 1) \end{bmatrix}, \tag{29}$$

so the intersection sets in the outer limit set on the left side of (24) are given by

$$-\nabla q_{h,x}(x') \cap \partial n_{\mu,x}(x') = \begin{cases} -\begin{bmatrix} 2\mu/(1+2\mu) \\ 2\mu/(1+2\mu) \end{bmatrix} & \text{if } x'_1 = x'_2 = \mu/(1+2\mu) \text{ and } \mu \neq -\frac{1}{2}, \\ \varnothing & \text{otherwise} \end{cases} \tag{30}$$

As long as the normal cone $N_X(\bar{x})$ is nonempty, it is clear that the intersection sets in (28) are contained in $N_X(\bar{x})$, so the closedness of the normal cone in this case ensures that the corresponding outer limit set in (24) is also contained in $N_X(\bar{x})$. According to Theorem 3.1 then, the successive approximation method, using the exact penalty functions $n_{\mu,x}$ de ned by (8) is valid at any $\bar{x} \in X$ for $\bar{U} := \bar{M} \times \{\nabla^2 f_0(\bar{x})\}$ with $\bar{M}$ any parameter target set.

Both of the examples given so far have involved families of penalty functions that are independent of the variable x. In the next subsections, we give examples involving well-known classes of functions, which are neither of these things.

3.4. SQP example. The SQP method for solving nonlinear programs is one of the most famous methods in numerical optimization. The standard nonlinear program is

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

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

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

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

$$\min f_0(x) + \delta_K(g(x)) \quad \text{over } x \in \mathbb{R}^n \tag{31}$$

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

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

The composite indicator term $\delta_K(g(x))$ in (31) is evidently a nonsmooth penalty function, which enforces the constraints $x \in X$ from the constrained version of the problem.

The classical SQP method takes a current primal-multiplier pair $(x_k, \mu_k) \in \mathbb{R}^n \times \mathbb{R}^m$ and solves the following approximation problem for the next primal iterate $x_{k+1} = x'$:

$$\min f_0(x_k) + \nabla f_0(x_k)[x' - x_k] + \tfrac{1}{2}[x' - x_k]^T \quad \nabla^2 f_0(x_k)[x' - x_k] + n_{\mu_k, x_k}(x'),$$

where the family of functions n_{μ_k, x_k} is de ned by

$$n_{\mu, x}(x') := \tfrac{1}{2}[x' - x]^T \quad \nabla^2 g_\mu(x)[x' - x] + \delta_K(\nabla g(x)[x' - x] + g(x)) \tag{32}$$

in terms of the functions

$$g_\mu(x) := \mu^T \quad (f_1(x), \ldots, f_m(x)) \tag{33}$$

The multiplier is then updated by adding the multiplier associated with the solution x' to the approximation problem and the process is repeated. However, our successive approximation methods allow general parameter updating, so we will consider a generalization of the classical SQP method, where only the updating of the primal variables x is speci ed.

3.4.1. The primal SQP method. Given a current iterate x_k and parameter μ_k, choose the next iterate $x_{k+1} = x'$ by solving the (unconstrained) approximation problem

$$\min f_0(x_k) + \nabla f_0(x_k)[x' - x_k] + \tfrac{1}{2}[x' - x_k]^T \quad \nabla^2 f_0(x_k)[x' - x_k] + n_{\mu_k, x_k}(x') \quad \text{over all } x' \in \mathbb{R}^n,$$

where the family of functions $n_{\mu, x}$ is de ned by (32).

For the classical SQP method with the usual multiplier updating, the multipliers typically converge to the multiplier $\bar{\mu}$ associated with the primal solution. The following lemma shows that a familiar constraint quali cation ensures that the primal SQP method is a valid successive approximation method at $\bar{x}$ for $\bar{U} := \bar{M} \times \{\nabla^2 f_0(\bar{x})\}$ with $\bar{M}$ any bounded parameter target set.

Lemma 3.1. *Under the Mangasarian-Fromovitz constraint quali cation at* $\bar{x}$:

$$\mu \in N_K(g(\bar{x})) \qquad \text{and} \qquad \nabla g(\bar{x})^T \quad \mu = 0 \Rightarrow \mu = 0,$$

the primal SQP method is valid at $\bar{x}$ *for* $\bar{U} := \bar{M} \times \{\nabla^2 f_0(\bar{x})\}$ *with* $\bar{M}$ *any bounded parameter target set.*

Proof. Under the Mangasarian-Fromovitz constraint quali cation, Rockafeller and Wets [25, Example 10.8] gives the subgradients of $n_{\mu, x}$ for x near $\bar{x}$ as

$$\partial n_{\mu, x}(x') = \nabla^2 g_\mu(x)[x' - x] + \nabla g(x)^T \quad N_K(\nabla g(x)[x' - x] + g(x))$$

as well as the formula

$$N_X(\bar{x}) = \nabla g(\bar{x})^T \quad N_K(g(\bar{x})) \tag{34}$$

for the normal cone to X at $\bar{x}$. According to Theorem 3.1, we need to verify the inclusion (24), and evidently, the intersection sets in the outer limit set on the left side of (24) in this case are given by

$$(-\nabla f_0(x) - \nabla^2 f_0(x)[x'-x]) \cap \;\nabla^2 g_\mu(x)[x'-x] + \nabla g(x)^T \quad N_K(\nabla g(x)[x'-x] + g(x))\big) \tag{35}$$

From the de nition of the functions g_μ in (33) and the fact that the parameter target set $\bar{M}$ is bounded, it follows that the term $\nabla^2 g_\mu(x)[x'-x]$ in the intersection (35) approaches zero as $(x, x') \to (\bar{x}, \bar{x})$ and $d(u, \bar{U}) \to 0$. The outer limit set of the intersection sets (35) is thus contained in the outer limit set

$$\limsup_{(x,x')\to(\bar{x},\bar{x})} \nabla g(x)^T \quad N_K(\nabla g(x)[x'-x] + g(x))$$

This, together with the formula (34), implies that the inclusion (24) from Theorem 3.1 follows in this case from the inclusion

$$\limsup_{(x,x')\to(\bar{x},\bar{x})} \nabla g(x)^T \quad N_K(\nabla g(x)[x'-x] + g(x)) \subseteq \nabla g(\bar{x})^T \quad N_K(g(\bar{x})) \tag{36}$$

We can deduce this inclusion from the Mangasarian-Fromovitz constraint quali cation as follows: Any element z in the outer limit set on the left side of the inclusion (36) is the limit point of a sequence $\nabla g(x)^T \quad v$ for vectors $v \in N_K(\nabla g(x)[x'-x] + g(x))$. If $z = 0$, it is trivially in the set $\nabla g(\bar{x})^T \quad N_K(g(\bar{x}))$ since the zero vector is always in the normal cone. We therefore suppose $z \neq 0$.

Case 1. The vectors v are bounded. Then, at least some subsequence of them has a limit point $\bar{v}$, which must satisfy $z = \nabla g(\bar{x})^T \quad \bar{v}$. Moreover, the outer semicontinuity of the normal cone mapping (c.f. Rockafeller and Wets [25, Proposition 6.6]) implies that $\bar{v} \in N_K(g(\bar{x}))$, so the inclusion (36) holds in this case.

Case 2. The vectors v are not bounded. Then, due to the fact that N_K is a cone, we know that the unit vectors $p := v/\|v\|$ are also in the normal cone $N_K(\nabla g(x)[x'-x] + g(x))$. Moreover, at least some subsequence of them has a limit point $\bar{p}$, which is a unit vector and must satisfy $\bar{p} \in N_K(g(\bar{x}))$ because of the outer semicontinuity of the normal cone mapping. However, from the convergence $\nabla g(x)^T \quad v \to z$, we know that

$$\lim \nabla g(x)^T \quad p = \lim \nabla g(x)^T \quad v/\|v\| = \lim z/\|v\| = 0$$

This implies that $\nabla g(x)^T \quad \bar{p} = 0$, which contradicts the Mangasarian-Fromovitz constraint quali cation at $\bar{x}$.

Remark. Note that our primal SQP method is still a valid successive approximation method when more general functions g_μ than the classical ones (33) are used, as long as the terms $\nabla^2 g_\mu(x)[x'-x]$ in the intersection (35) still approach zero as $(x, x') \to (\bar{x}, \bar{x})$ and $d(u, \bar{U}) \to 0$. For instance, we could instead use functions of the form $g_\mu(x) := \mu\|g(x)\|^2$ for $\mu \in [0, \infty)$, which act as penalty functions in the case when all the constraints are equations (i.e., $\mathcal{I} = \varnothing$).

3.5. Augmented Lagrangian example. Augmented Lagrangian methods have been developed for solving nonlinear programs like

$$\min f_0(x) \quad \text{over } x \in X := \{x \in \mathbb{R}^n \colon f_i(x) = 0 \text{ for } i = 1, 2, \ldots, m\},$$

where the f_i are smooth functions on $\mathbb{R}^n$. The standard approach is to construct a family of "augmented Lagrangian" functions $L_\mu(x) := f_0(x) + p_\mu(x)$ from the vector-scalar parameter pairs $\mu := (\lambda, \eta) \in \mathbb{R}^m \times \mathbb{R}^+$ and the functions

$$p_\mu(x) := \sum_{i=1}^{m} (\lambda)_i f_i(x) + \eta \sum_{i=1}^{m} (f_i(x))^2, \tag{37}$$

where $(\lambda)_i$ denotes the ith component of the vector $\lambda \in \mathbb{R}^m$. For a given parameter μ_k, the (unconstrained) augmented Lagrangian function L_{μ_k} is then approximately minimized via some appropriate scheme. If Newton's method is used to minimize the augmented Lagrangian, we get the following "primal" augmented Lagrangian method (so called because only an update on the primal variable x is speci ed).

The primal augmented Lagrangian method. Given a current iterate x_k and parameter μ_k, choose the next iterate $x_{k+1} = x'$ by solving the (unconstrained) approximation problem

$$\min f_0(x_k) + \nabla f_0(x_k)[x' - x_k] + \tfrac{1}{2}[x' - x_k]^T \ \nabla^2 f_0(x_k)[x' - x_k] + n_{\mu_k, x_k}(x'),$$

where the family of functions $n_{\mu, x}$ is de ned by

$$n_{\mu, x}(x') := p_\mu(x) + \nabla p_\mu(x)[x' - x] + \tfrac{1}{2}[x' - x]^T \ \nabla^2 p_\mu(x)[x' - x] \tag{38}$$

in terms of the functions p_μ from (37).

The single subgradient associated with each of the $n_{\mu, x}$ is given by

$$\partial n_{\mu, x}(x') = \nabla p_\mu(x) + \nabla^2 p_\mu(x)[x' - x] \tag{39}$$

and the normal cone to X at $\bar{x}$ is

$$N_X(\bar{x}) = \begin{cases} \left\{ \sum_{i=1}^{m} r_i \nabla f_i(\bar{x}) \,\middle|\, r_i \in \mathbb{R} \right\} & \text{if } \bar{x} \in X, \\ \varnothing & \text{otherwise} \end{cases} \tag{40}$$

From the de nition (37) of the functions p_μ, it is clear that their gradients satisfy

$$\nabla p_\mu(x) = \sum_{i=1}^{m} (\lambda_i + 2\eta f_i(x)) \nabla f_i(x) \tag{41}$$

If the Hessian matrices $\nabla^2 p_\mu(x)$ have uniformly bounded norm for all μ near the parameter target set $\bar{M}$ and all x near $\bar{x}$, then according to (39), any limit point from the sets $\partial n_{\mu, x}(x')$ with $(x, x') \to (\bar{x}, \bar{x})$ and $d(\mu, \bar{M}) \to 0$ is just the limit of the corresponding $\nabla p_\mu(x)$. The formulas (41) and (40) ensure that such a limit point will be in the normal cone $N_X(\bar{x})$, which leads to the following lemma.

Lemma 3.2. *If the Hessian matrices $\nabla^2 p_\mu(x)$ have uniformly bounded norm for all μ near the parameter target set $\bar{M}$ and all x near $\bar{x}$, then the primal augmented Lagrangian method is valid at $\bar{x}$ for $\bar{U} := \bar{M} \times \{\nabla^2 f_0(\bar{x})\}$.*

Proof. This follows from Theorem 3.1 and the discussion preceding the statement of the lemma since any element of the intersection on the left side of (24) is automatically in the set $\partial n_{\mu, x}(x')$.

Notice that we are taking a more general approach here to parameters than is typical. The classical augmented Lagrangian method involves essentially two stages at each step of the algorithm: The inner stage involves the complete approximate minimization of the augmented Lagrangian function

$$L_\mu(x) := f_0(x) + \sum_{i=1}^{m} (\lambda)_i f_i(x) + \eta \sum_{i=1}^{m} (f_i(x))^2$$

for a fixed value of the parameter $\mu = (\lambda_k, \eta_k)$, with as many steps as are necessary to achieve an acceptable approximate minimizer. The outer stage then updates the iterate x_k to be the approximate minimizer from the inner stage, possibly increases the parameter η_{k+1} from η_k and updates the multiplier according to

$$\lambda_{k+1} := \lambda_k + 2\eta_k (f_1(x_k), \ldots, f_m(x_k)) \tag{42}$$

Since it requires no special parameter update scheme, our primal augmented Lagrangian method allows a whole array of new possibilities beyond the classical updating scheme. For instance, our method allows new schemes, where the multiplier λ is updated via (42) after each step in the inner stage approximate minimization of the Lagrangian, or perhaps after only some of the steps in the inner stage, so that the two stages from the classical approach are no longer necessarily distinct in our method. In addition, our method allows entirely different parameter update schemes having no connection to the classical formula (42).

Of course, completely arbitrary multiplier updating will not necessarily produce nice convergence results. In the next section, we apply Theorem 2.2 to analyze the convergence rate for our primal augmented Lagrangian method and, consequently, develop conditions on the parameter updating that ensure nice convergence properties.

4. Convergence rate for the primal augmented Lagrangian method. To apply Theorem 2.2 to the primal augmented Lagrangian method, we first notice that the bound (16) in this case is satisfied by the convergence functions

$$\psi^q_{(h,\,h')}(r) := \epsilon r^2 + \|h - h'\| r, \tag{43}$$

where $\epsilon \geq 0$ satisfies

$$\|\nabla f_0(\bar{x}) - \nabla f_0(x) + \nabla^2 f_0(\bar{x})[x - \bar{x}]\| \leq \epsilon \|x - \bar{x}\|^2 \quad \text{for } x \in \quad (\bar{x}; \epsilon_x),$$

since then the following series of inequalities holds for all $x \in \quad (\bar{x}; \epsilon_x)$ and $h \in H$:

$$\begin{aligned} &\|\nabla f_0(\bar{x}) - \nabla f_0(x) - h[\bar{x} - x]\| \\ &\quad = \|\nabla f_0(\bar{x}) - \nabla f_0(x) + \nabla^2 f_0(\bar{x})[x - \bar{x}] + (h - \nabla^2 f_0(\bar{x}))[x - \bar{x}]\| \\ &\quad \leq \|\nabla f_0(\bar{x}) - \nabla f_0(x) + \nabla^2 f_0(\bar{x})[x - \bar{x}]\| + \|(h - \nabla^2 f_0(\bar{x}))[x - \bar{x}]\| \\ &\quad \leq \epsilon \|x - \bar{x}\|^2 + \|h - \nabla^2 f_0(\bar{x})\| \|x - \bar{x}\| \end{aligned}$$

The approximating inclusion $-\nabla q_{h,\,x}(x') \in \partial n_{\mu,\,x}(x')$ defining our primal augmented Lagrangian method is equivalent to the equation

$$-\nabla f_0(x) - h[x' - x] = \nabla p_\mu(x) + \nabla^2 p_\mu(x)[x' - x], \tag{44}$$

which can be immediately rewritten in terms of any fixed $\bar{x} \in \mathbb{R}^n$ as

$$(h + \nabla^2 p_\mu(x))[x' - \bar{x}] = -\nabla f_0(x) - h[\bar{x} - x] - \nabla p_\mu(x) - \nabla^2 p_\mu(x)[\bar{x} - x]$$

By adding and subtracting the term $\nabla f_0(\bar{x}) + \nabla p_\mu(\bar{x})$ on the right side of this equation, we get the equivalent expression

$$\begin{aligned} (h + \nabla^2 p_\mu(x))[x' - \bar{x}] = {} & \nabla f_0(\bar{x}) - \nabla f_0(x) - h[\bar{x} - x] + \nabla p_\mu(\bar{x}) - \nabla p_\mu(x) \\ & - \nabla^2 p_\mu(x)[\bar{x} - x] - \nabla f_0(\bar{x}) - \nabla p_\mu(\bar{x}) \end{aligned}$$

As long as the matrix $h+\nabla^2 p_\mu(x)$ is invertible with inverse having norm less than or equal to the same $\alpha \in (0,\infty)$ for all $(\mu, h, x) \in U \times \ (\bar{x}; \epsilon_x)$, we conclude that

$$\|x'-\bar{x}\| \le \alpha\|y\| + \alpha\|\nabla p_\mu(\bar{x}) - \nabla p_\mu(x) - \nabla^2 p_\mu(x)[\bar{x}-x]\| + \alpha\| -\nabla f_0(\bar{x}) - \nabla p_\mu(\bar{x})\|, \quad (45)$$

where we have substituted the identity $y = \nabla f_0(\bar{x}) - \nabla f_0(x) - h[\bar{x}-x]$ from the expression for the family of mappings $X_{(u,u')}$ de ned in (15). Because of the form of the function p_μ in this case, we know that the bound holds that

$$\|\nabla p_\mu(\bar{x}) - \nabla p_\mu(x) - \nabla^2 p_\mu(x)[\bar{x}-x]\| \le (a\|\lambda\| + b\eta)\|x-\bar{x}\|^2$$

for some scalar $a \ge 0$ corresponding to the Taylor linear approximation at $x \in \ (\bar{x}; \epsilon_x)$ of the gradient mapping associated with the mapping $x \mapsto (f_1(x), \ldots, f_m(x))$, and $b \ge 0$ corresponding to the Taylor linear approximation of the gradient mapping associated with the function $x \mapsto \sum_{i=1}^m (f_i(x))^2$. Moreover, as long as there exists a Lagrange multiplier $\bar{\lambda} \in \mathbb{R}^m$ for $\bar{x}$, we know that $-\nabla f_0(\bar{x}) = \sum_{i=1}^m \bar{\lambda}_i \nabla f_i(\bar{x})$ and $f_i(\bar{x}) = 0$, in which case the last term on the right side of (45) is bounded above by

$$\alpha\|\lambda - \bar{\lambda}\| \sum_{i=1}^m \|\nabla f_i(\bar{x})\|$$

Under these assumptions then, and according to the bound (45), the family of mappings $X_{(u,u')}(x,y)$ from Theorem 2.2 in this case has ϕ^2-calm selections for parameter target set

$$\bar{U} := \{\bar{\lambda}\} \times [\bar{\eta}, \infty) \times \{\nabla^2 f_0(\bar{x})\} \quad (46)$$

(for any $\bar{\eta} \in \mathbb{R}^+$) with bivariate convergence functions de ned by

$$\phi^2_{(u,u')}(r, r') := \alpha(a\|\lambda\| + b\eta)r^2 + \alpha c\|\lambda - \lambda'\| + \alpha r', \quad (47)$$

where $c := \sum_{i=1}^m \|\nabla f_i(\bar{x})\|$.

According to Theorem 2.2 and the preceding discussion of this example, we need to consider sequences of parameters that are target set appropriate for $(\psi, \|x_0 - \bar{x}\|)$ for the convergence functions de ned by

$$\psi_{(u,u')}(r) := 2\alpha(\epsilon + a\|\lambda\| + b\eta)r^2 + 2\alpha\|h-h'\|r + 2\alpha c\|\lambda - \lambda'\|$$

in terms of $a \ge 0$, $b \ge 0$, and $c \ge 0$ from (47) and $\epsilon \ge 0$ from (43). Thus, we need to establish the convergence to zero of the sequence of scalars de ned by $r_0 := \|x_0 - \bar{x}\|$ and

$$r_k := 2\alpha(\epsilon + a\|\lambda_{k-1}\| + b\eta_{k-1})r_{k-1}^2 + 2\alpha\|h_{k-1} - \nabla^2 f_0(\bar{x})\|r_{k-1} + 2\alpha c\|\lambda_{k-1} - \bar{\lambda}\|, \quad (48)$$

where the in mum over the parameter target set $\bar{U}$ is unnecessary since only the η-component of $\bar{U}$ is nonsingleton, and ψ does not depend on η'. To conclude from (48) that $r_k \to 0$, we need rst the following lemma for general quadratic iteration functions.

LEMMA 4.1. *Consider a family of quadratic functions q_k: $\mathbb{R} \to \mathbb{R}$ indexed by $k = 0, 1, 2, \ldots$ and de ned by $q_k(r) := A_k r^2 + B_k r + C_k$. Assume that there exist $A \in (0, \infty)$ and $B \in [0, 1)$ such that the coef cients satisfy $A_k \in (0, A]$, $B_k \in [0, B]$, and $C_k \in [0, (1-B)^2/(4A))$ for all k. If, in addition, $C_k \to 0$ and $r_0 \in [0, \bar{r}]$ for $\bar{r} := (1-B)/(2A)$, then the sequence of scalars de ned by $r_k = q_{k-1}(r_{k-1})$ converges to zero.*

PROOF. First, notice that since $C_k \to 0$ and C_k are strictly less than $(1-B)^2/(4A)$ for all k, we can assume that the coef cients C_k are actually all less than or equal to

$(1-B)^2/(4A)-\epsilon$ for some small $\epsilon>0$. It follows that the family of quadratic functions satis es

$$q_k(r)\in[0,\bar{r}-\epsilon]\quad \text{for all } r\in[0,\bar{r}] \text{ and all } k, \tag{49}$$

so that $r_k\in[0,\bar{r}-\epsilon]$ for all $k\geq 1$. If we de ne

$$r_k^* := \frac{(1-B_k)-\sqrt{(B_k-1)^2-4A_kC_k}}{2A_k},$$

it follows that $q_k(r_k^*)=r_k^*$ and that $r_k^*\to 0$. Since r_k^* is evidently in the interval $[0,\bar{r}]$, it follows from (49) and the fact that $r_k^*=q_k(r_k^*)$ that r_k^* is actually in the interval $[0,\bar{r}-\epsilon]$ for all k. From the mean value theorem, we get the following series of inequalities for all $k\geq 1$:

$$\begin{aligned}\|r_{k+1}-r_k^*\| &= \|q_k(r_k)-q_k(r_k^*)\|\\ &= \|q_k'(\xi_k)(r_k-r_k^*)\| \quad \text{for some } \xi_k\in[0,\bar{r}-\epsilon]\\ &\leq (1-2A\epsilon)\|r_k-r_k^*\|\end{aligned}$$

If we combine this with the fact that

$$\|r_{k+1}-r_{k+1}^*\|\leq \|r_{k+1}-r_k^*\|+\|r_k^*-r_{k+1}^*\|,$$

we get

$$\|r_{k+1}-r_{k+1}^*\|\leq (1-2A\epsilon)\|r_k-r_k^*\|+\|r_k^*-r_{k+1}^*\| \tag{50}$$

Since the sequence $\{r_k^*\}$ converges to zero, we know that $\|r_k^*-r_{k+1}^*\|\to 0$, so we can conclude from (50) that the sequence of differences $\{r_k-r_k^*\}$ converges to zero. Moreover, since $r_k^*\to 0$, this means that the sequence $\{r_k\}$ also converges to zero as claimed.

From Lemma 4.1 and the expression (48) for r_k as the output of a quadratic function at r_{k-1}, we can conclude that the sequence $\{r_k\}$ converges to zero as long as the following conditions are met:

(51)

- $\exists\delta_1\in[0,1/(2\alpha))$ such that $\|h_k-\nabla^2 f_0(\bar{x})\|\leq\delta_1 \quad \forall k=0,1,2,\ldots$
- $\exists\delta_2\in\mathbb{R}^+$ such that $\|\lambda_k\|\leq\delta_2 \quad \forall k=0,1,2,\ldots$
- $\exists\delta_3\in\mathbb{R}^+$ such that $\eta_k\leq\delta_3 \quad \forall k=0,1,2,\ldots$
- $\lambda_k\to\bar{\lambda}$ with $\|\lambda_k-\bar{\lambda}\|<\dfrac{(1-2\alpha\delta_1)^2}{16\alpha^2c(\epsilon+a\delta_2+b\delta_3)} \quad \forall k=0,1,2,\ldots$
- $r_0:=\|x_0-\bar{x}\|\leq\dfrac{1-2\alpha\delta_1}{4\alpha(\epsilon+a\delta_2+b\delta_3)}$

To conclude, we state the following corollary to Theorem 2.2, which consolidates what we have discovered about the primal augmented Lagrangian method.

Corollary 4.1 (Convergence of the Primal Augmented Lagrangian Method). *Assume that the following constraint quali cation holds*:

- *There exists a Lagrange multiplier* $\bar{\lambda}\in\mathbb{R}^m$ *for* $\bar{x}$,

and that there exists a scalar $\epsilon_x\geq 0$ *and a set* $U\subseteq\mathbb{R}^m\times\mathbb{R}^+\times\mathbb{R}^{n\times n}$ *containing the target set* $\bar{U}$ *de ned in* (46) *for some* $\bar{\eta}\in\mathbb{R}^+$ *such that the following generalized second-order condition holds*:

- *There exists an* $\alpha\in(0,\infty)$ *such that for all* $(\mu,h,x)\in U\times\ (\bar{x};\epsilon_x)$, *the matrix* $h+\nabla^2p_\mu(x)$ *is invertible with inverse having norm less than or equal to* α.

If the conditions (51) *hold, then any sequence of iterates* $\{x_k\}\subseteq\ (\bar{x};\epsilon_x)$ *generated by the primal augmented Lagrangian method for parameters* $u_k:=(\lambda_k,\eta_k,h_k)$ *with* $d(u_k,\bar{U})\to 0$

satis es $x_k \to \bar{x}$, *where* $\bar{x}$ *solves* $-\nabla f_0(x) \in N_X(x)$, *and moreover the convergence rate is governed by*

$$\|x_{k+1} - \bar{x}\| \le 2\alpha(\epsilon + a\|\lambda_k\| + b\eta_k)\|x_k - \bar{x}\|^2 + 2\alpha\|h_k - \nabla^2 f_0(\bar{x})\|\|x_k - \bar{x}\| + 2\alpha c\|\lambda_k - \bar{\lambda}\|$$

for the scalars $a \ge 0$, $b \ge 0$, *and* $c \ge 0$ *from* (47) *and* $\epsilon \ge 0$ *from* (43).

Proof. This follows from Lemma 3.2 and Theorem 2.2 since the assumption in Lemma 3.2 that the Hessian matrices $\nabla^2 p_\mu(x)$ have uniformly bounded norm for all μ near the parameter target set $\bar{M}$ and all x near $\bar{x}$ follows from the generalized second-order condition here.

Remark. Corollary 4.1 complements and extends existing results for the classical augmented Lagrangian method. For instance, see Nocedal and Wright [24, Theorem 17.6] where the standard choice of $h_k := \nabla^2 f_0(x_k)$ is used, the stronger constraint quali cation of linearly independent $\nabla f_i(\bar{x})$ is assumed, and the standard second-order suf cient condition that the Hessian in x at $(\bar{x}, \bar{\lambda})$ of the standard Lagrangian is positive de nite on the critical cone. In the proof of Nocedal and Wright [24, Theorem 17.5], these conditions are shown to imply the uniform positive de niteness (and hence invertibility) of the matrix $\nabla^2 f_0(\bar{x}) + \nabla^2 p_{\bar{\lambda},\eta}(x)$ for all $\eta \in [\bar{\eta}, \infty)$. In this situation, our second-order condition in Corollary 4.1 follows from the continuity of the mapping $(\mu, h, x) \mapsto (h + \nabla^2 p_\mu(x))$. The existing results like Nocedal and Wright [24, Theorem 17.6] for augmented Lagrangian methods use the standard multiplier update based on (42), but our Corollary 4.1 provides convergence results for *any* sequence of parameters approaching the target set.

Acknowledgments. The author would like to thank Jong-Shi Pang and the anonymous referees whose suggestions much improved the nal form of this paper.

References

[1] Ben-Tal, A., M. Zibulevsky. 1997. Penalty/barrier multiplier methods for convex programming problems. *SIAM J. Optim.* **7** 347–366.

[2] Boggs, P., J. Tolle. 1995. Sequential quadratic programming. *Acta Numerica*, Vol. 4. Cambridge University Press, Cambridge, UK, 1–51.

[3] Boukari, D., A. V. Fiacco. 1995. Survey of penalty, exact-penalty and multiplier methods from 1968 to 1993. *Optimization* **32** 301–334.

[4] Byrd, R., N. Gould, J. Nocedal, R. Waltz. 2004. An active set algorithm for nonlinear programming using linear programming and equality constrained subproblems. *Math. Programming* **100** 27–48.

[5] Contaldi, G., G. DiPillo, S. Lucidi. 1993. A continuously differentiable exact penalty function for nonlinear programming problems with unbounded feasible set. *Oper. Res. Lett.* **14** 153–161.

[6] Demyanov, V. F., G. DiPillo, F. Facchinei. 1998. Exact penalization via Dini and Hadamard conditional derivatives. *Optim. Methods Software* **9** 19–36.

[7] DiPillo, G. 1994. Exact penalty methods. *Algorithms for Continuous Optimization, NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences*, Vol. 434. Kluwer Academic Publishers, Dordrecht, The Netherlands, 209–253.

[8] DiPillo, G., F. Facchinei. 1992. Regularity conditions and exact penalty functions in Lipschitz programming problems. *Nonsmooth Optimization: Methods and Applications*. Gordon and Breach, Montreux, Switzerland, 107–120.

[9] DiPillo, G., F. Facchinei. 1995. Exact barrier function methods for Lipschitz programs. *Appl. Math. Optim.* **32** 1–31.

[10] DiPillo, G., F. Facchinei, L. Grippo. 1992. An RQP algorithm using a differentiable exact penalty function for inequality constrained problems. *Math. Programming Ser. A* **55** 49–68.

[11] DiPillo, G., L. Grippo, S. Lucidi. 1993. A smooth method for the nite minimax problem. *Math. Programming* **60** 187–214.

[12] DiPillo, G., L. Grippo, S. Lucidi. 1997. Smooth transformation of the generalized minimax problem. *J. Optim. Theory Appl.* **95** 1–24.

[13] DiPillo, G., S. Lucidi, L. Palagi. 1993. An exact penalty-Lagrangian approach for a class of constrained optimization problems with bounded variables. *Optimization* **28** 129–148.

[14] DiPillo, G., S. Lucidi, L. Palagi. 1999. A shifted-barrier primal-dual algorithm model for linearly constrained optimization problems. *Comput. Optim. Appl.* **12** 157–188.

[15] Dontchev, A. L. 1996. Uniform convergence of the Newton method for Aubin continuous maps: Well-posedness and stability of variational problems. *Serdica Math. J.* **22**(3) 283–296.

[16] Dontchev, A. L. 2000. Lipschitzian stability of Newton's method for variational inclusions. *System Modelling and Optimization*. Kluwer Academic Publishers, Boston, MA, 119–147.

[17] Facchinei, F. 1991. Exact penalty functions and Lagrange multipliers. *Optimization* **22** 579–606.

[18] Facchinei, F. 1997. Robust recursive quadratic programming algorithm model with global and superlinear convergence properties. *J. Optim. Theory Appl.* **92** 543–579.

[19] Facchinei, F., S. Lucidi. 1992. A class of penalty functions for optimization problems with bound constraints. *Optimization* **26** 239–259.

[20] Facchinei, F., S. Lucidi. 1998. Convergence to second-order stationary points in inequality constrained optimization. *Math. Oper. Res.* **23** 746–766.

[21] Facchinei, F., G. Liuzzi, S. Lucidi. 2003. A truncated Newton method for the solution of large-scale inequality constrained minimization problems. *Comput. Optim. Appl.* **25** 85–122.

[22] Fletcher, R. 1987. *Practical Methods of Optimization*. John Wiley and Sons, New York.

[23] Levy, A. B. 2004. A generic algorithm for solving inclusions. *SIAM J. Optim.* **15**(2) 430–455.

[24] Nocedal, J., S. J. Wright. 1999. *Numerical Optimization*. Springer-Verlag, New York.

[25] Rockafellar, R. T., R. J.-B. Wets. 1998. *Variational Analysis*. Springer-Verlag, Berlin, Germany.

[26] Sadjadi, S., K. Ponnambalam. 1999. Advances in trust region algorithms for constrained optimization. *Appl. Numer. Math.* **29**(3) 423–443.

[27] Zoutendijk, G. 1976. *Mathematical Programming Methods*. North-Holland Publishing Company, New York.