Preprint ANL/MCS-P793-0200

NONLINEAR PROGRAMS WITH UNBOUNDED LAGRANGE MULTIPLIER SETS

MIHAI ANITESCU*

Abstract. We investigate nonlinear programs that have a nonempty but possibly unbounded Lagrange multiplier set and that satisfy the quadratic growth condition. We show that such programs can be transformed, by relaxing the constraints and adding a linear penalty term to the objective function, into equivalent nonlinear programs that have differentiable data and a bounded Lagrange multiplier set and that satisfy the quadratic growth condition. As a result we can define, for this type of problem, algorithms that are linearly convergent, using only first-order information, and superlinearly convergent.

1. Introduction. Recently, there has been renewed interest in analyzing and modifying sequential quadratic programming (SQP) algorithms for constrained nonlinear optimization for cases where the traditional regularity conditions do not hold [5, 9, 15, 14, 26, 30, 31]. This research is partly motivated by the fact that large-scale nonlinear programming problems tend to be almost degenerate (have large condition numbers for the Jacobian of the active constraints). We term as degenerate those nonlinear programs (NLPs) for which the gradients of the active constraints are linearly dependent. In this case there may be several feasible Lagrange multipliers.

In addition, there are classes of problems that are explicitly formulated as degenerate nonlinear programs and whose Lagrange multiplier set not only is not a singleton, but also is unbounded. One such type of nonlinear program is mathematical programs with equilibrium constraints, or MPECs [19, 20, 25]. The complementarity part of the equilibrium constraints generally violate the Mangasarian-Fromovitz constraint qualification (MFCQ) [22]. MFCQ, in turn, is equivalent to the boundedness of the constraints [12], which means that such MPECs will have an unbounded Lagrange multiplier set. One approach that has been proposed to deal with these constraints is to enforce the complementarity conditions by a nondifferentiable penalty term added to the objective function and to restrict explicitly the size of the Lagrange multipliers [20]. Lack of differentiability is a serious problem for defining efficient algorithms. However, in the special case where the noncomplementarity constraints are linear, it is shown that the penalty term becomes differentiable. For this particular case, the approach becomes suitable for use with a nonlinear programming algorithm. Another problem with unbounded multipliers appears in the context of model reduction for chemical kinetics [32]. There the reduction equations are enforced by equality constraints whose gradients are 0 at a solution of the problem.

Many of the previous analysis and rate of convergence results for degenerate NLP [5, 9, 15, 14, 26, 30, 31] are based on the validity of some second-order conditions,

* Thackeray 301, Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15213 (anitescu@math.pitt.edu). Part of this work was completed while the author was the Wilkinson Fellow at the Mathematics and Computer Science Division, Argonne National Laboratory. This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38. This work was also supported by award DMS-9973071 of the National Science Foundation.

which imply the existence of a locally strictly convex augmented Lagrangian. In a recent approach, it has been shown that both linear convergence, using only first-order information, and superlinear convergence can be achieved for nonlinear programs even when there does not exist any locally strictly convex augmented Lagrangian [1, 2]. The results were obtained assuming only that the nonlinear program satisfies the boundedness of the set of Lagrange multipliers and the quadratic growth condition [6].

Here we extend the results from [1, 2] to the case of unbounded sets of Lagrange multipliers. We deal with the NLP problem

$$\min_{x} f(x) \quad \text{subject to } g(x) \le 0, \tag{1.1}$$

where $f : \mathsf{R}^n \to \mathsf{R}$ and $g : \mathsf{R}^n \to \mathsf{R}^m$. In this work, we assume only that

1. At a local solution x^* of (1.1), the set of Lagrange multipliers is not empty.
2. The quadratic growth condition [6, 17] is satisfied

   $$\max\{f(x) - f(x^*), g_1(x), g_2(x), \dots, g_m(x)\} \ge \sigma \|x - x^*\|^2$$

   for x in some neighborhood of x^* and $\sigma > 0$.
3. The data of the problem, f, g, are twice continuously differentiable.

These assumptions are related to a weaker form of the second-order sufficient conditions [16, 6], which does not imply the existence of a locally convex augmented Lagrangian as is the case in [5, 9, 15, 14, 26, 30, 31].

To accommodate the case where the Lagrange multiplier set is not bounded, we modify (1.1) by relaxing the constraints and adding a penalty term to the objective function:

$$\min_{x,\zeta} f(x) + c\zeta \quad \text{subject to } g_i(x) \le \zeta, \ \ i = 1, 2, \dots, m, \ \ \zeta \ge 0. \tag{1.2}$$

The modified program (1.2) is closely related to the use of the L_∞ exact penalty function for nonlinear programming [3, 4]. Clearly, (1.2) has twice differentiable data under our assumptions. We show that for a sufficiently large parameter c, the modified nonlinear program has the local solution $(x^*, 0)$, has bounded multipliers, and satisfies a corresponding quadratic growth constraint. Therefore, the algorithms from [1, 2] can be applied to (1.2).

1.1. Previous Work and Framework. We call x a stationary point of (1.1) if the Fritz-John conditions conditions hold: There exist the multipliers $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_m) \in \mathsf{R}^{m+1}$, such that

$$\nabla_x \mathcal{L}(x, \lambda) = 0, \ \ \lambda \ge 0, \ \ g(x) \le 0, \ \ \sum_{i=1}^{m} \lambda_i g_i(x) = 0, \ \ \|\lambda\|_1 = 1. \tag{1.3}$$

Here $\mathcal{L}$ is the Lagrangian function

$$\mathcal{L}(x, \lambda) = \lambda_0 f(x) + \sum_{i=1}^{m} \lambda_i g_i(x). \tag{1.4}$$

A local solution x^* of (1.1) is a stationary point [24]. We introduce the sets of multipliers

$$(1.5) \qquad \Omega_0(x) = \left\{\lambda \in \mathsf{R}^{m+1} \mid \ \lambda \text{ satisfies (1.3) at } x \ \right\},$$

$$(1.6) \qquad \Omega_1(x) = \left\{\lambda \in \Omega_0(x) \mid \lambda_0 > 0\right\}.$$

The active set at a stationary point x is

$$(1.7) \qquad \mathcal{A}(x) = \{i = 1, 2, \ldots, m \mid g_i(x) = 0\}.$$

The inactive set at x is the complement of $\mathcal{A}(x)$:

$$(1.8) \qquad \bar{\mathcal{A}}(x) = \{1, 2, \ldots, m\} - \mathcal{A}(x).$$

With this condition, the complementarity condition from (1.3), $\sum_{i=1}^{m} \lambda_i g_i(x) = 0$, becomes $\lambda_{\bar{\mathcal{A}}(x)} = 0$.

If certain regularity conditions hold at a stationary point x (discussed below), there exist $\mu \in \mathsf{R}^m$ that satisfy the Karush-Kuhn-Tucker conditions (or KKT conditions) [3, 4, 10]:

$$(1.9) \qquad \nabla_x f(x) + \sum_{i=1}^{m} \mu_i \nabla_x g_i(x) = 0, \ \ \mu \geq 0, \ \ g(x) \leq 0, \ \ \mu^T g(x) = 0.$$

In this case, μ are referred to as the Lagrange multipliers.

The regularity condition, or constraint qualification, ensures that a linear approximation of the feasible set in the neighborhood of a stationary point x captures the geometry of the feasible set. The regularity condition that we will use at times at a stationary point x is the Mangasarian-Fromovitz constraint qualification (MFCQ) [22, 21]:

$$(1.10) \qquad \nabla_x g_i(x)^T p < 0, \ \text{ for some } p \in \mathsf{R}^n \text{ and } i \in \mathcal{A}(x).$$

It is well known [12] that MFCQ is equivalent to the boundedness of the set $\mathcal{M}(x)$ of Lagrange multipliers that satisfy (1.9), that is,

$$(1.11) \qquad \mathcal{M}(x) = \{\mu \geq 0 \ \mid \ (x, \mu) \ \text{ satisfy (1.9)}\}.$$

Note that $\mathcal{M}(x)$ is certainly polyhedral in any case. It is immediate from (1.3) and (1.9) that

$$\mathcal{M}(x) \neq \emptyset \Leftrightarrow \Omega_1(x) \neq \emptyset$$

and that

$$(1.12) \qquad \mu \in \mathcal{M}(x) \Leftrightarrow \lambda = \frac{(1, \mu)}{||(1, \mu)||_1} \in \Omega_1(x) \subset \Omega_0(x).$$

The critical cone at a stationary point x is [8, 28]

$$(1.13) \qquad \mathcal{C}(x) = \left\{u \in \mathsf{R}^n \ \mid \nabla_x g_i(x)^T u \leq 0, \ i \in \mathcal{A}(x); \ \nabla_x f(x)^T u \leq 0\right\}.$$

The second-order necessary conditions for x^* to be a local minimum are that [16]

$$(1.14) \qquad \forall u \in \mathcal{C}(x^*), \ \exists \lambda^* \in \Omega_0(x^*), \text{ such that } u^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) u \geq 0.$$

The second-order sufficient conditions for x^* to be a local minimum are that $\Omega_0(x^*) \neq \emptyset$ and [16]

$$\forall u \in \mathcal{C}(x^*),\ \exists \lambda^* \in \Omega_0(x^*), \text{ such that } u^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) u > 0. \tag{1.15}$$

Further analysis shows that, in the presence of MFCQ (1.10), these conditions are necessary and sufficient for the quadratic growth condition to hold [6, 16, 17, 28].

We denote the L_∞ nondifferentiable penalty function by

$$P(x) = \max\left\{0, g_1(x), ...g_m(x)\right\}. \tag{1.16}$$

The nonlinear program (1.1) satisfies the quadratic growth condition with a parameter σ if

$$\max\left\{f(x) - f(x^*), g_1(x), g_2(x) \dots g_m(x)\right\} \geq \sigma \left|\left|x - x^*\right|\right|^2 \tag{1.17}$$

for some $\sigma > 0$ and all x in a neighborhood of x^*.

The quadratic growth condition can be rewritten in terms of $P(x)$ as

$$\min\left\{f(x) - f(x^*), P(x)\right\} \geq \sigma \left|\left|x - x^*\right|\right|^2 \tag{1.18}$$

for some $\sigma > 0$ and all x in a neighborhood of x^*.

Recent results have shown that, if MFCQ (1.10) and the quadratic growth condition (1.17) hold at x^*, then x^* is an isolated stationary point of (1.1) [1]. Moreover, an algorithm with a line-search procedure based on the direction that is the solution of the subproblem

$$\begin{array}{lll} \min_{d \in \mathsf{R}^n} & \nabla_x f(x)^T d + d^T d, & \\ \text{subject to} & g_i(x) + \nabla_x g_i(x)^T d \leq 0, & i = 1, 2, \dots, m \end{array} \tag{1.19}$$

induces the Q-linear convergence of the merit function

$$\phi(x) = f(x) + c_\phi P(x) \tag{1.20}$$

to $\phi(x^*)$ and the R-linear convergence of the iterates. The quantity c_ϕ is a parameter with the property [1]

$$c_\phi > \max_{\mu \in \mathcal{M}(x^*)} \left|\left|\mu\right|\right|_1. \tag{1.21}$$

In addition, the penalty function satisfies an unconstrained quadratic growth condition on a neighborhood $\mathcal{V}(x^*)$ with some parameter $\tilde{\sigma} > 0$ [1]

$$\phi(x) = f(x) + c_\phi P(x) \geq \tilde{\sigma} \left|\left|x - x^*\right|\right|^2. \tag{1.22}$$

Superlinear convergence can also be obtained under the same conditions by using as progress direction a stationary point of the following quadratically constrained quadratic program [2]:

$$\begin{array}{ll} \min\limits_{d \in \mathsf{R}^n} & f(x) + \nabla_x f(x)^T d + \frac{1}{2} d^T \nabla^2_{xx} f(x) d \\ \text{subject to} & g_i(x) + \nabla_x g_i(x)^T d + \frac{1}{2} d^T \nabla^2_{xx} g_i(x) d \leq 0, \quad i = 1, 2, \dots, m \\ & d^T d \leq \gamma^2. \end{array}$$

1.2. Assumptions. As we specified in the introduction, we do not assume that MFCQ (1.10) holds at x^*. Instead we assume only that

1. The Lagrange multiplier set of (1.1), $\mathcal{M}(x^*)$ is not empty, or, equivalently, $\Omega_1(x^*) \neq \emptyset$.
2. The quadratic growth condition (1.17) holds near x^*. From [16], this condition is equivalent to the sufficient second-order condition (1.15).
3. f, g are twice continuously differentiable.

The objective of this paper is to transform (1.1) into a nonlinear program (1.2) that satisfies the same conditions at x^*, and MFCQ (1.10) in addition to those. As a result, the algorithms from [1, 2] can be used on the modified nonlinear program.

1.3. Notation. To distinguish between quantities associated with the original NLP (1.1) and the modified NLP (1.2), we use separate notations.

- The point at which we conduct the analysis is x^* for (1.1) and $(x^*, 0)$ for (1.2).
- The set of generalized multipliers (1.5) is $\Omega_0(x^*) \subset \mathsf{R}^{m+1}$ for (1.1) and $\Omega_0^c((x^*, 0)) \in \mathsf{R}^{m+2}$ for (1.2).
- The set of generalized multipliers with a positive first component (1.6) is $\Omega_1(x^*) \subset \mathsf{R}^{m+1}$ for (1.1) and $\Omega_1^c((x^*, 0)) \subset \mathsf{R}^{m+2}$ for (1.2).
- The set of Lagrange multipliers (1.1) is $\mathcal{M}(x^*) \subset \mathsf{R}^m$ for (1.1) and $\mathcal{M}^c((x^*, 0)) \subset \mathsf{R}^{m+1}$ for (1.2).
- The critical cone (1.13) is $\mathcal{C}(x^*) \subset \mathsf{R}^n$ for (1.1) and $\mathcal{C}^c((x^*, 0)) \subset \mathsf{R}^{n+1}$ for (1.2).
- The active set (1.7) $\mathcal{A}(x^*)$ for (1.1), and $\mathcal{A}^c((x^*, 0))$ for (1.2). It is immediate that $\mathcal{A}^c((x^*, 0)) = \mathcal{A}(x^*) \cup \{m+1\}$.

In general, we use the superscript ^{c} to denote a quantity connected to (1.2).

We also define the reduced set of Lagrange multipliers of (1.2), $\mathcal{M}_r^c(x^*)$, to be the projection of the Lagrange multiplier set of (1.2), $\mathcal{M}^c((x^*, 0))$, on its first m components:

$$(1.23)\quad \mathcal{M}_r^c(x^*) = \{\mu \in \mathsf{R}^m \mid \exists \mu_{m+1} \in \mathsf{R} \text{ such that } (\mu, \mu_{m+1}) \in \mathcal{M}^c((x^*, 0))\}.$$

2. Multiplier Sets of the Penalized Problem. We show that the penalty term in (1.2) has the effect of filtering the Lagrange multipliers of (1.1): The Lagrange multipliers of (1.2) are essentially the Lagrange multipliers of (1.1) whose 1 norm is less than or equal c.

We characterize the properties of the nonlinear program (1.2) at $x = x^*$, and $\zeta = 0$ or $(x^*, 0)$. In the next lemma we show that, for a sufficiently large c, (1.1) and (1.2) have essentially the same critical cone and closely related multiplier sets.

Lemma 2.1. *Let $\mu^* \in \mathcal{M}(x^*)$ be a Lagrange multiplier of (1.1). Then for c such that $c > ||\mu^*||_1$, we have that*

(i) $\mathcal{C}^c((x^*, 0)) = \{(u, 0) \mid u \in \mathcal{C}(x^*)\}$.

(ii) $\lambda^* \in \Omega_0(x^*), \lambda_0^* \geq \frac{1}{1+c} \Leftrightarrow \exists \lambda^c = (\lambda_0^c, \lambda_1^c, \ldots, \lambda_{m+1}^c) \in \Omega_0^c((x^*, 0))$, *such that*

$$\lambda^* = \frac{(\lambda_0^c, \lambda_1^c, \ldots \lambda_m^c)}{||(\lambda_0^c, \lambda_1^c, \ldots \lambda_m^c)||_1}.$$

Proof Let (u, y) be in the critical cone for (1.2) at $(x^*, 0)$, $(u, y) \in \mathcal{C}^c((x^*, 0))$, $u \in \mathsf{R}^n$, $y \in \mathsf{R}$. This means that (u, y) satisfies the critical cone conditions (1.13)

$$(2.1)\qquad \nabla_x f(x^*)^T u + cy \leq 0$$

$$(2.2) \qquad \nabla_x g_i(x^*)^T u - y \leq 0, \quad i \in \mathcal{A}(x^*)$$
$$(2.3) \qquad -y \leq 0.$$

We now take the Lagrange multiplier $\mu^* \in \mathcal{M}(x^*)$ defined in the hypothesis of the Lemma. We add the inequality (2.1) with the inequalities (2.2), each multiplied with the corresponding $\mu_i^* \geq 0$. Since from (1.9) $\mu_i^* = 0$, $\forall i \notin \mathcal{A}(x^*)$, we obtain the following inequality:

$$(2.4) \qquad \left(\nabla_x f(x^*) + \sum_{i=1}^m \mu_i^* \nabla_x g_i(x^*) \right)^T u + \left(c - \sum_{i=1}^m \mu_i^* \right) y \leq 0.$$

Since $\mu^* \in \mathcal{M}(x^*)$ satisfies the KKT conditions (1.9) for (1.1), we must have in particular that

$$\nabla_x f(x^*) + \sum_{i=1}^m \mu_i^* \nabla_x g_i(x^*) = 0.$$

Using this relation in (2.4), we obtain that

$$\left(c - \sum_{i=1}^m \mu_i^* \right) y \leq 0.$$

Since, from our assumptions, $c > \sum_{i=1}^m \mu_i^*$, this results in $y \leq 0$, which, together with (2.3), implies that $y = 0$. It now follows by inspection of (2.1) and (2.2) that, since $y = 0$, u is in the critical cone $\mathcal{C}(x^*)$ (1.13) of (1.1) at x^*. Therefore

$$(2.5) \qquad (u, y) \in \mathcal{C}^c((x^*, 0)) \Rightarrow u \in \mathcal{C}(x^*), \quad y = 0.$$

It is immediate that for any $u \in \mathcal{C}(x^*)$, we must have $(u, 0) \in \mathcal{C}^c((x^*, 0))$ (by examination of the critical cone conditions (2.1), (2.2), and (2.3)), which together with (2.5) proves part i.

Now let $\lambda^c = (\nu^c, \lambda_{m+1}^c) \in \Omega_0^c((x^*, 0))$, where

$$\nu^c = (\nu_0^c, \nu_1^c, \ldots \nu_m^c).$$

Therefore, $\nu^c \geq 0$ and $\lambda_{m+1}^c \geq 0$ satisfy the Fritz-John conditions (1.3) for (1.2):

$$(2.6) \qquad \nu_0^c \nabla_x f(x^*) + \sum_{i=1}^m \nu_i^c \nabla_x g_i(x^*) = 0,$$

$$(2.7) \qquad \nu_{\bar{\mathcal{A}}(x^*)}^c = 0, \quad \sum_{i=0}^m \nu_i^c + \lambda_{m+1}^c = 1,$$

$$(2.8) \qquad c\nu_0^c - \sum_{i=1}^m \nu_i^c - \lambda_{m+1}^c = 0.$$

From (2.8), (2.7), we have that $(c+1)\nu_0^c = \sum_{i=0}^m \nu_i^c + \lambda_{m+1}^c = 1$. Therefore

$$(2.9) \qquad \nu_0^c = \frac{1}{1+c},$$

and ν^c satisfies

$$0 < \frac{1}{1+c} = \nu_0^c \leq ||\nu^c||_1 = 1 - \lambda_{m+1}^c \leq 1. \tag{2.10}$$

Since $||\nu^c||_1 \neq 0$, we can thus define

$$\lambda^* = \frac{\nu^c}{||\nu^c||_1}, \tag{2.11}$$

which satisfies $\lambda^* \geq 0$, $||\lambda^*||_1 = 1$. Also, by dividing (2.6) and (2.7) by $||\nu^c||_1$, we obtain $\lambda^*_{\bar{\mathcal{A}}} = 0$ and

$$\lambda_0^* \nabla_x f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_x g_i(x^*) = 0,$$

which shows that $\lambda^* = (\lambda_0^*, \lambda_1^*, \ldots \lambda_m^*)$ satisfies the Fritz-John conditions (1.3) for (1.1). Therefore $\lambda^* \in \Omega_0(x^*)$. In addition, from (2.9), (2.10) and (2.11) we have that

$$\lambda_0^* = \frac{\nu_0^c}{||\nu^c||_1} \geq \frac{1}{1+c}.$$

We have thus shown that

$$(\nu^c, \lambda_{m+1}^c) \in \Omega_0^c((x^*, 0)) \Rightarrow \lambda^* = \frac{\nu^c}{||\nu^c||_1} \in \mathcal{C}(x^*), \quad \lambda_0^* \geq \frac{1}{1+c}. \tag{2.12}$$

Assume now that $\lambda^* = (\lambda_0^*, \lambda_1^*, \ldots, \lambda_m^*) \in \Omega_0(x^*)$, with $\lambda_0^* \geq \frac{1}{1+c}$. Therefore λ^* satisfies (1.3) at x^*:

$$\lambda_0^* \nabla_x f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_x g(x^*) = 0, \quad \lambda^* \geq 0, \quad \lambda^*_{\bar{\mathcal{A}}(x^*)} = 0, \quad ||\lambda^*||_1 = 1. \tag{2.13}$$

From $||\lambda^*||_1 = 1$ it follows that

$$\sum_{i=1}^{m} \lambda_i^* = 1 - \lambda_0^* \leq 1 - \frac{1}{1+c} = \frac{c}{1+c} \leq c\lambda_0^*.$$

Therefore we can define

$$\lambda_{m+1}^* = c\lambda_0^* - \sum_{i=1}^{m} \lambda_i^* \geq 0,$$

which ensures that

$$c\lambda_0^* - \sum_{i=1}^{m+1} \lambda_i^* = 0. \tag{2.14}$$

Define now

$$\lambda^c = \frac{(\lambda^*, \lambda_{m+1}^*)}{||(\lambda^*, \lambda_{m+1}^*)||_1}. \tag{2.15}$$

We denote the components of λ^c by $\lambda_0^c, \lambda_1^c, \ldots, \lambda_{m+1}^c$. Since from (2.13) we have that $||\lambda^*||_1 = 1$, it follows from (2.15) that

$$\lambda^* = \frac{(\lambda_0^c, \lambda_1^c, \ldots \lambda_m^c)}{||(\lambda_0^c, \lambda_1^c, \ldots \lambda_m^c)||_1}.$$

We divide the relations (2.13) and (2.14), which are linear in λ^*, by $\left|\left|(\lambda^*, \lambda_{m+1}^*)\right|\right|_1$. From (2.15) we obtain that λ^c satisfies (2.6), (2.7), and (2.8), with $\nu_i^c = \lambda_i^c$, for $i = 0, 1, \ldots, m$, which are precisely the Fritz-John conditions for (1.2). Therefore, $\lambda^c \in \Omega_0^c((x^*, 0))$. We have thus proved that

$$\lambda^* \in \Omega_0(x^*), \lambda_0^* \geq \frac{1}{1+c} \Rightarrow \exists \lambda^c \in \Omega_0^c((x^*, 0)), \text{ such that } \lambda^* = \frac{(\lambda_0^c, \lambda_1^c, \ldots \lambda_m^c)}{||(\lambda_0^c, \lambda_1^c, \ldots \lambda_m^c)||_1}. \tag{2.16}$$

From (2.12) and (2.16), the conclusion of part ii follows. The proof is complete. $\diamond$

We now show that the penalty term results in the reduced Lagrange multiplier set $\mathcal{M}_r^c(x^*)$ (1.23) of (1.2) being a bounded subset of the set of Lagrange multipliers $\mathcal{M}(x^*)$ of (1.1).

LEMMA 2.2. *The set of Lagrange multipliers of (1.2) satisfies* $\mathcal{M}^c((x^*, 0)) = \overline{\mathcal{M}}^c(x^*)$, *where*

$$\overline{\mathcal{M}}^c(x^*) = \{\mu^c \in \mathsf{R}^{m+1} \mid \mu^* = (\mu_1^c, \mu_2^c, \ldots \mu_m^c) \in \mathcal{M}(x^*), \\ ||\mu^*||_1 \leq c, \ \mu_{m+1}^c = c - ||\mu^*||_1\}.$$

In particular $\mu^c \in \mathcal{M}^c((x^*, 0)) \Rightarrow ||\mu^c||_1 = c$. *The reduced set of Lagrange multipliers (1.23) thus satisfies*

$$\mathcal{M}_r^c(x^*) = \{\mu^* \in \mathsf{R}^m \mid \mu^* \in \mathcal{M}(x^*), \ \ ||\mu^*||_1 \leq c\}.$$

Note In the case where there is no $\mu^* \in \mathcal{M}(x^*)$ such that $||\mu^*||_1 \leq c$, we have that $\overline{\mathcal{M}}^c(x^*) = \emptyset$, and $\mathcal{M}_r^c(x^*) = \emptyset$.

Proof We will prove the results first assuming that both $\mathcal{M}^c((x^*, 0))$ and $\overline{\mathcal{M}}^c(x^*)$ are not empty. Let $\mu^c \in \mathcal{M}^c((x^*, 0))$. From the KKT conditions (1.9) for (1.2) at $(x^*, 0)$, $\mu^c = (\mu_1^c, \mu_2^c, \ldots \mu_{m+1}^c)$ satisfies

$$\begin{aligned} \nabla_x f(x^*) + \textstyle\sum_{i=1}^m \mu_i^c \nabla_x g_i(x^*) = 0, \ \ \sum_{i=1}^{m+1} \mu_i^c &= c, \\ \mu^c \geq 0, \ \ g(x^*) \leq 0, \ \ \textstyle\sum_{i=1}^m \mu_i^c g_i(x^*) &= 0, \end{aligned} \tag{2.17}$$

Then, in particular, $\mu^* = (\mu_1^c, \mu_2^c, \ldots \mu_m^c)$ satisfies $||\mu^*||_1 \leq c$, $\mu_{m+1}^c = c - ||\mu^*||_1$:

$$\begin{aligned} \nabla_x f(x^*) + \textstyle\sum_{i=1}^m \mu_i^c \nabla_x g_i(x^*) = 0, \ \ \mu^* &\geq 0, \\ g(x^*) \leq 0, \ \ \textstyle\sum_{i=1}^m \mu_i^c g_i(x^*) &= 0, \end{aligned} \tag{2.18}$$

which represents the KKT conditions (1.9) for (1.1). Therefore, $\mu^* \in \mathcal{M}(x^*)$, and this proves that $\mathcal{M}^c((x^*, 0)) \subset \overline{\mathcal{M}}^c(x^*)$.

For the reverse inclusion, if $\mu^* \in \overline{\mathcal{M}}^c(x^*)$, then $\mu^* \in \mathcal{M}(x^*)$, and $||\mu^*||_1 \leq c$. We can thus define $\mu_{m+1}^c = c - ||\mu^*||_1 \geq 0$ and $\mu_i^c = \mu_i^*, i = 1, 2, \ldots, m$. It is then immediate by inspection that $\mu^c = (\mu_1^c, \mu_2^c, \dots \mu_{m+1}^c)$ satisfies (2.17). Therefore, this proves the reverse inclusion, $\bar{\mathcal{M}}(x^*) \subset \mathcal{M}^c((x^*, 0))$.

From our proof, if either of $\mathcal{M}^c((x^*, 0))$ and $\bar{\mathcal{M}}^c(x^*)$ is not empty, the other also is not empty. Therefore, if one is empty, the other is empty. Thus, we have proved that $\mathcal{M}^c((x^*, 0)) = \bar{\mathcal{M}}^c(x^*)$ even when one of the sets is empty.

From (2.17) we have in particular that if $\mu^c \in \mathcal{M}^c((x^*, 0))$, then $||\mu^c||_1 = c$. Thus, the second part of the claim is proved.

The statement concerning $\mathcal{M}_r^c(x^*)$ follows by inspection of the definition of $\bar{\mathcal{M}}(x^*)$. The proof is complete. $\diamond$

The main consequence of the preceding lemma is that the projection of the Lagrange multiplier set $\mathcal{M}^c((x^*, 0))$ of (1.2) on the first m components is the Lagrange multipliers μ^* of (1.1) satisfying $||\mu^*||_1 \leq c$. Therefore, adding a penalty term in (1.1) results in retaining only those multipliers of (1.1) that are less than c in $||\cdot||_1$.

Example Consider the following nonlinear programming problem [27]:

$$\begin{array}{lr} \min_x & x^2 \\ \text{subject to} & x^6 \sin \frac{1}{x} = 0. \end{array} \tag{2.19}$$

The data of the problem are twice continuously differentiable. To put the problem in the framework we used so far, we replace the equality constraint by two inequality constraints

$$\begin{array}{lcccl} \min_x & f(x) & = & x^2 & \\ \text{subject to} & g_1(x) & = & x^6 \sin \frac{1}{x} & \leq 0. \\ & g_2(x) & = & -x^6 \sin \frac{1}{x} & \leq 0. \end{array} \tag{2.20}$$

The global solution of the problem is $x^* = 0$. At x^* we have that $\nabla_x f(x^*) = 0$, $\nabla_x g_1(x^*) = 0$, and $\nabla_x g_2(x^*) = 0$ and that both constraints are active.

Therefore, the Lagrange multipliers $\mu^* = (\mu_1^*, \mu_2^*)$ are those μ^* that satisfy the KKT conditions (1.9), or

$$\mu^* \geq 0, \quad 0 = 0 + \mu_1^* \times 0 + \mu_2^* \times 0,$$

and the Lagrange multiplier set is thus

$$\mathcal{M}(0) = \left\{\mu^* \in \mathsf{R}^2 \mid \mu_1^* \geq 0,\ \mu_2^* \geq 0\right\}. \tag{2.21}$$

Since the nonlinear program (2.20) does not satisfy (1.10), its Lagrange multiplier set $\mathcal{M}(0)$ is unbounded. We now construct the corresponding penalized nonlinear program (1.2) for this case, for $c = 1$. We obtain

$$\begin{array}{lcccl} \min_{x,\zeta} & f^c(x,\zeta) & = & x^2 + \zeta & \\ \text{subject to} & g_1^c(x,\zeta) & = & x^6 \sin \frac{1}{x} - \zeta & \leq 0 \\ & g_2^c(x,\zeta) & = & -x^6 \sin \frac{1}{x} - \zeta & \leq 0 \\ & g_3^c(x,\zeta) & = & -\zeta & \leq 0. \end{array} \tag{2.22}$$

At $(0,0)$, the gradients are $\nabla_{(x,\zeta)} f^c(0,0) = (0,1)^T$, $\nabla_{(x,\zeta)} g_1^c(0,0) = (0,-1)^T$, $\nabla_{(x,\zeta)} g_2^c(0,0) = (0,-1)^T$, and $\nabla_{(x,\zeta)} g_3^c(0,0) = (0,-1)^T$. The Lagrange multipliers $\mu^c = (\mu_1^c, \mu_2^c, \mu_3^c)$ of (2.20) at $(0,0)$ satisfy (1.9), or $\mu^c \geq 0$ and

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \mu_1^c \begin{pmatrix} 0 \\ -1 \end{pmatrix} + \mu_2^c \begin{pmatrix} 0 \\ -1 \end{pmatrix} + \mu_3^c \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

We thus have that the set of Lagrange multipliers of (2.22) is

$$\mathcal{M}^c((0,0)) = \left\{\mu^c \in \mathsf{R}^3 \mid \mu^c \geq 0,\ \ \mu_1^c + \mu_2^c + \mu_3^c = 1\right\},$$

and the projection of the Lagrange multiplier set on its first two coordinates, or the reduced Lagrange multiplier set, (1.23), thus becomes

$$\mathcal{M}_r^c(0) = \left\{\mu \in \mathsf{R}^2 \mid \mu \geq 0,\ \ \mu_1 + \mu_2 \leq 1\right\}.$$

It is immediate that

$$\mathcal{M}_r^c(0) = \{\mu \in \mathcal{M}(x^*) \mid ||\mu||_1 \leq 1\},$$

which is the claim of the Lemma 2.2: The projection of the Lagrange multiplier set $\mathcal{M}^c((0,0))$ of the penalized problem (2.22) on its first two coordinates (the original Lagrange multiplier variables) consists of the elements of the Lagrange multiplier set $\mathcal{M}(0)$ of the original problem (2.20) whose 1 norm is less than the penalty parameter, $c = 1$. In this sense, the penalty term $c\zeta$ of (1.1) acts like a filter: it retains only the Lagrange multipliers of the original problem whose 1 norm is less than c.

3. The Quadratic Growth Condition for the Penalized Problem. We now discuss the connection between the parameter σ involved in the definition of (1.17) and the corresponding parameter for the second-order conditions (1.14) and (1.15).

LEMMA 3.1. *A necessary condition for the quadratic growth condition to hold with parameter σ in a neighborhood of a stationary point x^* of (1.1) is*

$$\forall u \in \mathcal{C}(x^*),\ \exists \lambda^* \in \Omega_0(x^*) \text{ such that } u^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) u \geq 2\sigma \left|\left|u\right|\right|^2. \tag{3.1}$$

A sufficient condition for the quadratic growth condition to hold with parameter σ in a neighborhood of a stationary point x^* is

$$\forall u \in \mathcal{C}(x^*),\ \exists \lambda^* \in \Omega_0(x^*),\ \text{such that } u^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) u > 2\sigma \left|\left|u\right|\right|^2. \tag{3.2}$$

Proof If the nonlinear program (1.1) satisfies the quadratic growth condition with a parameter σ (1.17), it follows that the modified nonlinear program

$$\begin{array}{c} \min_x\ f(x) - \sigma \left|\left|x - x^*\right|\right|^2 \\ \text{subject to } \ g_i(x) - \sigma \left|\left|x - x^*\right|\right|^2 \leq 0,\ \forall i = 1, 2, \ldots, m \end{array} \tag{3.3}$$

satisfies

$$\begin{array}{c} \max\Big\{f(x) - f(x^*) - \sigma \left|\left|x - x^*\right|\right|^2, g_1(x) - \sigma \left|\left|x - x^*\right|\right|^2, \\ g_2(x) - \sigma \left|\left|x - x^*\right|\right|^2 \ldots g_m(x) - \sigma \left|\left|x - x^*\right|\right|^2\Big\} \geq 0 \end{array} \tag{3.4}$$

in a neighborhood of x^*. Thus, in particular, x^* is a local minimum for (3.3). Since $\nabla_x \left|\left|x - x^*\right|\right|^2 = 0$ at x^*, it follows that (1.1) and (3.3) have the same multiplier set $\Omega_0(x^*)$ and critical cone $\mathcal{C}^c(x^*)$. If $\mathcal{L}^\sigma$ is the Lagrangian of (3.3), it immediately follows that, since $\left|\left|\lambda^*\right|\right|_1 = 1$ for $\lambda^* \in \Omega_0(x^*)$, $\mathcal{L}^\sigma(x, \lambda^*) = \mathcal{L}(x, \lambda^*) - \sigma \left|\left|x - x^*\right|\right|^2$. As a result, we have from the second-order necessary condition (1.14) applied to the local minimum x^* that

$$\begin{array}{c} \forall u \in \mathcal{C}(x^*),\ \exists \lambda^* \in \Omega_0(x^*) \text{ such that} \\ u^T \nabla_{xx}^2 \mathcal{L}(x, \lambda^*) u - 2\sigma u^T u = u^T \nabla_{xx}^2 \mathcal{L}^\sigma(x^*, \lambda^*) u \geq 0, \end{array} \tag{3.5}$$

which proves (3.1), the necessary condition part of the lemma. Assume now that x^* is a stationary point of (1.1) satisfying (3.2). It follows that (3.3) satisfies

$$\begin{aligned} &\forall u \in \mathcal{C}(x^*),\ \exists \lambda^* \in \Omega_0(x^*), \text{ such that} \\ &u^T \nabla^2_{xx} \mathcal{L}^\sigma(x^*, \lambda^*) u = u^T \nabla^2_{xx} \mathcal{L}(x, \lambda^*) u - 2\sigma u^T u > 0. \end{aligned} \tag{3.6}$$

This means that (3.3) satisfies the second-order conditions (1.15) at x^*, and x^* is, as a result, a strict local minimum of (3.3). Therefore, there exists a neighborhood of x^* such that

$$\begin{aligned} \max \Big\{ & f(x) - f(x^*) - \sigma \left\|x - x^*\right\|^2, g_1(x) - \sigma \left\|x - x^*\right\|^2, \\ & g_2(x) - \sigma \left\|x - x^*\right\|^2 \dots g_m(x) - \sigma \left\|x - x^*\right\|^2 \Big\} > 0. \end{aligned} \tag{3.7}$$

It is then immediate that (1.1) satisfies the quadratic growth condition (1.17) with parameter σ, which completes the proof. $\diamond$

We now show that under our assumptions, the second order sufficient conditions (1.15) hold with multipliers $\lambda^* \in \Omega_1(x^*)$. From (1.12) this will ultimately imply that the modified nonlinear program (1.2) satisfies the second-order conditions and has a bounded Lagrange multiplier set.

LEMMA 3.2. *Let $\mu \in \mathcal{M}(x^*)$ and*

$$\tilde{\lambda} = \frac{(1, \mu)}{1 + \left\|\mu\right\|_1}, \quad \theta = \frac{\sigma + \left\|\nabla^2_{xx} \mathcal{L}(x^*, \tilde{\lambda})\right\|}{2\sigma + \left\|\nabla^2_{xx} \mathcal{L}(x^*, \tilde{\lambda})\right\|}.$$

Then $\forall u \in \mathcal{C}(x^*)$, $\exists \lambda^* = (\lambda^*_0, \lambda^*_1, \dots \lambda^*_m) \in \Omega_0(x^*)$ such that

$$\lambda^*_0 \geq \lambda_\zeta = (1 - \theta) \frac{1}{1 + \left\|\mu\right\|_1} > 0, \quad u^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) u \geq \sigma \left\|u\right\|^2.$$

Proof Let $u \in \mathcal{C}(x^*)$. Since (1.1) satisfies the quadratic growth condition (1.17), it follows from Lemma 3.1 that, $\forall u \in \mathcal{C}(x^*)$, there exists $\lambda^+ \in \Omega_0(x^*)$ such that

$$u^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^+) u \geq 2\sigma \left\|u\right\|^2. \tag{3.8}$$

Let now

$$\lambda^* = (1 - \theta)\tilde{\lambda} + \theta \lambda^+. \tag{3.9}$$

From the linearity of the Lagrangian (1.4) with respect to the multipliers λ, it follows that

$$\mathcal{L}(x^*, \lambda^*) = (1 - \theta)\mathcal{L}(x^*, \tilde{\lambda}) + \theta \mathcal{L}(x^*, \lambda^+).$$

Therefore from the definition of θ in our hypothesis and (3.8), we will have that

$$\begin{aligned} u^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) u = (1 - \theta) u^T \nabla^2_{xx} \mathcal{L}(x^*, \tilde{\lambda}) u + \theta u^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^+) u \geq \\ -(1 - \theta) \left\|\nabla^2_{xx} \mathcal{L}(x^*, \tilde{\lambda})\right\| \left\|u\right\|^2 + \theta 2\sigma \left\|u\right\|^2 = \\ \theta \left(\left\|\nabla^2_{xx} \mathcal{L}(x^*, \tilde{\lambda})\right\| + 2\sigma \right) \left\|u\right\|^2 - \left\|\nabla^2_{xx} \mathcal{L}(x^*, \tilde{\lambda})\right\| \left\|u\right\|^2 = \\ \left(\left\|\nabla^2_{xx} \mathcal{L}(x^*, \tilde{\lambda})\right\| + \sigma \right) \left\|u\right\|^2 - \left\|\nabla^2_{xx} \mathcal{L}(x^*, \tilde{\lambda})\right\| \left\|u\right\|^2 = \sigma \left\|u\right\|^2. \end{aligned} \tag{3.10}$$

Since $\Omega_0(x^*)$ (1.5) is a convex set, and $\tilde{\lambda} \in \Omega_0(x^*)$ from (1.12), $\lambda^+ \in \Omega_0(x^*)$, from (3.9) and since $0 < \theta < 1$, it follows that $\lambda^* \in \Omega_0(x^*)$. From the definition of $\tilde{\lambda}$ in our hypothesis, we have that

$$\tilde{\lambda}_0 = \frac{1}{1 + ||\mu||_1}.$$

From (3.9) it follows that

$$\lambda_0^* = (1-\theta)\tilde{\lambda}_0 + \theta\lambda_0^+ \geq (1-\theta)\frac{1}{1+||\mu||_1} = \lambda_\zeta.$$

The conclusion follows from the preceding equation, from (3.10). and from the fact that $0 < \theta < 1$. $\diamond$

We now show that the modified nonlinear program (1.2) satisfies (1.14), with a parameter possibly different from σ. The Lagrangian function for (1.2) is, following (1.4),

$$\mathcal{L}^c(x, \zeta, \lambda^c) = \lambda_0^c f(x) + \sum_{i=1}^{m} \lambda_i^c g_i^c(x) + \lambda_{m+1}^c \zeta. \tag{3.11}$$

We keep denoting by $\mathcal{L}(x, \lambda)$ the Lagrangian function of (1.1). We write

$$\bar{\lambda} = (\lambda_0^c, \lambda_1^c, \ldots \lambda_m^c). \tag{3.12}$$

Note that $\bar{\lambda}$ is not necessarily an element of $\Omega_0(x^*)$, since $||\bar{\lambda}||_1$ may not be equal to 1.

Then the Hessian of the Lagrangian $\mathcal{L}^c(x, \lambda^c)$ (3.11) becomes

$$\nabla^2_{(x,\zeta)(x,\zeta)} \mathcal{L}^c(x, \zeta, \lambda^c) = \begin{bmatrix} \nabla^2_{xx}\mathcal{L}(x, \bar{\lambda}) & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.13}$$

Lemma 3.3. *Let $\mu \in \mathcal{M}(x^*)$ and*

$$c_\zeta = \max\left\{ ||\mu||_1, \frac{1}{\lambda_\zeta} - 1 \right\},$$

where λ_ζ is the parameter defined in Lemma 3.2. Then for any c satisfying $c > c_\zeta$ we have the following property: For all $u^c \in \mathcal{C}^c((x^*, 0))$, $\exists \lambda^c \in \Omega^c((x^*, 0))$ such that

$$(u^c)^T \nabla^2_{(x,\zeta)(x,\zeta)} \mathcal{L}^c(x^*, 0, \lambda^c) u^c \geq \frac{\sigma}{(1+c)} ||u^c||^2.$$

Proof Let $u^c \in \mathcal{C}^c((x^*, 0))$. Then from Lemma 2.1, since $c > c_\zeta \geq ||\mu||_1$, it follows that $u^c = (u, 0)$, where $u \in \mathcal{C}(x^*)$. From Lemma 3.2 it follows that there exists $\lambda^* \in \Omega_0(x^*)$ such that $\lambda_0^* \geq \lambda_\zeta$ and

$$u^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) u \geq \sigma ||u||^2. \tag{3.14}$$

Since $c > c_\zeta \geq \frac{1}{\lambda_\zeta} - 1$ and $\lambda_0^* \geq \lambda_\zeta$, we have that

$$(1+c)\lambda_0^* \geq 1 \;\;\Rightarrow\;\; \lambda_0^* \geq \frac{1}{1+c}.$$

Therefore from Lemma 2.1 (ii) it follows that there exists $\lambda^c = (\bar{\lambda}^c, \lambda^c_{m+1}) \in \Omega^c_0((x^*, 0))$ such that $\lambda^* = \frac{\bar{\lambda}^c}{||\bar{\lambda}^c||_1}$, where

$$\bar{\lambda}^c = (\lambda^c_0, \lambda^c_1, \cdots, \lambda^c_m). \tag{3.15}$$

From (3.13) and the linearity of the Lagrangians (1.4) and (3.11), since $u^c = (u, 0)$ and from (3.14) it follows that

$$(u^c)^T \frac{1}{||\bar{\lambda}^c||_1} \nabla^2_{(x,\zeta)(x,\zeta)} \mathcal{L}^c(x^*, 0, \lambda^c) u^c = (u^c)^T \nabla^2_{(x,\zeta)(x,\zeta)} \mathcal{L}^c(x^*, 0, \frac{\lambda^c}{||\bar{\lambda}^c||_1}) u^c =$$
$$u^T \nabla^2_{xx} \mathcal{L}(x, \lambda^*) u \geq \sigma ||u||^2 = \sigma ||u^c||^2$$

or

$$(u^c)^T \nabla^2_{(x,\zeta)(x,\zeta)} \mathcal{L}^c(x^*, 0, \lambda^c) u^c \geq \sigma ||\bar{\lambda}^c||_1 ||u^c||^2 . \tag{3.16}$$

Since $\lambda^c \in \Omega_0((x^*, 0))$, λ^c must satisfy the Fritz-John condition (1.3) for (1.2), which leads to (2.8), or

$$c\lambda^c_0 = \sum_{i=1}^{m+1} \lambda^c_i,$$

and $||\lambda^c||_1 = 1$. Therefore,

$$(c+1)\lambda^c_0 = \lambda_0 + \sum_{i=1}^{m+1} \lambda^c_i = ||\lambda^c||_1 = 1,$$

which results in

$$\lambda^c_0 = \frac{1}{1+c}.$$

As a result we have from (3.15) that $||\bar{\lambda}^c||_1 \geq \lambda^c_0 = \frac{1}{1+c}$. Using this inequality in (3.16), we obtain that, $\forall u^c \in \mathcal{C}^c((x^*, 0))$, there exists $\lambda^c \in \Omega^c_0((x^*, 0))$:

$$(u^c)^T \nabla^2_{(x,\zeta)(x,\zeta)} \mathcal{L}^c(x^*, 0, \lambda^c) u^c \geq \frac{\sigma}{(1+c)} ||u^c||^2 .$$

The proof is complete. ◇

We are now ready to state our main result of this work.

Theorem 3.4. *Let x^* be a minimum of the nonlinear program (1.1) at which the quadratic growth condition with parameter σ (1.17) holds, and for which the set of Lagrange multipliers $\mathcal{M}(x^*)$ is not empty. There exists $c_\zeta > 0$ such that, for any $c > c_\zeta$*

- *i The nonlinear program (1.2) satisfies the Mangasarian-Fromovitz constraint qualification at $(x^*, 0)$.*
- *ii The nonlinear program (1.2) satisfies the quadratic growth condition at $(x^*, 0)$ with any parameter $\bar{\sigma} < \frac{\sigma}{2(1+c)}$.*
- *iii There exists a neighborhood $\mathcal{V}^c(x^*)$ of x^* and $\tilde{\sigma} > 0$ such that*

$$\forall x \in \mathcal{V}^c(x^*), \quad f(x) + (c+1)P(x) - f(x^*) \geq \tilde{\sigma} ||x - x^*||^2 .$$

Proof Let c_ζ be the quantity introduced in Lemma 3.3. For any $c > c_\zeta$, from Lemma 2.2, it follows that the Lagrange multiplier set of (1.2), $\mathcal{M}^c((x^*,0))$ is not empty and bounded, which, from [12], is equivalent to (1.2) satisfying MFCQ (1.10). The same conclusion can also be drawn by noting that the direction $p^c = (p, p_\zeta)$ with $p \in \mathsf{R}^n$, $p = 0$, and $p_\zeta = 1$ satisfies $\nabla_x g_i(x^*)^T p - p_\zeta < 0$, $i = enumm$, $-p_\zeta < 0$ and thus satisfies (1.10) proving part i.

Now let $u^c \in \mathcal{C}^c((x^*,0))$. From Lemma 3.3, it follows that, since $c > c_\zeta$, there exists $\lambda^c \in \Omega^c(c^*,0)$ such that

$$(u^c)^T \nabla^2_{(x,\zeta)(x,\zeta)} \mathcal{L}^c(x^*,0,\lambda^c) u^c \geq \frac{\sigma}{(1+c)} \left|\left|u^c\right|\right|^2 .$$

From Lemma 3.1, it follows that the preceding relation is a sufficient condition for (1.2) to satisfy the quadratic growth condition (1.17) with any parameter $\bar{\sigma} > 0$, $\bar{\sigma} < \frac{\sigma}{2(1+c)}$. This proves part ii.

Since $c > c_\zeta$, it follows from Lemma 2.2 that if μ^c is a Lagrange multiplier of (1.1), $\mu^c \in \mathcal{M}^c((x^*,0))$, then $\left|\left|\mu^c\right|\right|_1 = c$. We therefore have that

$$c + 1 > \max_{\mu^c \in \mathcal{M}^c((x^*,0))} \left|\left|\mu^c\right|\right|_1 .$$

Since from parts (i) and (ii), (1.2) satisfies MFCQ (1.10), (1.17) and (1.21) from the preceding relation, it follows that from (1.22) applied to (1.2) that, for some $r > 0$, $\tilde{\sigma} > 0$ and, for any $x \in B((x^*,0),r)$,

$$\begin{aligned}\phi(x,\zeta) = f(x) + c\zeta + (c+1)\max\{0, g_1(x) - \zeta,& \\ g_2(x) - \zeta, \ldots, g_m(x) - \zeta, -\zeta\} - f(x^*) \quad &\geq \quad \tilde{\sigma}(\left|\left|x - x^*\right|\right|^2 + \zeta^2).\end{aligned} \tag{3.17}$$

The conclusion of part iii follows after taking $\zeta = 0$ in (3.17): For any $x \in B(x^*)$,

$$f(x) + (c+1)\max\{0, g_1(x), g_2(x), \ldots, g_m(x)\} - f(x^*) \geq \tilde{\sigma} \left|\left|x - x^*\right|\right|^2 .$$

The proof is complete. ◇

4. Algorithms for Problems with Unbounded Multipliers. Although this framework was developed for problems with inequality constraints, it can be easily extended to accommodate equality constraints. Indeed if we have the nonlinear program

$$\min_x f(x) \quad \text{subject to } g(x) \leq 0, \;\; h(x) = 0,$$

$h : \mathsf{R}^l \to \mathsf{R}$, it can be transformed into an inequality constrained nonlinear program

$$\min_x f(x) \quad \text{subject to } g(x) \leq 0, \;\; h(x) \leq 0, \;\; -h(x) \leq 0.$$

Even if the original problem satisfies at x^* the variant of MFCQ (1.10) that includes equality constraints [22],

$$\begin{aligned} \nabla_x h_j(x^*),\ j = 1,\ldots,l \ \text{ are linearly independent and} \\ \nabla_x h_j^T(x^*)p = 0,\ j = 1,\ldots,l,\ \ \nabla_x g_i(x^*)^T p < 0,\ \ i \in \mathcal{A}(x^*),\ \text{ for some } p \in \mathsf{R}^n. \end{aligned}$$

the transformed problem does not satisfy MFCQ (1.10). However, since in this paper we do not assume MFCQ (1.10), this does not create a difficulty. An important consequence of this fact is that we can accommodate even those cases for which the gradients of the equality constraints are linearly dependent.

To unify notation, we put (1.2) in the same form as (1.1). We denote $y = (x, \zeta)$, and we obtain that (1.2) can be written as

$$
(4.1)\qquad \begin{array}{rrcll}
\min_{x,\zeta} & f^c(y) & = & f(x) + c\zeta & \\
\text{subject to} & g_1^c(y) & = & g_1(x) - \zeta & \leq 0, \\
 & g_2^c(y) & = & g_2(x) - \zeta & \leq 0, \\
 & & \vdots & & \\
 & g_m^c(y) & = & g_m(x) - \zeta & \leq 0, \\
 & g_{m+1}^c(y) & = & -\zeta & \leq 0.
\end{array}
$$

As specified in the beginning of this work, we assume that (1.1) has a nonempty Lagrange multiplier set $\mathcal{M}(x^*)$, that it satisfies the quadratic growth condition (1.17), and that f, g are twice continuously differentiable. Under these assumptions, we can apply Theorem 3.4 to problem (1.1) to obtain that for $c > c_\zeta$ we have the following properties.

1. At $y^* = (x^*, 0)$, the modified problem (1.2) and its equivalent form (4.1) satisfy MFCQ (1.10). Therefore (1.2) and (4.1) have bounded multipliers.
2. At $y^* = (x^*, 0)$, (1.2) and (4.1) satisfy the quadratic growth condition (1.17).
3. The data of the problem are twice continuously differentiable.
4. From Lemma 2.2, if $\mu^c \in \mathcal{M}^c(x^*, 0)$ is a Lagrange multiplier of (1.2) and (4.1), then $||\mu^c||_1 = c$.

Therefore the convergence results from [1, 2] can be applied. However, we will assume that c_ζ and thus c are already determined.

The algorithm in [1] is based on the merit function of (1.2)

$$
(4.2)\qquad \phi(y) = f^c(y) + c_\phi \max\left\{g_1^c(y), g_2^c(y), \dots g_{m+1}^c(y)\right\},
$$

where $P(x)$ is the L_∞ penalty function defined in (1.16). We choose $c_\phi = 1 + c$, which, from the outlined properties of (4.1), satisfies

$$
c_\phi > \max_{\mu^c \in \mathcal{M}^c((x^*,0))} ||\mu^c||_1 = c.
$$

We use the following sequential quadratic programming (SQP) algorithm [1]:

1. Start with $k = 0$, $y = y^k$.
2. Compute the direction (d^k) the solution of the problem

$$
(4.3)\qquad \begin{array}{rl}
\min_d & \nabla_x f^c(y)^T d + \frac{1}{2} d^T d \\
\text{subject to} & g_i^c(y) + \nabla_x g_i^c(y)^T d \leq 0 \quad i = 1, 2, \dots, m+1.
\end{array}
$$

3. Take $y^{k+1} = y^k + \alpha^k y^k$, where α^k is a stepsize obtained by the Armijo rule [3, 4] applied to the merit function $\phi(y)$.
4. Take $k = k + 1$ and restart with step 1.

Since, as outlined above, the nonlinear program (4.1) has bounded Lagrange multipliers and satisfies the quadratic growth constraint, from [1] it follows that, when started sufficiently close to the point $y^* = (x^*, 0)$, this algorithm induces Q-linear convergence of $\phi(x^k) \to \phi(x^*)$ and R-linear convergence of $x^k \to x^*$. Given that the algorithm does not use second-order information, it is expected that the order of convergence will be generally linear. Obviously the same algorithm can be applied directly to (1.1), by replacing y with x, f^c with f, g^c with g and m by $m+1$ in the definition of the algorithm. However, if we apply the algorithm directly for (1.1), we do not have an initial estimate of the size of the Lagrange multiplier set, which is necessary to define the merit function $\phi(x)$ (1.20) with the appropriate c_ϕ (1.21). When applying this algorithm to (1.1) we use an updating procedure for c_ϕ that is common for the L_∞ penalty function [3, 4].

To illustrate the difficulties that appear in the context of problems with unbounded Lagrange multiplier sets, we consider an example on which we run the following well-established algorithms for nonlinear programming:

- LANCELOT [7], a Lagrange multiplier algorithm.
- LOQO [29], an interior-point approach.
- SNOPT [13], a sequential quadratic programming algorithm.
- FilterSQP [11], a sequential quadratic programming algorithm with a special merit criterion.
- LINF [1], the algorithm presented in the beginning of this section based on the descent direction (4.3) with the merit function $\phi(x)$ (1.20).

Except for LINF, which was coded in Matlab, all other algorithms were used with AMPL input on the NEOS server [23]. All tolerance parameters were set to 10^{-16}.

The example we are considering is (2.20) [27], which has unbounded Lagrange multipliers, which follows from (2.21). Since several algorithms for nonlinear programming are initiated at 0, and to avoid accidental convergence, we translate (2.20) by 1. We obtain

$$\begin{array}{rrcll} \min_x & f(x) & = & (x-1)^2 & \\ \text{subject to} & g_1(x) & = & (x-1)^6 \sin\frac{1}{x-1} & \leq 0. \\ & g_2(x) & = & -(x-1)^6 \sin\frac{1}{x-1} & \leq 0. \end{array} \tag{4.4}$$

The solution of the problem is $x=1$. From the form of the objective function it is immediate that the nonlinear program (4.4) satisfies the quadratic growth condition (1.17). From (2.21) the Lagrange multiplier set is not empty, and the data of the problem are at least twice continuously differentiable. Therefore, Theorem 3.4 applies.

This example is important because it shows that problems with unbounded multipliers do not generally have isolated stationary points. As outlined in [1], an accumulation of stationary points cannot occur at a local solution with bounded Lagrange multipliers and where the quadratic growth condition is satisfied.

Indeed, the feasible set of (4.4) is made of the points where $\sin\frac{1}{x-1}=0$, or $1+\frac{1}{k\pi}$, for integer k, $k\neq 0$, which accumulate at the solution $x=1$. Each such point is a local minimum and a stationary point. Hence, it is likely that an algorithm started close to the solution $x=1$ will, in fact stop at some of the other stationary points that are close to $x=1$.

The results of all algorithms on (4.4) are summarized in Table 4.1. The algorithm LINF is started at 0. With the exception of LANCELOT, all algorithms converge to $1-\frac{1}{\pi}$. LANCELOT converges to the solution of (4.4). LANCELOT enforces the nonlinear constraints by means of a penalty function, which may be responsible for avoiding the other local minima.

We now transform the nonlinear program (4.4), based on the Theorem 3.4. Specifically, we relax the constraints and add a penalty term with $c=1$ as indicated by

Table 4.1
Results for problem (4.4)

Solver Type	$\|x - x^*\|$	Iterations	Message
LANCELOT	3.09e-12	60	Step got too small
LOQO	3.18e-01	149	Primal and/or dual infeasible
SNOPT	3.18e-01	1	Optimal solution found
FilterSQP	3.18e-01	13	Optimal solution found
LINF	3.18e-01	13	Step got too small

Table 4.2
Results for the modified problem (4.5)

Solver Type	$\|x - x^*\|$	Iterations	Message
LANCELOT	2.18e-12	297	Step got too small
LOQO	2.9e-2	1000	Iteration limit (1000 iterations)
SNOPT	5.6e-12	10	The current point cannot be improved
FilterSQP	7.45 e-13	42	Optimal solution found
LINF	0	39	Optimal solution found

(1.2). We obtain the nonlinear program

$$\begin{array}{rrcrrl} \min_{x,\zeta} & f^c(y) & = & (x-1)^2+\zeta & & \\ \text{subject to} & g_1^c(y) & = & (x-1)^6 \sin\frac{1}{x-1} & -\zeta & \leq 0 \\ & g_2^c(y) & = & -(x-1)^6 \sin\frac{1}{x-1} & -\zeta & \leq 0 \\ & g_3^c(y) & = & & -\zeta & \leq 0, \end{array} \tag{4.5}$$

where $y = (x, \zeta)$, as outlined in the generic modified nonlinear program (4.1). The results of applying the algorithms to the modified nonlinear program (4.5) (whose solution is $(1, 0)$, from Theorem 3.4) are illustrated in Table 4.2. The algorithm LINF is started at $(0, 0)$, which is the analogue of starting LINF at 0 for (4.4). We monitor only the accuracy in determining the first variable, x^*, since this indicates how close we are both to the solution of (4.4), which is the problem we are trying to solve, and (4.5), the modified problem. For the algorithms on NEOS [23] no modification (such as a specific initial point) was attempted. One conclusion from Table 4.2 is that the modification (1.2) is beneficial for all algorithms, which now all converge to the global solution of the original problem (4.4) (with the exception of LOQO, which terminates early, though increased accuracy can be observed for that case as well).

The fact that (2.22) has bounded Lagrange multipliers results in the fact that y^* is an isolated stationary point [1]. This enables all algorithms considered in our experiment to converge to the solution. The Q-linear convergence of the merit function $\phi(y)$ is demonstrated in Table 4.3.

From Theorem 3.4, we can also obtain superlinear convergence for (1.2) at a cost of computing the second-order derivatives for f and g and solving more complicated subproblems with quadratic constraints. We use the sequential quadratically constrained quadratic programming (SQCQP) algorithm from [2] on (4.1):

1. Choose a starting point y^0, $k = 0$.

TABLE 4.3
Q-linear convergence of $\phi(y)$

Iteration	$\frac{\phi(y^k)-\phi(y^*)}{\phi(y^{k+1})-\phi(y^*)}$
5	3.79
10	7.88
15	8.99
20	1.18
25	8.99
30	9.00
35	9.00
39	9.00

2. Let $y = x^k$, and determine d^k, a stationary point of

$$\begin{array}{llrl} & \min_d & \nabla_x f^c(y)^T d + \frac{1}{2} d^T \nabla^2_{xx} f^c(y) d & \\ (4.6) & \text{subject to} & g_i^c(y) + \nabla_x g_i^c(y)^T d + \frac{1}{2} d^T \nabla^2_{xx} g_i^c(y) d \leq & 0, \\ & & i = 1, 2, \ldots, m & \\ & & d^T d \leq & \gamma^2. \end{array}$$

3. Take $y^{k+1} = y^k + d^k$ and $k = k + 1$ and restart.

The quantity γ defines a trust region constraint. Since, from Theorem 3.4, (1.2) satisfies MFCQ (1.10) and the quadratic growth condition, it follows that if the algorithm is started sufficiently close to $(x^*, 0)$ for a sufficiently small γ, then [2]

1. The trust region constraint will be inactive.
2. The sequence (y^k) is superlinearly convergent to $(x^*, 0)$:

$$\lim_{k\to\infty} \frac{||y^{k+1} - (x^*, 0)||}{||y^k - (x^*, 0)||} = 0.$$

Undoubtedly, the subproblems of SQCQP are not easy to solve, since both the objective function and the constraints are nonconvex and nonlinear. However, recent approaches have shown that an efficient solution of the subproblems can be obtained by semidefinite relaxation [18].

For both algorithms, the main issue is how to determine an appropriate value of c that satisfies the conclusions of Theorem 3.4. This problem is typical of penalty functions approaches [3, 4, 20, 24], which are the justification for the modified nonlinear program (1.2). One could, of course, pick a c by a trial-and-error procedure. But a too large c would distort the nonlinear program (1.2) by overemphasizing the importance of the constraints and possibly slowing progress of optimization algorithms.

For the traditional constraint qualification, it is shown that a simple update procedure based on the multipliers obtained at every step of a sequential quadratic programming algorithm will identify a valid value of c [4]. For that case (1.2) and (1.1) are equivalent in the sense that the sets of Lagrange multipliers corresponding to the first m constraints are identical for sufficiently large c (as can also be seen from Lemma 2.2). In the case discussed in this work, however, this does not occur, since the Lagrange multiplier set $\mathcal{M}(x^*)$ of (1.1) may be unbounded, whereas the multiplier set $\mathcal{M}^c((x^*, 0))$ of (1.2) is bounded.

5. Conclusions. We construct nonlinear programming algorithms that are locally linearly convergent, using only first-order information, and superlinearly convergent under only the assumptions of quadratic growth and a nonempty but not necessarily bounded Lagrange multiplier set. An important class of problems that do not generally have bounded Lagrange multiplier sets are the mathematical programs with equilibrium constraints [19, 25]. The results are achieved by relaxing the constraints and adding a linear penalty term with a sufficiently large parameter c to the objective function. The effect of the penalty term is to retain from the Lagrange multipliers of the original problem only those whose 1 norm is less than or equal to c. An important achievement of our approach compared with [20] is that the new formulation involves only differentiable functions which makes it substantially easier to solve.

The modified problem has the same solution as the original one (in the first n variables) and satisfies the quadratic growth condition as well. In addition, however, the modified problem has a nonempy and bounded Lagrange multiplier set. For nonlinear programs that satisfy these conditions, we can use the algorithms from [1, 2] to obtain linear and superlinear convergence to the solution of the modified problem.

We ran some well-established algorithms, as well as the algorithm from [1], on an example whose solution, because of the unboundedness of the Lagrange multiplier set, is not an isolated stationary point. We demonstrate that most of the algorithms stop at one of the neighboring stationary points. Applying these algorithms to the modified nonlinear program, however, results in convergence to the solution of the problem. This is a highly desired outcome for any nonlinear programming problem.

One of the issues related to this approach, as for any other method originating from a penalty method, is that the appropriate c (and γ for SQCQP) has to be estimated in practice. This question is fundamentally connected to inducing global convergence to a local minimum of (1.1), which will be approached in a future work.

Acknowledgments. Thanks to Sven Leyffer and Danny Ralph for valuable pointers concerning MPEC and the quadratic growth condition. Thanks to Steve Wright for the countless discussions regarding degenerate nonlinear programming.

REFERENCES

[1] M. Anitescu, *Degenerate nonlinear programming with a quadratic growth condition.* SIAM Journal on Optimization, 10:4 (2000), pp. 1116-1135.

[2] M. Anitescu, *A superlinearly convergent sequential quadratically constrained quadratic programming algorithm for degenerate nonlinear programming*, preprint.

[3] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[4] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusets, 1995.

[5] J. F. Bonnans, *Local analysis of Newton-type methods for variational inequalities and nonlinear programming*, Applied Mathematics and Optimization, 29 (1994), pp. 161–186.

[6] J. F. Bonnans and A. Ioffe, *Second-order sufficiency and quadratic growth for nonisolated minima*, Mathematics of Operations Research , 20:4 (1995), pp. 801–819.

[7] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization*, Springer Verlag, Berlin, 1992.

[8] A. V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York, 1983.

[9] A. Fischer, *Modified Wilson's method for nonlinear programs with nonunique multipliers*, Mathematics of Operations Research 24:3 (1999), pp. 699–727.

[10] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Chichester, 1987.

[11] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Numerical Analysis Report, NA/171, Department of Mathematics, University of Dundee, UK.

[12] J. GAUVIN, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Mathematical Programming 12 (1977), pp. 136–138.

[13] P. E. GILL, W. MURRAY AND M. A. SAUNDERS, *User's guide for SNOPT 5.3: A Fortran package for large-scale nonlinear programming*, Report NA 97-5, Department of Mathematics, University of California, San Diego, 1997.

[14] W. W. HAGER, *Stabilized sequential quadratic programming*, To appear in Computational Optimization and Applications .

[15] W. W HAGER AND M. S. GOWDA, *Stability in the presence of degeneracy and error estimation*, Technical Report, Department of Mathematics, University of Florida, 1997.

[16] A. IOFFE, *Necessary and sufficient conditions for a local minimum.3: Second order conditions and augmented duality*, SIAM Journal of Control and Optimization, 17:2 (1979), pp. 266–288.

[17] A. IOFFE, *On Sensitivity analysis of nonlinear programs in Banach spaces: the approach via composite unconstrained optimization*, SIAM Journal of Optimization, 4:1 (1994), pp. 1–43.

[18] S. KRUK AND H. WOLKOWICZ, *SQQP, Sequential Quadratic Constrained Quadratic Programming*, Research Report CORR 97-01, University of Waterloo, Waterloo, 1997.

[19] Z.-Q. LUO, J.-S. PANG AND D. RALPH, *Mathematical Programs with Equilibrium Constraints* Cambridge University Press, 1996.

[20] Z.-Q. LUO, J.-S. PANG, D. RALPH AND S.-Q. WU, *Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints.* Mathematical Programming, 34(2):142-162, 1986.

[21] O. L. MANGASARIAN, *Nonlinear Programming*, McGraw-Hill, New York 1969.

[22] O. L. MANGASARIAN AND S. FROMOVITZ, *The Fritz John necessary optimality conditions in the presence of equality constraints*, J. Math. Anal. and Appl. 17 (1967), pp. 34-47.

[23] *The NEOS Guide*. Available online at *http://www.mcs.anl.gov/otc/Guide*.

[24] E. POLAK, *Optimization*, Springer, New York, 1997.

[25] J. OUTRATA, M. KOCVARA and J. ZOWE, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, Kluwer Academic Publishers, Dordrecht, 1998.

[26] D. RALPH AND S. J. WRIGHT, *Superlinear convergence of an interior-point method despite dependent constraints*, Preprint ANL/MCS-P622-1196, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., 1996.

[27] S. M. ROBINSON, *Generalized equations and their solutions, Part II: Applications to nonlinear programming* Mathematical Programming Study 19 (1980), 200–221.

[28] A. SHAPIRO, *Sensitivity analysis of nonlinear programs and differentiability properties of metric projections*, SIAM J. Control and Optimization , 26:3 (1988), pp. 628–645.

[29] R. J. VANDERBEI, *LOQO: An interior-point code for quadratic programming*, Technical Report SOR-94-21, Statistics and Operations Research, Princeton University, Princeton, 1994.

[30] S. J. WRIGHT, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Computational Optimization and Applications 11 (1998), pp. 253–275.

[31] S. J. WRIGHT, *Modifying SQP for degenerate problems*, Preprint ANL/MCS-P699-1097, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill, 1997.

[32] W. ZHU and L. PETZOLD, *Model reduction for chemical kinetics: An optimization approach*, AIChE Journal April 1999, pp. 869–886.