Preprint ANL/MCS-P761-0699.

Mihai Anitescu*

Degenerate Nonlinear Programming with a Quadratic Growth Condition

July 7, 1999

Abstract. We show that the quadratic growth condition and the Mangasarian-Fromowitz constraint qualification imply that local minima of nonlinear programs are isolated stationary points. As a result, when started sufficiently close to such points, an L_∞ exact penalty sequential quadratic programming algorithm will induce at least R-linear convergence of the iterates to such a local minimum. We construct an example of a degenerate nonlinear program with a unique local minimum satisfying the quadratic growth and the Mangasarian-Fromowitz constraint qualification but for which no positive semidefinite augmented Lagrangian exists. We present numerical results obtained using several nonlinear programming packages on this example, and discuss its implications for some algorithms.

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 [14,13, 20,25]. This research has been 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). It is therefore important to establish to what extent the convergence properties of the SQP methods are dependent on the ill-conditioning of the constraints. In this work, 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.

Many of the previous analysis and rate of convergence results for degenerate NLP are based on the validity of second-order conditions. These are essentially equivalent to the condition in unconstrained optimization that, for a critical point of a function $f(x)$ to be a local minimum, $f_{xx} \succeq 0$ is a necessary condition and $f_{xx} \succ 0$ is a sufficient condition. Here $\succeq$ is the positive semidefinite ordering. The place of f_{xx} in constrained optimization is taken for these conditions by L_{xx}, the Hessian of the Lagrangian, which is now required to be positive definite on the critical cone for one or all of the Lagrange multipliers [6,21].

Department of Mathematics, University of Pittsburgh

* This work was completed while the author was the Wilkinson Fellow at the Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439. 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 differs from previous approaches in that we assume only that

1. At a local solution x^* of the constrained nonlinear program, the first-order Mangasarian-Fromowitz constraint qualification holds.
2. The quadratic growth condition (QG) [4,16] is satisfied:

$$f(x) \geq f(x^*) + \sigma ||x - x^*||^2 \tag{1}$$

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

These assumptions are equivalent to a weaker form of the second-order sufficient conditions [15,4] which do not require the positive semidefinitenes of the Hessian of the Lagrangian on the entire critical cone.

We prove that these conditions guarantee that x^* is the only local stationary point (3) of the nonlinear program. This is an important issue because it guarantees that descent-like algorithms will not stop arbitrarily close to x^*, except at x^*. This extends a result from [21] that required some second-order sufficient conditions to be satisfied for all multipliers. In particular, our work implies that if MFCQ holds and the second-order sufficient conditions hold for one multiplier, then x^* is a strict local minimum and an isolated stationary point.

We also show that, under the same assumptions, the L_∞ exact penalty sequential quadratic program (SQP) induces at least Q linear convergence [19] of the penalized objective to $f(x^*)$ and R-linear convergence of the iterates. Finally, we provide an example of a nonlinear program that satisfies our assumptions for which it is not possible to construct an augmented Lagrangian such that x^* will be an unconstrained local minimum. This may present an adverse case to algorithms based on this assumption, such as Lagrange multiplier methods. However, we show that it is possible to construct a nondifferentiable function that has x^* as its minimum, namely the L_∞ penalty function (which can also be inferred from the results in [4]). We describe our computational experience with several nonlinear programming packages applied to this example and discuss the expected and observed behavior of Lagrangian multiplier methods.

Our convergence analysis for the L_∞ exact penalty function suggests that it is possible to construct a convergence theory with much more general second-order conditions. This may result in algorithms with superior robustness, because their properties depend on significantly fewer assumptions.

1.1. Previous Work, Framework, and Notations

We deal with the NLP problem

$$\min_x f(x) \quad \text{subject to} \quad g(x) \leq 0, \tag{2}$$

where $f : \mathsf{R}^n \to \mathsf{R}$ and $g : \mathsf{R}^n \to \mathsf{R}^m$ are twice continuously differentiable.

We call x a stationary point if the following conditions hold for some $\lambda \in \mathsf{R}^m$:

$$\mathcal{L}_x(x, \lambda) = 0, \;\; \lambda \geq 0, \;\; g(x) \leq 0, \;\; \lambda^T g(x) = 0. \tag{3}$$

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

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T g(x). \tag{4}$$

If certain regularity conditions hold (discussed below), then a local solution x^* of (2) is a stationary point. In that case (3) are referred to as the KKT (Karush-Kuhn-Tucker) conditions.

Since our analysis will be limited to a neighborhood of a point x^* that is a strict minimum, we will assume that all constraints are active at x^*, or $g(x^*) = 0$. Such a situation can be obtained by simply dropping the constraints i for which $g_i(x^*) < 0$, since this relationship holds in an entire neighborhood of x^*. This does not reduce the generality of our results, but it simplifies the notation because now we do not have to refer separately to the active set.

The regularity condition, or constraint qualification, ensures that a linear approximation of the feasible set in the neighborhood of x^* captures the geometry of the feasible set. Often in local convergence analysis of constrained optimization algorithms, it is assumed that the constraint gradients $\nabla g_i(x^*)$, $i = 1..m$ are linearly independent, so that the Lagrange multiplier in (3) is unique. We assume instead the Mangasarian-Fromowitz constraint qualification (MFCQ):

$$\nabla g_i(x^*)^T p < 0, \quad \text{for all } i \text{ and some } p \in \mathsf{R}^n. \tag{5}$$

It is well known [9] that MFCQ is equivalent to boundedness of the set $\mathcal{M}(x^*)$ of Lagrange multipliers that satisfy (3), that is,

$$\mathcal{M}(x^*) \stackrel{\text{def}}{=} \{\lambda \geq 0 \,|\, (x^*, \lambda) \ \text{ satisfy (3)}\}. \tag{6}$$

Note that $\mathcal{M}(x^*)$ is certainly polyhedral in any case.

The critical cone at x^* is [6,22]

$$\mathcal{C} = \{u \in \mathsf{R}^n \,|\nabla g_i(x^*)^T u \leq 0, \ i = 1, \ldots, m; \ \nabla f(x^*)^T u = 0\} \tag{7}$$

We briefly review the some of the second-order conditions in the literature, although they are not an assumption for our analysis but only a basis for comparison. In the framework of [6], the second-order sufficient conditions for x^* to be an isolated local solution of (2) are:

$$\exists \lambda^* \in \mathcal{M}(x^*), \ \exists \sigma > 0 \text{ such that } v^T \mathcal{L}_{xx}(x^*, \lambda^*) v \geq \sigma \|v\|_2^2, \ \ \forall v \in \mathcal{C}. \tag{8}$$

If these conditions hold at x^* for some λ^*, then the quadratic growth condition is satisfied, irrespective of the validity of the first-order constraint qualification [6,7]. However, this does not imply that x^* is an isolated stationary point, as shown by a simple example [21], which may prevent an optimization algorithm that uses only first derivative information from reaching x^* even when started arbitrarily close to x^*.

In [21] it is shown that if MFCQ holds, and the relation (8) is satisfied for all $\lambda^* \in \mathcal{M}(x^*)$ then x^* is an isolated stationary point and a minimum of (2). Also, with these conditions, the exact solution is Lipschitz stable with respect to perturbations. By compactness of $\mathcal{M}(x^*)$, we can choose σ independently of

λ^* in this case. In [1] it is proven that, under these assumptions, the L_∞ exact penalty SQP will converge Q-linearly to $f(x^*)$, when the descent direction is computed by a QP using only first-order information.

A refinement of the second-order conditions was introduced in [15]. In the presence of MFCQ, those conditions require that

$$\forall u \in \mathcal{C},\ \exists \lambda^* \in \mathcal{M}(x^*), \text{ such that } u^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) u > 0. \tag{9}$$

Further analysis shows that, in presence of MFCQ, these conditions are necessary and sufficient for the quadratic growth condition to hold [4,15,16,22]. Also, the exact solution is Lipschitz stable with respect to certain classes of perturbations [22], though not to any perturbation (see an example in [10, p.308]). In this paper we assume only the quadratic growth condition and MFCQ, and thus we do not use the perturbation results.

If the condition (9) holds, but (8) does not, then there is no positive semidefinite augmented Lagrangian, as we will show with an example. This is an interesting aspect since it invalidates the usual working assumption of Lagrange multiplier methods [3].

Finally, we review some of the facts concerning the L_∞ nondifferentiable exact penalty function:

$$P(x) = \max\{0, g_1(x), ...g_m(x)\}. \tag{10}$$

We are looking for an unconstrained minimum of the function

$$\phi(x) = f(x) + c_\phi P(x),$$

where c_ϕ is a sufficiently large constant. Descent directions d of $\phi(x)$ at the point x can be obtained by solving the following quadratic program (QP) [3]:

$$\begin{array}{ll} \text{minimize} & \nabla f(x)^T d + \frac{1}{2} d^T H d + c_\phi \zeta \\ \text{subject to} & g_j(x) + \nabla g_j(x)^T d \leq \zeta, \qquad j = 0, 1, 2...m, \end{array} \tag{11}$$

where H is some positive definite matrix and $g_0(x) = 0$. In this paper the analysis will be restricted to the case $H = I$, although the same results apply for any other positive definite matrix.

At the current point x^k of an iterative procedure that attempts to determine x^*, the QP (11) generates the descent direction d^k. The next iterate is $x^{(k+1)} = x^k + \alpha^k d^k$, where α^k is obtained by a line search procedure. Usual stepsize rules are the minimization rule, the limited minimization rule, and the Armijo rule [3]. For these rules, any limit point of $\{x^k\}$ is a stationary point of $\phi(x)$, and the descent procedure is therefore globally convergent in this sense [3].

If, in addition,

$$c_\phi > \sum_{j=1}^{m} \lambda_j^* \tag{12}$$

for some $\lambda^* \in \mathcal{M}(x^*)$, then x^* is a stationary point of $\phi(x)$ [2]. A suitable value for c_ϕ is not available in the early stages of the algorithm, but a good estimate can be found via an update scheme [2]. Here we assume that c is constant and

$$c_\phi > \sum_{j=1}^{m} \lambda_j^* + 2\gamma \tag{13}$$

for all $\lambda^* \in \mathcal{M}(x^*)$, where γ is some prescribed safety factor.

Consider the quadratic program

$$\begin{array}{ll} \text{minimize} & \nabla f(x)^T d + \frac{1}{2} d^T d \\ \text{subject to} & g_j(x) + \nabla g_j(x)^T d \leq 0 \quad j = 0, 1, 2 \ldots m. \end{array} \tag{14}$$

We denote the unique solution of this program by $d(x)$ and the set of its multipliers by $\mathcal{M}(x)$. At x^* (14) has the same multiplier set as (2), which are both denoted by $\mathcal{M}(x^*)$. Since MFCQ is satisfied at x^*, this QP is feasible in a neighborhood of x^*. The KKT conditions for this QP require

$$\begin{array}{c} d(x) + \nabla f(x) + \nabla g(x)\lambda = 0 \\ \lambda \geq 0, \;\; g(x) + \nabla g(x)^T d(x) \leq 0, \;\; \lambda^T(g(x) + \nabla g(x)^T d(x)) = 0. \end{array} \tag{15}$$

With these notations, $d(x^*) = 0$. If the QP (14) were unconstrained, then its solution would be $d(x) = -\nabla f(x)$. We name a descent-like algorithm a sequential quadratic program that solves instances of the above QP.

At x^*, the QP (14) satisfies MFCQ and some second-order sufficient conditions. From [21] there exists c_d such that, in a neighborhood of x^* we have $||d|| \leq c_d ||x - x^*||$ and, $\forall \lambda \in \mathcal{M}(x)$, there exists $\lambda^* \in \mathcal{M}(x^*)$ such that

$$||\lambda - \lambda^*|| \leq c_d ||x - x^*||. \tag{16}$$

Therefore, from the definition of c_ϕ, there exists a neighborhood of x^* such that

$$c_\phi > \gamma + \sum_{i=1}^{m} \lambda_i \tag{17}$$

for all multipliers $\lambda \in \mathcal{M}(x)$. For such x, it can be verified by inspection that $(d(x), \zeta = 0)$ is a solution of (11) [2, p. 195]. We therefore concentrate on the QP (14), because, if c_ϕ is large enough and we are sufficiently close to x^*, it generates the same descent direction as (11), thus sharing its global convergence property.

For some function $h : \mathsf{R}^n \to \mathsf{R}^k$ we denote by c_{1h}, c_{2h} bounds depending on the first and second derivatives of h. The positive and negative parts of $h(x)$ are $h^+(x) = \max\{h(x), 0\}$ and respectively, $h^-(x) = \max\{-h(x), 0\}$, both taken componentwise. With this notation $h(x) = h^+(x) - h^-(x)$.

2. Stationary Points of NLPs Satisfying MFCQ

In this section, we assume that x is in a sufficiently small neighborhood of x^*, whose size or properties are specified in each of the following results. In particular the standing assumptions hold on all neighborhoods considered here and

$$\nabla g_i(x)^T p < -\zeta_0, \text{ for all } i \text{ and } x \in W(x^*). \tag{18}$$

Here p with $||p|| = 1$ is one of the vectors satisfying (5), $\zeta_0 > 0$ and $W(x^*)$ is a suitable neighborhood of x^*.

Lemma 1. *There exist $\bar{\alpha}_P > 0$, $c_P > 0$, and a neighborhood $W(x^*)$ such that*

$$g(x) \leq 0,\, g_i(x) = 0 \textit{ for some } i,\ 1 \leq i \leq m \Rightarrow P(x - \alpha p) \geq c_P \alpha,\ \forall \alpha \in [0, \bar{\alpha}_P].$$

Here $P(x)$ is the usual L_∞ penalty function (10).

Proof. We have by Taylor's theorem

$$g_i(x - \alpha p) \geq -\alpha \nabla g_i(x)^T p - c_{2g}\alpha^2 \geq \alpha\zeta_0 - \alpha^2 c_{2g}$$

We choose

$$\bar{\alpha}_P = \frac{\zeta_0}{2c_{2g}}. \tag{19}$$

For $0 \leq \alpha \leq \alpha_P$ we have

$$g_i(x - \alpha p) \geq \alpha\zeta_0 - c_{2g}\alpha^2 = \alpha(\zeta_0 - \alpha c_{2g}) \geq \alpha\frac{\zeta_0}{2}.$$

The claim follows after choosing $c_P = \frac{\zeta_0}{2}$.

The proof of the following lemma can be inferred from [4]. We include it here for completeness.

Lemma 2. *There exists a c_ϕ such that*

$$f(x) + c_\phi P(x) - f(x^*) \geq \frac{\sigma}{2}||x - x^*||^2 \tag{20}$$

for all x in a neighborhood of x^*.

Proof. Let $r > 0$ be such that $B(x^*, r) \subset W(x^*)$. We choose $r_1 < \frac{r}{2}$ such that $\alpha = \frac{P(x)}{\zeta_0} < \min\{\bar{\alpha}_P, r/2\}$ for $x \in B(x^*, r_1)$. This is always possible because $P(x^*) = 0$. We then have that, for any $x \in B(x^*, r_1)$,

$$||x + \alpha p - x^*|| \leq ||x - x^*|| + \alpha \leq \frac{r}{2} + \frac{r}{2} = r \tag{21}$$

and thus $x + \alpha p \in B(x^*, r)$. By the intermediate value theorem, we have that $g_i(x + \alpha p) = g_i(x) + \alpha \nabla g_i(x + \alpha^* p)^T p$, where $0 \leq \alpha^* \leq \alpha$ and thus $x + \alpha^* p \in B(x^*, r)$, implying in turn that $\nabla g_i(x + \alpha^* p)^T p \leq -\zeta_0$. Therefore $g_i(x + \alpha p) \leq g_i(x) - \alpha\zeta_0 = g_i(x) - P(x) \leq 0$. Therefore $x + \alpha p$ is feasible.

Take now

$$\alpha_1 = \min\{\hat{\alpha} \geq 0 \,|\, g(x + \hat{\alpha} p) \leq 0\}. \tag{22}$$

If x is infeasible, then $\alpha_1 > 0$ and there exists i such that $g_i(x + \alpha_1 p) = 0$. Since $x + \alpha_1 p$ is feasible, and $0 \leq \alpha_1 \leq \bar{\alpha}_P$, Lemma 1 applies to give

$$P(x) \geq c_P \alpha_1. \tag{23}$$

If x is feasible, then $\alpha_1 = 0$ and $P(x) = 0$, and the previous bound still applies.

From the quadratic growth assumption (1) and the feasibility of $x + \alpha_1 p$, we must have that

$$f(x + \alpha_1 p) - f(x^*) \geq \sigma ||x - x^* + \alpha_1 p||^2$$

or

$$f(x) - f(x^*) \geq \sigma ||x - x^* + \alpha_1 p||^2 - (f(x + \alpha_1 p) - f(x)). \tag{24}$$

By (23) and Taylor's theorem we have

$$f(x + \alpha_1 p) - f(x) \leq c_{1f} \alpha_1 \leq \frac{c_{1f}}{c_P} P(x) \tag{25}$$

Choose

$$c_\phi = \frac{c_{1f}}{c_P} + \frac{\sigma \bar{\alpha}_P}{c_P}$$

Then by (23)

$$c_\phi P(x) = \frac{c_{1f}}{c_P} P(x) + \frac{\sigma \bar{\alpha}_P}{c_P} P(x) \geq \frac{c_{1f}}{c_P} P(x) + \sigma \bar{\alpha}_P \alpha_1 \geq \frac{c_{1f}}{c_P} P(x) + \sigma \alpha_1^2. \tag{26}$$

Using (25), (24) and (26) we get

$$f(x) - f(x^*) + c_\phi P(x) \geq \sigma ||x - x^* + \alpha_1 p||^2 + \sigma \alpha_1^2 = \sigma ||x - x^* + \alpha_1 p||^2 + \sigma ||\alpha_1 p||^2.$$

The conclusion follows, because

$$\sigma ||x - x^* + \alpha_1 p||^2 + \sigma ||\alpha_1 p||^2 \geq \frac{\sigma}{2} ||x - x^*||^2$$

from the Cauchy-Schwartz inequality.

We can assume that c_ϕ from the previous lemma satisfies (17), or otherwise we replace it with the right-hand side of (17) and the conclusion of the lemma still holds for the new c_ϕ.

To prove the following results, we will use the results from [12] concerning sets defined by linear inequalities:

$$\mathcal{P} = \{x \in \mathsf{R}^n \,|\, a_i^T x + b_i \leq 0, i = 1..m, ||a_i|| = 1\}. \tag{27}$$

For such a set, denote by $d(x, \mathcal{P})$ the distance from a point $x \in \mathsf{R}^n$ to the set $\mathcal{P}$. Also, denote by $d_{\mathcal{P}}(x)$ the maximum value of the infeasibility:

$$d_{\mathcal{P}}(x) = \max\{0, a_1^T x + b_1, \ldots a_m^T x + b_m\} \tag{28}$$

Then there exists a number $\mu^*(\mathcal{P}) > 0$ such that

$$0 < \mu^*(\mathcal{P}) d(x, \mathcal{P}) \leq d_{\mathcal{P}}(x) \leq d(x, \mathcal{P}), \forall x \in \mathsf{R}^n. \tag{29}$$

If we have equality constraints, we recast them as two inequality constraints. The following lemma uses the fact that $\mathcal{M}(x^*)$ is polyhedral and can thus be expressed in the form (27).

Lemma 3. *Let $\mathcal{I}$ be an index set such that there exists a multiplier $\bar{\lambda} \in \mathcal{M}(x^*)$ with $lam\bar{b}da_{\mathcal{I}} = 0$. Then there exists a constant $c_{\mathcal{I}}$ such that $\forall \lambda \in \mathcal{M}(x^*)$ there exists a $\lambda^* \in \mathcal{M}(x^*)$ with $\lambda^*_{\mathcal{I}} = 0$ and such that $||\lambda - \lambda^*|| \leq c_{\mathcal{I}} ||\lambda_{\mathcal{I}}||_\infty$.*

For a vector λ we have denoted by $\lambda_{\mathcal{I}}$ the restriction of the vector to the index set $\mathcal{I}$.

Proof. Let $\mathcal{M}_{\mathcal{I}}(x^*)$ be the set of all $\lambda^* \in \mathcal{M}(x^*)$ such that $\lambda^*_{\mathcal{I}} = 0$. Then $\nu \in \mathcal{M}_{\mathcal{I}}(x^*)$ satisfies

$$\sum_{j=1}^m \nabla g_j(x^*) \nu_j = -\nabla f(x^*), \tag{30}$$

$$\nu_{\mathcal{I}} = 0, \tag{31}$$

$$\nu \geq 0. \tag{32}$$

From our assumptions, $\mathcal{M}_{\mathcal{I}}(x^*)$ is not empty. By eventually rescaling the x space, we can assume, without loss of generality, that the vectors defining the equality constraints in (30) are of norm 1; otherwise, if all entries are 0, we remove that row, and the feasible λ set remains unchanged. $\mathcal{M}_{\mathcal{I}}$ can be described in the following, alternative, way:

$$\sum_{j=1}^m \nabla g_j(x^*) \nu_j + \nabla f(x^*) \leq 0, \tag{33}$$

$$\sum_{j=1}^m -\nabla g_j(x^*) \nu_j - \nabla f(x^*) \leq 0, \tag{34}$$

$$\nu_{\mathcal{I}} \leq 0, \tag{35}$$

$$-\nu_{\mathcal{I}} \leq 0, \tag{36}$$

$$\nu \geq 0, \tag{37}$$

where each row is described by a unit vector, which puts the set in the form (27). Thus from [12] there exists a $\mu^*(\mathcal{M}_{\mathcal{I}}) > 0$ such that

$$\mu^*(\mathcal{M}_{\mathcal{I}}) d(\lambda, \mathcal{M}_{\mathcal{I}}) \leq d_{\mathcal{M}_{\mathcal{I}}}(\lambda). \tag{38}$$

However, since $\lambda \in \mathcal{M}(x^*)$ is a valid multiplier set, we have that only the constraints $\lambda_{\mathcal{I}} \leq 0$, (35), are violated. Thus $d_{\mathcal{M}_{\mathcal{I}}} = ||\lambda_{\mathcal{I}}||_\infty$. The conclusion follows from (38) by taking $c_{\mathcal{I}} = \frac{1}{\mu^*(\mathcal{M}_{\mathcal{I}})}$. The proof is complete.

We define

$$c_\lambda = \max_{\mathcal{I} \subset \{1,..,m\}} c_{\mathcal{I}}, \quad \text{for feasible } \mathcal{M}_{\mathcal{I}}(x^*). \tag{39}$$

Lemma 4. *There exists a neighborhood W of x^* such that, $\forall x \in W, \lambda \in \mathcal{M}(x)$, $\lambda_{\mathcal{I}} = 0$ implies that there exists a $\lambda^* \in \mathcal{M}(x^*)$ with $\lambda^*_{\mathcal{I}} = 0$.*

Proof. Assume the contrary. Then there exists a sequence $x^k \to x^*$ such that there exists $\lambda^k \in \mathcal{M}(x)$ and an index set $\mathcal{I}$ for which $\lambda_{\mathcal{I}} = 0$, but $\lambda^*_{\mathcal{I}} \neq 0$, $\forall \lambda^* \in \mathcal{M}(x^*)$. Since there is only a finite set of index sets, we can extract an infinite subsequence for which the above happens for a fixed set $\mathcal{I}$. By extracting another subsequence, we can assume that λ^k is convergent, from (16) and the fact that $\mathcal{M}(x^*)$ is compact.

But then $\lambda^k \to \lambda^* \in \mathcal{M}(x^*)$ and $\lambda^*_{\mathcal{I}} = 0$, a contradiction.

From here on we will use extensively that, for h twice continuously differentiable, we have

$$||h(x)-h(x^*)-\frac{(\nabla h(x)+\nabla h(x^*))^T}{2}(x-x^*)|| \leq \psi_{3h}(||x-x^*||)||x-x^*||^2, \quad (40)$$

where $\psi_{3h}(z) : \mathsf{R} \to \mathsf{R}$ is a continuous function with $\psi_{3h}(0) = 0$. Indeed by Taylor's theorem we have that there exist continuous functions $\psi^1_{3h}(z) : \mathsf{R} \to \mathsf{R}$ and $\psi^2_{3h}(z) : \mathsf{R} \to \mathsf{R}$ with $\psi^1_{3h}(0) = \psi^2_{3h}(0) = 0$

$$\begin{aligned} ||h(x) - h(x^*) - \nabla_x h(x^*)^T(x-x^*) - \frac{1}{2}(x-x^*)^T \nabla_{xx} h(x^*)(x-x^*)|| \\ \leq \psi^1_{3h}(||x-x^*||)||x-x^*||^2, \end{aligned}$$

and

$$\begin{aligned} || \ & \frac{(\nabla_x h(x) + \nabla_x h(x^*))^T}{2}(x-x^*) - \frac{(\nabla_x h(x^*) + \nabla_x h(x^*))^T}{2}(x-x^*) - \\ & \frac{1}{2}(x-x^*)^T \nabla_{xx} h(x^*)(x-x^*)|| \leq \psi^2_{3h}(||x-x^*||)||x-x^*||^2. \end{aligned}$$

The relation (40) now follows by comparing the last two equations.

Theorem 1. *There exists a constant $c_\sigma > 0$ such that in a neighborhood of x^* we have that*

$$||d||^2 + P(x) + \lambda^T g^-(x) \geq c_\sigma ||x-x^*||^2, \quad (41)$$

where (d, λ) is the solution of the QP (14).

Proof. From (16), there exists a $\lambda^* \in \mathcal{M}(x^*)$ such that $||\lambda - \lambda^*|| \leq c_d ||x - x^*||$. Let $\mathcal{I}$ be the set of indices i for which $\lambda_i = 0$. We have $||\lambda^*_{\mathcal{I}}||_\infty = ||\lambda^*_{\mathcal{I}} - \lambda_{\mathcal{I}}||_\infty \leq c_d||x-x^*||$. From (39) and Lemmas 3 and 4 there exists a $\tilde{\lambda} \in \mathcal{M}(x^*)$ with $\tilde{\lambda}_{\mathcal{I}} = 0$ and $||\tilde{\lambda} - \lambda^*|| \leq c_\lambda ||\lambda^*_{\mathcal{I}}||_\infty \leq c_\lambda c_d ||x-x^*||$. As a result

$$||\lambda - \tilde{\lambda}|| \leq ||\lambda - \lambda^*|| + ||\lambda^* - \tilde{\lambda}|| \leq (c_d + c_d c_\lambda)||x-x^*|| \quad (42)$$

and $\lambda_i = 0 \Rightarrow \tilde{\lambda}_i = 0$. The important consequence of this fact, using the complementarity relations from (15), is that

$$\begin{gathered} (\lambda_i + \tilde{\lambda}_i) g_i(x) = 2\lambda_i g(x) + (\tilde{\lambda}_i - \lambda_i) g(x) = \\ -(\tilde{\lambda}_i - \lambda_i)\nabla g_i(x)^T d + 2\lambda_i g_i(x) \quad \forall i,\ 1 \leq i \leq m. \end{gathered} \quad (43)$$

Indeed, $\lambda_i > 0$ implies $g_i(x) + \nabla g_i(x)^T d = 0$, whereas $\lambda_i = 0$ implies $\tilde{\lambda}_i = 0$ and all the above equalities are 0.

From Lemma 2 we have that

$$
\begin{aligned}
\frac{\sigma}{2}\|x-x^*\|^2 \leq f(x) - f(x^*) + c_\phi P(x) \leq && (44)\\
\psi_{3f}(\|x-x^*\|)\|x-x^*\|^2 + \frac{1}{2}(\nabla f(x) + \nabla f(x^*))(x-x^*) + c_\phi P(x) = && (45)\\
\psi_{3f}(\|x-x^*\|)\|x-x^*\|^2 + c_\phi P(x) + && (46)\\
\frac{1}{2}(-d - \nabla g(x)\lambda - \nabla g(x^*)\tilde{\lambda})^T(x-x^*). && (47)
\end{aligned}
$$

Here (d, λ) is a solution of (15), and $\tilde{\lambda} \in \mathcal{M}(x^*)$ satisfies (3). We also used (40). We now employ the identity $ab + cd = \frac{1}{2}((a+c)(b+d) + (a-c)(b-d))$, (42), and Taylor's theorem for $\nabla g(x)$ to get, by continuing from the previous equation,

$$
\begin{aligned}
\frac{\sigma}{2}\|x-x^*\|^2 \leq \psi_{3f}(\|x-x^*\|)\|x-x^*\|^2 && (48)\\
+\frac{1}{2}(-d - \frac{1}{2}(\nabla g(x) + \nabla g(x^*))(\lambda + \tilde{\lambda}) && (49)\\
-\frac{1}{2}(\nabla g(x) - \nabla g(x^*))(\lambda - \tilde{\lambda}))^T(x-x^*) + c_\phi P(x) \leq && (50)\\
(\psi_{3f}(\|x-x^*\|) + c_{2g}c_d(1+c_\lambda)\|x-x^*\|)\|x-x^*\|^2 + && (51)\\
c_\phi P(x) - \frac{1}{2}d^T(x-x^*) - \frac{1}{4}[(\nabla g(x) + \nabla g(x^*))(\lambda + \tilde{\lambda})]^T(x-x^*). && (52)
\end{aligned}
$$

We denote

$$
\lambda_\infty = \max_{\lambda^* \in \mathcal{M}(x^*)} \max_{i=1..m} \lambda_i^*. \qquad (53)
$$

From (16) $\|\lambda + \tilde{\lambda}\|_\infty \leq 2\lambda_\infty$ for x sufficiently close to x^*. By using λ_∞, (40), (43) and (42), we get

$$
\begin{aligned}
-\frac{1}{4}[(\nabla g(x) + \nabla g(x^*))(\lambda + \tilde{\lambda})]^T(x-x^*) = && (54)\\
-\frac{1}{4}(x-x^*)^T[(\nabla g(x) + \nabla g(x^*))(\lambda + \tilde{\lambda})] \leq && (55)\\
2\lambda_\infty \psi_{3g}(\|x-x^*\|)\|x-x^*\|^2 - \frac{1}{2}(\lambda + \tilde{\lambda})^T g(x) = && (56)\\
2\lambda_\infty \psi_{3g}(\|x-x^*\|)\|x-x^*\|^2 + \frac{1}{2}(\tilde{\lambda} - \lambda)^T \nabla g(x)^T d - \lambda^T g(x) \leq && (57)\\
2\lambda_\infty \psi_{3g}(\|x-x^*\|)\|x-x^*\|^2 + c_{1g}(c_d + c_d c_\lambda)\|x-x^*\|\|d\| + \lambda^T g^-(x), && (58)
\end{aligned}
$$

since $-\lambda^T g(x) = \lambda^T g^-(x) - \lambda^T g^+(x)$. Using the above bound in (52), together with $-d^T(x-x^*) \leq \|d\|\|x-x^*\|$, we get

$$
\begin{aligned}
\frac{\sigma}{2}\|x-x^*\|^2 \leq (\psi_{3f}(\|x-x^*\|) + c_{2g}c_d(1+c_\lambda)\|x-x^*\| + && (59)\\
2\lambda_\infty^* \psi_{3g}(\|x-x^*\|))\|x-x^*\|^2 + && (60)\\
c_\phi P(x) + \|d\|\|x-x^*\| + c_{1g}(c_d + c_d c_\lambda)\|x-x^*\|\|d\| + \lambda^T g^-(x) = && (61)\\
c_\phi P(x) + \lambda^T g^-(x) + B\|x-x^*\|\|d\| + \psi(\|x-x^*\|)\|x-x^*\|^2, && (62)
\end{aligned}
$$

where $B = \frac{1}{2} + c_{1g}(c_d + c_d c_\lambda)$ and $\psi(\|x-x^*\|) = (\psi_{3f}(\|x-x^*\|) + c_{1g}c_d(1 + c_\lambda)\|x-x^*\| + 2\lambda_\infty^* \psi_{3g}(\|x-x^*\|))$.

We can now choose a sufficiently small neighborhood of x^* such that $\psi(||x - x^*||) \leq \frac{\sigma}{4}$ and subtract the last term of the last relation from the lower bound $\frac{\sigma}{2}||x - x^*||^2$. We take $A = \lambda^T g^-(x) + c_\phi P(x)$, and with this new notation, we get that

$$\frac{\sigma}{4}||x - x^*||^2 \leq A + B||d||||x - x^*||. \tag{63}$$

We treat $||x - x^*||$ as a variable and, by using the formulas for the quadratic equation, we get that

$$||x - x^*|| \leq \frac{2}{\sigma}(B||d|| + \sqrt{B^2||d||^2 + A\sigma}). \tag{64}$$

By using the arithmetic-quadratic mean inequality, we get that

$$||x - x^*||^2 \leq \tfrac{8}{\sigma^2}(2B^2||d||^2 + A\sigma) = \tfrac{16}{\sigma^2}B^2||d||^2 + \tfrac{8}{\sigma}(\lambda^T g^-(x) + c_\phi P(x)) \tag{65}$$

$$\leq \max\{\tfrac{16}{\sigma^2}B^2, \tfrac{8}{\sigma}, \tfrac{8}{\sigma}c_\phi\}(||d||^2 + P(x) + \lambda^T g^-(x)). \tag{66}$$

Choosing

$$c_\sigma = \frac{1}{\max\{\frac{16}{\sigma^2}B^2, \frac{8}{\sigma}, \frac{8}{\sigma}c_\phi\}} \tag{67}$$

we prove the claim.

Corollary 1. *x^* is an isolated stationary point.*

Proof. Let x be another stationary point of the NLP in the neighborhood of x^* where the above theorem holds. Therefore there exists a $\lambda \in \mathcal{M}(x)$ satisfying (3). Hence $(d = 0, \lambda)$ is a solution of (15) and $d = 0$ is the unique solution of the strictly convex QP (14). Since $d = 0$, x is feasible from (14) and $P(x) = 0$ or $g = -g^-(x)$. Now from the complementarity conditions in (15) we get $\lambda^T g^- = -\lambda^T g = 0$. From the previous theorem we get $x = x^*$, which proves the claim.

Corollary 2. *If the second-order sufficient condition (8) is satisfied for one multiplier, and if MFCQ holds at x^*, then x^* is an isolated stationary point.*

Proof. Since x^* is satisfies the quadratic growth condition (1) under these assumptions [6,7] and MFCQ holds, Corollary 1 applies.

3. An Example Without a Locally Convex Augmented Lagrangian

Consider the matrix

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}. \tag{68}$$

Take $u = (\frac{\sqrt{3}}{2}, \frac{1}{2})$. We then have that $u^T Q u = \frac{1}{4}$ and $||u||^2 = 1$. Since the vector $u_0 = (1, 0)$ corresponds to the positive eigenvalue, we have that for any u at an angle of at most $\frac{\pi}{6}$ from u_0, $u^T Q u \geq \frac{1}{4}||u||^2$. Consider now the rotation matrix

$$U_k = \begin{pmatrix} \cos(\frac{k\pi}{4}) & \sin(\frac{k\pi}{4}) \\ -\sin(\frac{k\pi}{4}) & \cos(\frac{k\pi}{4}) \end{pmatrix}. \tag{69}$$

Define $Q_k = U_k^T Q U_k$, for $k = 0, ...3$. We then have $Q_0 + Q_2 = Q_1 + Q_3 = -I_2$, since Q_0 and Q_2 have the same axes of symmetry, but with the eigenvalues switched. Also, for any $u \in \mathsf{R}^2$, there exists a k such that $u^T Q_k u \geq \frac{1}{4}||u||^2$, since the $\frac{\pi}{3}$ wide cones centered at the axis of the positive eigenvalues of Q_k now sweep the entire R^2.

Consider now the optimization problem

$$\min z \quad \text{subject to } z \geq (x\, y) Q_k (x\, y)^T \quad k = 0...3. \tag{70}$$

By the previous observation, we have that $z \geq \frac{1}{4}(x^2 + y^2)$ on the feasible set; thus $z \geq 0$. Clearly, the only solution of the problem is $(0,0,0)$. Since $z \geq \frac{z^2}{4}$, if $0 \leq z \leq 4$, we have that $z \geq \frac{1}{8}(x^2 + y^2 + z^2)$, for all x, y, z feasible and $0 < z < 4$.

Therefore at $x^* = (0,0,0)$ the quadratic growth condition is satisfied for the above NLP, with constant $\frac{1}{8}$. Obviously, MFCQ holds at $(0,0,0)$, and a simple calculation shows that $\sum_{k=0..3} \lambda_k = 1$, for λ_k a multiplier of (70). In particular, at least one multiplier has to be positive. Also, at $(0,0,0)$, all constraints are active and their gradients are $(0,0,-1)$ for any of them. As a result, the linear constraints in (8) now become either $z \geq 0$ or $z = 0$, with at least one being $z = 0$. Therefore the critical cone at x^* is $\mathcal{C} = \{(x,y,z) \in \mathsf{R}^3 | z = 0\}$. Also, from (3), if $\lambda \in \mathcal{M}(x^*)$, then $\sum_{i=1}^4 \lambda_i = 1$.

Assume that there is a choice $\lambda \in \mathcal{M}(x^*)$ such that L_{xx}, the Hessian of the Lagrangian, is positive semidefinite on the critical cone:

$$(x\, y\, z) \begin{pmatrix} \sum_{k=0..3} \lambda_k Q_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \geq 0, \quad \forall (x,y,z), \text{ such that } z = 0. \tag{71}$$

This is equivalent to

$$\sum_{k=0..3} \lambda_k Q_k \succeq 0. \tag{72}$$

Since our construction is invariant to rotations with $\frac{\pi}{4}$ $(U_1^T Q_3 U_1 = Q_0)$, it follows that the positive semi-definiteness holds for any circular permutation σ of this multiplier set:

$$\sum_{k=0..3} \lambda_{\sigma(k)} Q_k \succeq 0. \tag{73}$$

We denote by $\mathcal{A}_c(4)$ the set of circular permutations of four elements. Since the set of positive definite matrices is a convex cone, and

$$\sum_{\sigma \in \mathcal{A}_c(4)} \lambda_{\sigma(k)} = 1, \tag{74}$$

we must have

$$0 \preceq \frac{1}{4} \sum_{\sigma \in \mathcal{A}_c(4)} \sum_{k=0..3} \lambda_{\sigma(k)} Q_k = \frac{1}{4} \sum_{k=0..3} Q_k \sum_{\sigma \in \mathcal{A}_c(4)} \lambda_{\sigma(k)} = \frac{1}{4} \sum Q_k = \frac{-1}{2} I_2, \tag{75}$$

which is impossible. Therefore L_{xx} cannot be positive semidefinite on the critical cone for any choice $\lambda \in \mathcal{M}(x^*)$. Hence the second-order conditions from [6,21] will not hold for any choice of the multipliers.

3.1. Augmented Lagrangian Approaches

Here we discuss the expected behavior of augmented Lagrangian techniques when applied to this example. For these methods, the inequalities of the NLP (2) are converted into equalities [3,5]. The feasible set can be represented as [5]

$$g_i(x) + t_i = 0,\ t_i \geq 0 \text{ for } \ i = 1, \ldots, m.$$

The NLP is replaced by a bound-constrained optimization problem. The equality constraints are incorporated in the objective function based on an estimate λ of the multipliers and a penalty term,

$$\begin{array}{rl} \min & f(x) + \sum_{i=1}^{4} [\lambda_i (g_i(x) + t_i) + \frac{1}{\mu}(g_i(x) + t_i)^2] \\ \text{subject to } t_i \geq 0, & i = 1, \ldots, m. \end{array} \tag{76}$$

Here μ is the barrier parameter. The objective function in (76) is the augmented Lagrangian. The problem is subjected to an additional trust-region constraint [5] to enforce global convergence.

The desired outcome is to have μ bounded away from zero and the trust-region inactive as λ approaches $\mathcal{M}(x^*)$ and the solution of the above problem approaches x^*.

If that happens for our example, then, by a continuity argument following the lower boundedness of μ, $(x^*, t = 0)$ should be a solution of (76) for an appropriate choice of λ, μ. Since (76) has linearly independent gradients of the constraints, both the first and second order necessary conditions must hold [7]. The first order necessary condition results in

$$\nabla f(x) + \lambda \nabla g(x) = 0 \quad \lambda + \nu = 0 \quad \nu \leq 0,$$

where ν, with components $\nu_i \leq 0$, are the multipliers associated with the variables t_i. As a result $\lambda \in \mathcal{M}(x^*)$. The second order necessary conditions require that

$$\nabla_{(x,t)(x,t)} L|_{(x^*,0)} = \begin{pmatrix} F_{xx} + \sum_{i=1}^{4} (\lambda_i G_{xx} + \frac{2}{\mu} \nabla g_i(x^*) \nabla g_i(x^*)^T) & \frac{2}{\mu} \nabla g(x^*) \\ \frac{2}{\mu} \nabla g(x^*)^T & \frac{2}{\mu} I_4 \end{pmatrix}$$

be positive semidefinite, at least on the subspace of $(\delta x, \delta t)$ with $\delta t = 0$. This results in

$$0 \preceq F_{xx} + \sum_{i=1}^{4} (\lambda_i G_{xx} + \frac{2}{\mu} \nabla g_i(x^*) \nabla g_i(x^*)^T) = \begin{pmatrix} \sum_{i=1}^{4} \lambda_i Q_i & 0 \\ 0 & \frac{8}{\mu} \end{pmatrix} \tag{77}$$

or

$$0 \preceq \begin{pmatrix} \sum_{i=1}^{4} \lambda_i Q_i & 0 \\ 0 & \frac{8}{\mu}. \end{pmatrix} \tag{78}$$

We proved that the last matrix cannot be positive semidefinite for our example and we thus get a contradiction. This shows that, either the trust region will be active arbitrarily close to x^*, or $\mu \to 0$.

This also shows that the Hessian of the augmented Lagrangian of the equality constrained problem

$$F_{xx} + \sum_{i=1}^{4} (\lambda_i G_{xx} + \frac{2}{\mu} \nabla g_i(x^*) \nabla g_i(x^*)^T)$$

is not positive semidefinite and thus the augmented Lagrangian of the equality constrained problem cannot be locally convex.

4. Linear Convergence of the SQP with Nondifferentiable Exact Penalty $P(x)$

The points x considered in thus subsection are assumed to be sufficiently close to x^*. The notation d and $\lambda \in \mathcal{M}(x)$ will refer to the solutions of (14) and (15). Also, $P(x)$ is the L_∞ penalty function (10) and $\phi(x) = f(x) + c_\phi P(x)$.

4.1. Proof of the Technical Results

Lemma 5.

$$P(x + \alpha d) \leq (1 - \alpha) P(x) + c_{2g} \alpha^2 ||d||^2, \quad \forall \alpha \in [0, 1].$$

Proof. Since d is a feasible point of (14), we have that $\nabla g_i(x)^T d \leq -g_i(x), \forall i \in \{1, \ldots, m\}$. By Taylor's remainder theorem

$$g_i(x + \alpha d) \leq (1 - \alpha) g_i(x) + c_{2g} \alpha^2 ||d||^2, \forall \alpha \in [0, 1], \forall i = 1, \ldots, m.$$

Hence

$$\begin{aligned} \max_{1 \leq i \leq m} \{g_i(x + \alpha d)\} &\leq (1 - \alpha) \max_{1 \leq i \leq m} \{g_i(x)\} + c_{2g} \alpha^2 ||d||^2 \\ &\leq (1 - \alpha) P(x) + c_{2g} \alpha^2 ||d||^2, \quad \forall \alpha \in [0, 1]. \end{aligned}$$

This completes the proof.

Lemma 6. *There exist $\bar{\alpha}$, $0 < \bar{\alpha} \leq 1$, and $c_2 > 0$ such that, for some $(\lambda) \in \mathcal{M}(x)$*

$$\begin{gathered} \phi(x + \alpha d) - \phi(x) \leq -\alpha \tfrac{1}{2} ((d)^T d + \gamma P(x) + \lambda^T g^-(x)) \leq \\ -c_2 \alpha (||d||^2 + P(x) + \lambda^T g^-(x)), \quad \forall \alpha \in [0, \bar{\alpha}]. \end{gathered}$$

Proof. Writing the KKT conditions for (14), we obtain

$$d + \nabla f(x) + \sum_{i=1,m} \lambda_i \nabla g_i(x) = 0$$

and, hence,

$$\begin{gathered} (d)^T d + \nabla f(x)^T d + \textstyle\sum_{i=1,m} \lambda_i \nabla g_i(x)^T d = 0 \\ (d)^T d + \nabla f(x)^T d - \textstyle\sum_{i=1,m} \lambda_i g_i(x) = 0, \end{gathered}$$

since, by the complementarity conditions satisfied by the solution of (14), $\lambda^T \nabla g(x)^T d = -\lambda^T g(x), \forall i = 1, m$. Therefore, since $g_i(x) = g_i^+(x) - g_i^-(x)$,

$$\begin{aligned} \nabla f(x)^T d = -(d)^T d + \textstyle\sum_{i=1,m} \lambda_i (g_i^+(x) - g_i^-(x)) \leq \\ -(d)^T d + P(x)(\textstyle\sum_{i=1,m} \lambda_i) - \lambda^T g^-(x) \leq \\ -(d)^T d + (c_\phi - \gamma) P(x) - \lambda^T g^-(x) \end{aligned} \tag{79}$$

by (10), (17). By Taylor's remainder theorem,

$$f(x + \alpha d) \leq f(x) + \alpha \nabla f(x)^T d + c_{2f} \alpha^2 ||d||^2.$$

Hence, for $\alpha \in [0, 1]$,

$$\begin{aligned} f(x + \alpha d) + c_\phi P(x + \alpha d) \leq f(x) + \alpha \nabla f(x)^T d + c_{2f} \alpha^2 ||d||^2 + \\ (1 - \alpha) c_\phi P(x) + c_\phi c_{2g} \alpha^2 ||d||^2 \leq f(x) + (1 - \alpha) c_\phi P(x) + \\ \alpha(-(d)^T d + (c_\phi - \gamma) P(x) - \lambda^T g^-(x)) + (c_\phi c_{2g} + c_{2f}) \alpha^2 ||d||^2 = \\ f(x) + c_\phi P(x) - \alpha((d)^T d + \gamma P(x) + \lambda^T g^-(x)) + (c_\phi c_{2g} + c_{2f}) \alpha^2 ||d||^2 \end{aligned}$$

from (79) and Lemma 5. Therefore, for $\alpha \in [0, 1]$,

$$\phi(x + \alpha d) - \phi(x) \leq -\alpha((d)^T d + \gamma P(x) + \lambda^T g^-(x)) + (c_\phi c_{2g} + c_{2f}) \alpha^2 ||d||^2.$$

The result of the statement follows by choosing $\bar{\alpha} = \min\{1, \frac{1}{2(c_\phi c_{2g} + c_{2f})}\}$ and $c_2 = \frac{1}{2} \min\{\gamma, 1\}$.

Lemma 7. *There exists a constant* c_5 *such that,* $\forall (\lambda) \in \mathcal{M}(x)$,

$$\phi(x) - \phi(x^*) \leq c_5 (P(x) + ||x - x^*||^2 + \lambda^T g^-(x) + ||d||^2).$$

Proof. From (15) and the definition of the Lagrangian (4) it follows, using Taylor's theorem, that, for a sufficiently small neighborhood of x,

$$\mathcal{L}(x, \lambda^*) - \mathcal{L}(x^*, \lambda^*) \leq \Sigma ||x - x^*||^2 \quad \forall (\lambda^*) \in \mathcal{M}(x^*),$$

where $\Sigma = \max\{c_{2f}, c_\phi c_{2g}\}$. Also, by (16), we can choose $\lambda^* \in \mathcal{M}(x^*)$ such that

$$|g(x)^T (\lambda^* - \lambda)| \leq c_{1g} c_d ||x - x^*||^2.$$

Since $\mathcal{L}(x^*, \lambda^*) = f(x^*)$, we have that

$$f(x) - f(x^*) - \Sigma ||x - x^*||^2 + (\lambda)^T g(x) \leq \tag{80}$$

$$(\lambda - \lambda^*)^T g(x) \leq c_{1g} c_d (||x - x^*||^2) \tag{81}$$

and, thus

$$f(x) - f(x^*) \leq (\Sigma + c_{1g} c_d) ||x - x^*||^2 - (\lambda)^T g(x) \tag{82}$$

$$\leq (\Sigma + c_{1g} c_d) ||x - x^*||^2 + \lambda^T g^-(x). \tag{83}$$

Therefore

$$f(x) + c_\phi P(x) - f(x^*) \leq (\Sigma + c_{1g} c_d) ||x - x^*||^2 + c_\phi P(x) + \lambda^T g^-(x).$$

The conclusion of the lemma follows by choosing $c_5 = \max\{\Sigma + c_{1g} c_d, c_\phi, 1\}$.

4.2. Nondifferentiable Exact Penalty Algorithms and the Linear Convergence Theorem

The linearization algorithm [3, p.372] has the following form:

1. Set $k = 0$, choose x^0.
2. Compute d^k from (11).
3. Choose α^k from a line search procedure, and set $x^{(k+1)} = x^k + \alpha^k d^k$.
4. Set $k = k + 1$ and return to Step 2.

The stepsize α^k is chosen by one of the following procedures [3, p.372].

(a) *Minimization rule* Here α^k is chosen such that

$$\phi(x^k + \alpha^k d^k) = \min_{\alpha \geq 0}\{\phi(x^k + \alpha d^k).\}$$

(b) *Limited minimization rule* Here a fixed scalar $s > 0$ is selected, and α^k is chosen such that

$$\phi(x^k + \alpha^k d^k) = \min_{\alpha \in [0,\, s]}\{\phi(x^k + \alpha d^k)\}.$$

(c) *Armijo rule* Here fixed scalars s, τ, and σ with $s > 0$, $\tau \in (0,1)$, and $\sigma \in (0, \frac{1}{2})$ are chosen and we set $\alpha^k = \tau^{m_k} s$, where m_k is the first nonnegative integer m for which

$$\phi(x^k) - \phi(x^k + \tau^m s d^k) \geq \sigma \tau^m s (d^k)^T d^k.$$

It can be shown that the Armijo rule yields a stepsize after a finite number of iterations.

The following theorem establishes the convergence properties of the linearization algorithm. The global convergence properties, established in [2, Prop. 4.3.3], are also stated here for completeness.

Theorem 2. *Let x^k be a sequence generated by the linearization algorithm, where the stepsize α^k is chosen by the minimization rule, limited minimization rule or the Armijo rule. Then any accumulation point of the sequence x^k is a stationary point of $\phi(x) = f(x) + c_\phi P(x)$. If $x^k \to x^*$, where x^* is a strict local minimum of the problem (2) satisfying the local quadratic growth (1) and the Mangasarian-Fromowitz constraint qualification (5), then $\phi(x^k) \to \phi(x^*)$ Q-linearly and $x^k \to x^*$ R-linearly.*

Proof. The first part is an immediate consequence of [2, Prop. 4.3.3]. We prove the linear convergence statement only for the Armijo rule, the proof being similar for the other stepsize selection mechanisms. By Lemma 6

$$\begin{aligned} \phi(x^k) - \phi(x^k + \alpha d^k) &\geq \alpha \tfrac{1}{2}((d^k)^T d^k + \tfrac{\gamma}{2} P(x^k) + (\lambda^k)^T g^-(x^k) \\ &\geq \alpha \tfrac{1}{2}(d^k)^T d^k > \sigma \alpha (d^k)^T d^k \end{aligned} \tag{84}$$

for all $\alpha \in [0, \bar{\alpha}]$. Since m_k is the smallest integer m for which

$$\phi(x^k) - \phi(x^k + \tau^m s d^k) \geq \sigma \tau^m s (d^k)^T d^k,$$

it follows that $\tau^m s \geq \tau\bar{\alpha}$. This therefore ensures that the stepsize is at least $\tau\bar{\alpha}$ for k sufficiently large. As a result of Lemma 6, we have that

$$\phi(x^k) - \phi(x^{(k+1)}) \geq c_2\tau\bar{\alpha}(||d^k||^2 + P(x^k) + (\lambda^k)^T g^-(x^k)). \tag{85}$$

On the other hand, by Lemma 7 we have that

$$\phi(x^k) - \phi(x^*) \leq c_5(P(x^k) + ||x^k - x^*||^2 + ||d^k||^2 + (\lambda^k)^T g^-(x^k)).$$

By Theorem (1) and the previous relation it follows that there exists $c_6 = c_5(1 + c_\sigma)$ such that

$$\phi(x^k) - \phi(x^*) \leq c_6((\lambda^k)^T g^-(x^k) + P(x^k) + ||d^k||^2) \leq \tag{86}$$

$$\tfrac{c_6}{\tau\bar{\alpha}c_2}(\phi(x^k) - \phi(x^{k+1})) = \delta(\phi(x^k) - \phi(x^{k+1})) = \tag{87}$$

$$\delta(\phi(x^k) - \phi(x^*)) - \delta(\phi(x^{(k+1)}) - \phi(x^*)), \tag{88}$$

by using Lemma 6 and where $\delta = \frac{c_6}{\tau\bar{\alpha}c_2}$. After some obvious manipulation, it follows that

$$\delta(\phi(x^{(k+1)}) - \phi(x^*)) \leq (\delta - 1)(\phi(x^k) - \phi(x^*)),$$

which proves the Q-linear convergence [19] of the sequence $\phi(x^k)$ to $\phi(x^*)$ with a linear rate of at most $\frac{\delta-1}{\delta}$. Therefore

$$\limsup_{k\to\infty} \sqrt[k]{\phi(x^k) - \phi(x^*)} \leq \frac{\delta - 1}{\delta}.$$

From Lemma 2

$$\phi(x^k) - \phi(x^*) \geq \beta||x^k - x^*||^2.$$

Therefore

$$\limsup_{k\to\infty} \sqrt[k]{||x^k - x^*||} \leq (\frac{\delta - 1}{\delta})^{\frac{1}{2}},$$

which proves the R-linear convergence [19] to 0 of the sequence $x^k - x^*$. The proof is complete.

Following the techniques from [1], we can extend the result for the case where the matrix H of the QP is not I but changes from iteration to iteration. The only condition is that the sequence of strictly convex H^k be uniformly upper and lower bounded.

Iteration	$\frac{\phi(x^k)-\phi(x^*)}{\phi(x^{k+1})-\phi(x^*)}$
4	4.00
9	4.00
14	3.99
19	3.99
24	4.00
27	4.00

Table 1. Rates of convergence for the L_∞ penalty algorithm

Iteration	(New) Penalty Parameter	Trust Region Radius $\|\|\|\|_\infty$
16	1e-2	3.81 e-02
43	1e-4	1.1 e-02
85	1e-6	1.35 e-03
141	1e-8	4.22 e-05
203	1e-10	5.28 e-06
241	1e-12	1.70 e-06
268	1e-14	1.93
283	1e-16	4.41 e02
323	1e-18	2.19 e04
336	STOP	

Table 2. Reduction of the penalty parameter μ for LANCELOT

5. Numerical Experiments with Degenerate NLP

We experimented with several nonlinear programming packages on the example from Section 3. Certainly, comparing the behavior of NLP algorithms on a unique degenerate example cannot result in a complete characterization. Nevertheless, it may be of interest to determine whether methods using augmented Lagrangians will really encounter problems when solving an example without a positive semidefinite augmented Lagrangian. We also desire to validate the theoretical conclusions of the preceding sections.

We have shifted the origin for our example, to avoid one step convergence of algorithms that start at $0, 0, 0$ by default. The algebraic form of the example is

$$\begin{array}{rll} & \min z & \\ \text{sbj.to:} & g_0(x,y,z) = (x-1)^2 - 2(y-1)^2 - z & \leq 0 \\ & g_1(x,y,z) = -\frac{1}{2}((x-1)^2 + (y-1)^2) + 3(x-1)(y-1) - z & \leq 0 \\ & g_2(x,y,z) = -2(x-1)^2 + (y-1)^2 - z & \leq 0 \\ & g_3(x,y,z) = -\frac{1}{2}((x-1)^2 + (y-1)^2) - 3(x-1)(y-1) - z & \leq 0. \end{array} \tag{89}$$

From our analysis, we have that $w^* = (1,1,0)$ is a minimum satisfying the quadratic growth condition (1) with $z - 0 \geq \frac{1}{8}((x-1)^2 + (y-1)^2 + z^2)$ for feasible (x, y, z) near w^*. The feasible set is described in Figure 5. In the lateral view, the quadratic growth at $(1,1,0)$ is fairly obvious from the curvature of the ridges that appear at the intersection of two constraints. From the shape of the feasible set it is also clear that $(1,1,0)$ is the unique stationary point of the NLP.

Among the solvers we used, MINOS [17] and SNOPT [11] use quasi-Newton methods that do not require second-order derivatives of the constraints. They

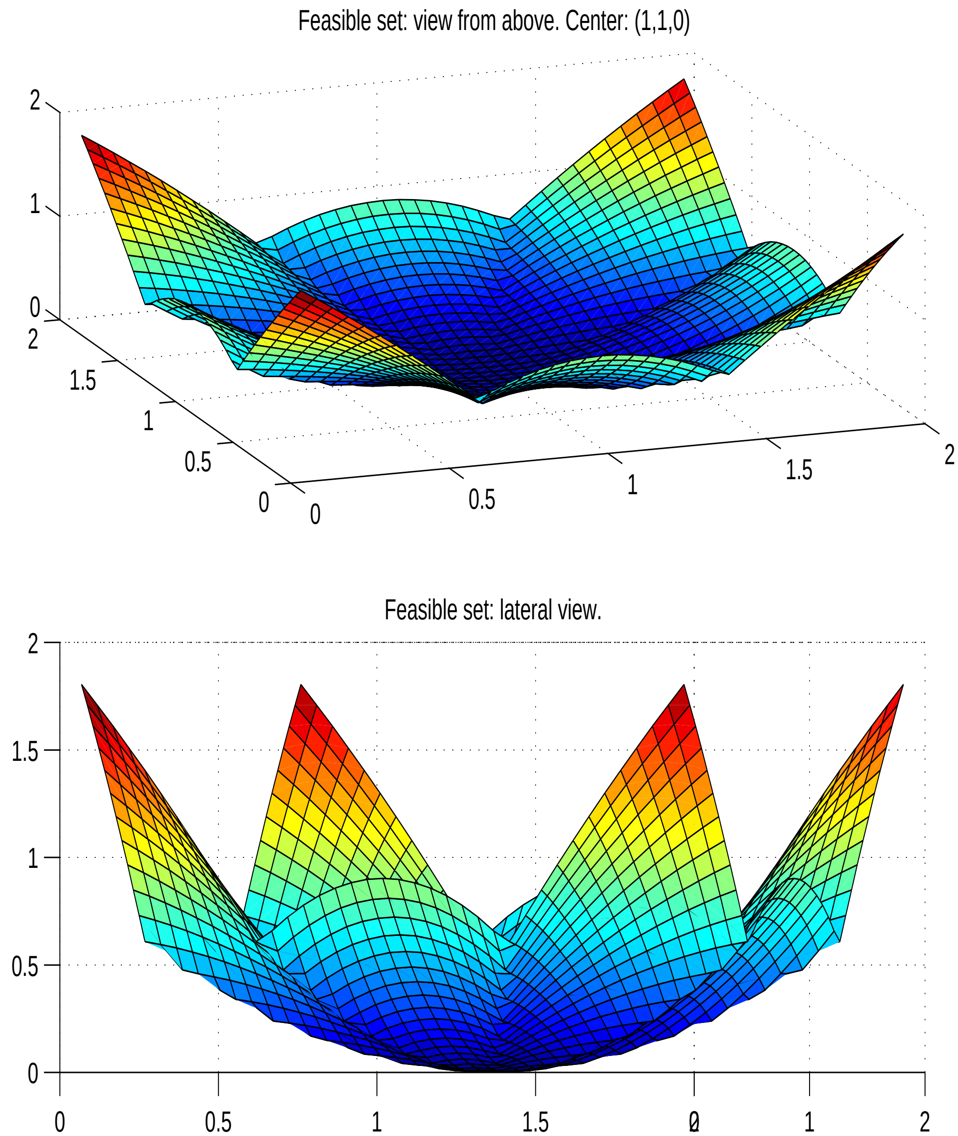

Fig. 1. Feasible set of the nonlinear program (89) (1,1,0) is the local minimum satisfying the quadratic growth condition (1). The jagged edges in the lateral view are a meshing effect.

also use an augmented Lagrangian as a merit function. DONLP2 [23] solves a linear system instead of a Quadratic Program at each iteration and uses an L_1 penalty function. LANCELOT [5] uses an augmented Lagrangian technique in conjunction with a trust-region. FilterSQP [8] also uses a trust region approach but with a special classification of the relative merits of the iterates instead of

Nonlinear solver	$\|\|x^{final} - x^*\|\|_2$	Iterations	Message at termination
DONLP2	1.45e-16	4	Success
FilterSQP	5.26e-09	28	Convergence
LANCELOT	8.65e-07	336	Step size too small
LINF	1.05e-08	28	Step size too small
LOQO	1.60e-07	200	Iteration limit
LOQO	5.50e-07	1000	Iteration limit
MINOS	4.76e-06	27	Current point cannot be improved
SNOPT	3.37e-07	3	Optimal Solution Found

Table 3. Runs with various nonlinear solvers on the problem (89)

a penalty or merit function. LOQO [24] is an interior-point approach. Finally, LINF is an ad hoc Matlab implementation of the L_∞ exact penalty function described in the preceding section, with an Armijo rule. The latter algorithm is started at $(0,0,0)$. All runs, except for the L_∞ penalty and FilterSQP algorithms, were done on the NEOS server [18], where additional documentation can be found for all of the above solvers.

For such a small example the time of execution is not relevant in comparing the behavior of the solvers. Since the solution of the problem is known, we chose as a criteria for comparison the best achievable solution. We set all relevant tolerances to $1e-16$, via the AMPL interface of NEOS. Smaller tolerances may interfere with the machine precision, though most of the solvers gave comparable answers even when the tolerances are set to $1e-20$. Larger tolerances ($1e-12$–$1e-15$) again resulted in very similar results. Whenever allowed, we also changed other limiting parameters until an intrinsic stopping decision was issued. The only exception was DONLP2 which converged to all digits in the mantissa with the default settings.

Table 1 shows the ratios $\frac{\phi(x^k)-\phi(x^*)}{\phi(x^{k+1})-\phi(x^*)}$ at various iterations for our implementation LINF. All are close to 4.00, which is consistent with the Q-linear convergence claim for $\phi(x)$.

Table 2 shows that LANCELOT decreases succesively the value of the penalty parameter (by 16 orders of magnitude), until it stops with the message 'Step size too small'. This was indeed one of the alternatives allowed by our analysis in Subsection 3.1 ($\mu \to 0$). This is an undesirable outcome since the subproblems (76) may become harder to solve.

The results for all runs are illustrated in Table 3. It can be seen that the solvers that use augmented Lagrangians MINOS, SNOPT, LANCELOT exhibit an error of at least one order of magnitude larger compared to all other algorithms. However, one would expect that SNOPT and MINOS would have had at least as good a behavior as LINF if they would use a different merit function, since the nature of the QP solved is very similar to (14). Increasing the iteration limit in LOQO did not result in a better outcome. It is interesting to note that the outcome in FilterSQP and LINF differ by only a factor of 2 in the same number of iterations, though FilterSQP uses second-order information whereas LINF does not. Both LINF and FilterSQP solve quadratic programs at each iteration. DONLP2 has a remarkable behavior, though further investigation is necessary to determine whether this has some general implications.

It is impossible to draw a general conclusion from one example. However, there seems to be an adverse bias for methods using augmented Lagrangians on degenerate NLPs as the one above. We are not advocating the use of LINF on general NLP, since its similarity to steepest descent makes it very sensitive to ill-conditioning. But the fact that it gives an outcome comparable to the one of solvers using second-order information shows that, for better results, a different way of incorporating second-order derivatives may be necessary.

6. Conclusions

In this work we analyze the behavior of nonlinear programs in presence of constraint degeneracy: linear dependence of the gradients of the active constrains. The problems of interest exhibit minima with a quadratic growth property that satisfy the Mangasarian-Fromowitz constraint qualification. The novelty of our approach is that, while studying the SQP convergence properties, we do not assume the positive semidefiniteness of the Hessian of the Lagrangian on the critical cone for any of the feasible Lagrange multipliers. Our conditions are equivalent to a weak second-order sufficient condition [15,22].

We prove that, under these assumptions, if the data of the problem are twice continuously differentiable, the target minimum will be an isolated stationary point of the NLP. We also show that, when started sufficiently close to the minimum, the L_∞ exact penalty SQPs induce Q-linear convergence of the values of the penalized objective $\phi(x) = f(x) + c_\phi P(x)$ and R-linear convergence of the iterates. This shows that such methods are robust with respect to constraint degeneracy.

We give an example of a nonlinear program with a unique minimum that satisfies our conditions for which the Hessian of the Lagrangian is not positive semidefinite on the critical cone for any feasible choice of the multipliers. The direct consequence of this fact is that there is no augmented Lagrangian that will be positive semidefinite at the solution. Therefore, Lagrange multipliers algorithms will have to drive the penalty parameter to zero for such examples unless the trust region is active even at convergence.

We provide our computational experience with this small nonlinear program. As a criteria for comparison we used the best achievable solution, which was obtained after tuning the parameters of the algorithms. We observed that, for this example, algorithms that use augmented Lagrangians resulted in errors of one order of magnitude or larger when compared to the other approaches. The Lagrange multiplier package that we used (LANCELOT [5]), was confined to decrease substantially the value of the penalty parameter (16 orders of magnitude), which is one of the outcomes allowed by our analysis. The linear convergence results concerning the L_∞ penalty function were also validated by our experiments.

Undoubtedly, such a small experiment is insufficient to draw any conclusions, especially about the approaches for which we have no theory under these assumptions, such as interior-point algorithms. However, both from our theory and our experiments, it does appear that methods that use augmented Lagrangians are less robust with respect to constraint degeneracy when compared to SQP.

We believe that attempting to develop a convergence theory in absence of the usual second-order conditions is interesting because it may result in algorithms that are more robust by virtue of the fact that their properties depend on fewer assumptions. However, how to improve on the current results, and especially how to define reliable variants of the Newton method (if possible) for this case, is a subject of future research.

Acknowledgments

Thanks to Stephen Wright, Jorge Moré and Danny Ralph for the many discussions on the subject. David Gay, Sven Leyffer and Chi-Jen Lin have kindly provided me information and support for the numerical examples.

References

1. M. Anitescu, "On the rate of convergence of sequential quadratic programming with nondifferentiable exact penalty function in the presence of constraint degeneracy", Preprint ANL/MCS-P760-0699, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill, 1999.
2. D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
3. D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusets, 1995.
4. J. F. Bonnans and A. Ioffe, "Second-order sufficiency and quadratic growth for nonisolated minima", *Mathematics of Operations Research*, **20**:4 (1995), 801–819.
5. A. R. Conn, N. I. M. Gould and Ph. L. Toint, *LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization*, Springer Verlag, Berlin, 1992.
6. A. V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York, 1983.
7. R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Chichester, 1987.
8. R. Fletcher and S. Leyffer, "Nonlinear programming without a penalty function", Numerical Analysis Report, NA/171, Department of Mathematics, University of Dundee, UK.
9. J. Gauvin, "A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming", *Mathematical Programming* 12 (1977), 136–138.
10. J. Gauvin and J. W. Tolle, "Differential stability in nonlinear programming", *SIAM Journal of Control and Optimization*, **15**(1977), 294-311.
11. 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.
12. J.L. Goffin, "The relaxation method for solving systems of linear inequalities", *Mathematics of Operation Research*, **5**:3 (1980), 388-414.
13. W. W. Hager, "Stabilized sequential quadratic programming", To appear in *Computational Optimization and Applications.*
14. 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.
15. 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), 266-288.

16. A. Ioffe, "On Sensitivity analysis of nonlinear programs in Banach spaces: the approach via composite unconstrained optimization", *Siam Journal of Optimization*, **4**:1 (1994), 1-43.
17. B. Murtagh and M. A. Saunders, "MINOS 5.0 user's guide", Technical Report SOL 83-20, Stanford University, Stanford, 1983.
18. *The NEOS Guide*. Available Online at *http://www.mcs.anl.gov/otc/Guide.*
19. J. Ortega and W. Rheinboldt, *Iterative Solutions of Nonlinear Equations in Several Variables*, Academic Press, New York, 1972.
20. 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.
21. S. M. Robinson, "Generalized equations and their solutions, Part II: Applications to nonlinear programming" *Mathematical Programming Study* 19 (1980), 200–221.
22. A. Shapiro, "Sensitivity analysis of nonlinear programs and differentiability properties of metric projections", *SIAM Journal of Control and Optimization*, **26**:3 (1988), 628–645.
23. P. Spelucci, "An SQP method for general nonlinear programs using only equality constrained subproblems". Available online from http://www.mathematik.tu-darmstadt.de/ags/ag8/spellucci/.
24. Vanderbei, R. J., LOQO: "An interior-point code for quadratic programming", Technical Report SOR-94-21, Statistics and Operations Research, Princeton University, Princeton, 1994.
25. S. J. Wright, "Superlinear convergence of a stabilized SQP method to a degenerate solution", Preprint ANL/MCS-P463-0297, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill, 1997.

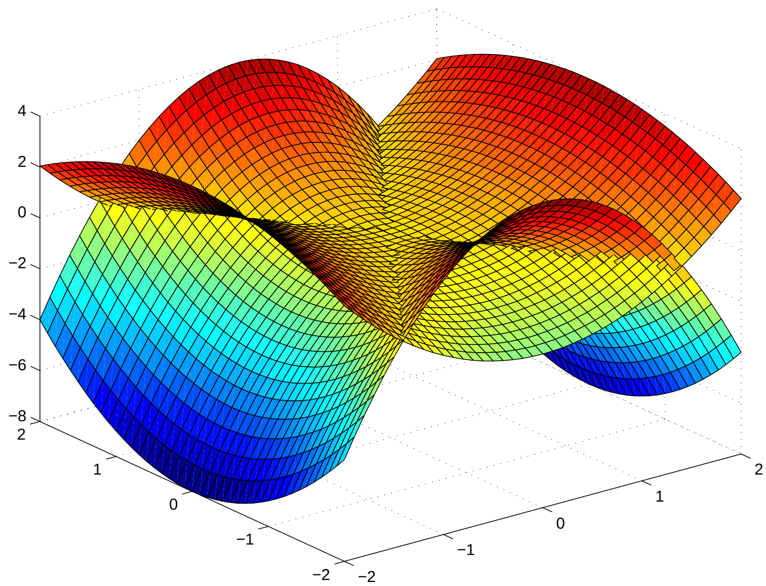

Fig. 2. Feasible set of (90) $c = 0.49$: Top of the image

A. Other runs with degenerate LCP

This section will not appear in the journal submission.

A.1. Limited Validity of Second Order Conditions

The example used here is

$$\begin{array}{lll} & \min z & \\ \text{subject to:} & g_1(x,y,z) = (x-1)^2 - 2(y-1)^2 - z & \leq 0 \\ & g_2(x,y,z) = -c(x-1)^2 + (y-1)^2 - z & \leq 0 \\ & g_3(x,y,z) = -(x-1)^2 - (y-1)^2 - z & \leq 0. \end{array} \tag{90}$$

The last constraint is redundant. If $c < 0.5$, the problem has a unique solution and a unique stationary point, $(1,1,0)$ in the feasible region. Also, at the optimal point there is a specific choice of multipliers that makes the Hessian matrix positive definite on the null space of the gradients of the active constraints. This choice is $\frac{1}{1+\gamma}(1,\gamma,0)$ with γ of appropriate value ($\gamma \approx 2$). What is particular about this problem is that the size of the multipliers that satisfy the second-order sufficient conditions is small compared with the size of the multiplier set (if $c = 0.5$, no multiplier satisfies the second-order sufficient conditions).

The difficulty of the problem resides in the geometry of the feasible set. This consists of two linear (in projection on the (x,y) plane) ridges with steep walls, but with slow descent, as shown in Figure A.1. The slope of the descent along this ridges is controlled by c. If $c = 0.5$, that slope becomes 0. Thus the danger is that an optimization algorithm will get stuck in the ridge and, unless the ratio of the constrained gradient to the curvature is evaluated correctly, a stop-optimality decision may be issued by the NLP solver before reaching the true minimum. Note that along the ridge the objective function is convex quadratic; thus, one Newton step (of the equality-constrained problem defining the ridge) would get any point on the ridge to $(1, 1, 0)$, the true solution.

The problem becomes increasingly ill-conditioned as $c \to 0.5$. However, to demonstrate that the difficulty of the problem lies with the description of the feasible primal-dual set (as done by most NLP solvers) and not with the size of the curvature, we construct what we call the convexified problem. Convexification involves replacing all constraints with a single constraint generated by adding all constraints with weights that are proportional to a set of multipliers satisfying the second-order sufficient condition $((1, \gamma, 0))$:

$$\begin{array}{rc} & \min z \\ \text{subject to:} & g_1(x, y, z) + \gamma g_2(x, y, z) \leq 0. \end{array} \tag{91}$$

The performance of the algorithm on this problem is arguably a good estimate of the performance of the method when a set of multipliers that satisfies the second-order sufficient conditions is chosen, since the curvature of the convexified problem is now close to the one in the ridges. The results of the algorithms on this problem are reported in the result tables under "Convexified".

The problem was solved by the same nonlinear programming packages. The results are reported in the following tables, indexed by the value of c, for all solvers except the $L\infty$ penalty algorithm. Although convergence to the correct solution was observed for $c = 0.49$ and $c = 0.499$, the number of iterations to reach a precision of $1e - 4$ was in the thousands for this algorithm, because of the absence of scaling with second-order derivatives. Although the theoretical results are confirmed, this algorithm would fail if the stop criteria from the other solvers were to be applied.

All other algorithms reported success (" optimal solution found "). This problem seems to have the biggest impact on SNOPT. While the effect described above is at a moderate level, $c = 0.49$, SNOPT gives almost 2% relative error, while on the convexified problem the solution is exact to almost all digits reported. At $c = 0.4999$, the error is almost 14%, while on the convexified problem, again, the result is correct to all reported digits. Similar effects are seen with MINOS and, to a smaller extent, LANCELOT and DONLP2. For MINOS and SNOPT the result is expected to a certain extent because none of them uses second-order derivative information, which is important if the algorithm is to advance correctly from one of the ridges.

At the time of this experiment, it was not possible to change the tolerance from within the NEOS server. This is another reason why it is difficult to judge the relative performance of the solvers. However, we note that LOQO and FilterSQP proved more robust for the following reason. For a given ideal algorithm the expectation is that maintaining the tolerance level but worsening

Solver Type	x1	x2	x3	ITERATIONS
True Solution	1.0	1.0	0.0	
DonLP2	.999844	1.000486	.4177973E-21	8
LANCELOT	1.00001	1.00002	-6.61704e-07	48
LOQO	0.999302	0.999505	2.30844e-15	17
MINOS	0.997436	0.998193	1.34441e-17	17
SNOPT	0.987785	0.99139	1.47451e-17	6
FilterSQP	0.999999	0.999999	0	27
Convexified with $\gamma = 2.01$				
MINOS	1	0.999994	1.32168e-13	
SNOPT	1.00001	1	-2.57968e-07	

Table 4. Example (90) with $c = 0.49$

Solver Type	x1	x2	x3	Iterations
True Solution	1.0	1.0	0.0	
DONLP2	.999999	.999999	.677626E-20	8
LANCELOT	0.998335	0.998838	9.59784e-09	98
LOQO	0.999201	0.999435	1.10422e-15	26
MINOS	0.994454	0.99608	-4.10642e-18	19
SNOPT	0.96652	0.976334	-1.2902e-17	6
FilterSQP	0.999999	0.999999	0	27
Convexified with $\gamma = 2.001$				
MINOS	1.00002	1.00001	-4.55105e-10	
SNOPT	1.00001	1	-2.36249e-08	

Table 5. Example (90) with $c = 0.499$

Solver Type	x1	x2	x3	Iterations
True Solution	1.0	1.0	0.0	
DONLP2	.999779	.999804	-.361395E-10	8
LANCELOT	0.999692	0.999781	6.21676e-06	203
LOQO	0.99938	0.999562	7.45876e-17	32
MINOS	0.984589	0.989103	9.93053e-14	24
SNOPT	0.861688	0.902202	6.41441e-17	6
FilterSQP	0.999999	0.999999	0	27
Convexified with $\gamma = 2.0001$				
MINOS	0.999529	0.999531	1.84292e-11	
SNOPT	1	1	-8.47945e-08	

Table 6. Example (90) with $c = 0.4999$

the conditioning of the problem will result in more iterations, but comparable accuracy (at least in the initial stages, when the effect of errors in the data is not visible). However, SNOPT and DONLP2, and to a lesser extent MINOS and LANCELOT, had a small variation in the number of iterations, but large variations in the accuracy of the outcome. On the other hand, FilterSQP and LOQO had a substantial increase in the number of iterations, but produced iterates of similar quality. Because of the complexity of the packages, it is hard to pinpoint the reason for this behavior; however, it is probably related to the fact that DONLP2, MINOS, LANCELOT, and SNOPT use only approximations of the second-order derivatives of the constraints.

Solver Type	x1	x2	x3	Iterations
True Solution	1.0	1.0	0.0	
DONLP2	-0.115711	0.211076	8.29874e-06	4
LANCELOT	0.4657	0.622194	3.04601e-06	183
LOQO	0.999309	0.999512	1.05234e-16	92
MINOS	0.865743	0.905066	1.05126e-07	22
SNOPT	0.723417	0.804427	6.64565e-17	7
FilterSQP	0.999999	0.999999	0	44
Convexified with $\gamma = 2.00001$				
MINOS	1.00765	1.00556	3.95604e-10	
SNOPT	1.00424	0.993557	-5.74026e-09	

Table 7. Example (90) with $c = 0.49999$

Solver Type	x1	x2	x3	Iterations
True Solution	1.0	1.0	0.0	
DONLP2	-0.115724	0.211064	8.29893e-07	4
LANCELOT	-0.128934	0.201723	1.47398e-06	40
LOQO	0.999296	0.999502	7.16259e-17	157
MINOS	-0.1992	0.152038	9.58652e-07	7
SNOPT	-0.110258	0.214929	8.21782e-07	6
FilterSQP	0.999999	0.999999	0	44
Convexified with $\gamma = 2.000001$				
LOQO1	1.00052	1.0011	1.66011e-08	
MINOS	0.915435	0	3.36909e-07	
SNOPT	1e-06	0	8.33332e-07	

Table 8. Example (90) with $c = 0.499999$

A.2. Discussion

Figures 5 and A.1 show that the feasible set is contained in a strictly convex rotation paraboloid and, at least for the first example, is reasonably well behaved ($f(x)-f(x^*) \geq \frac{1}{8}||x-x^*||^2$ for x feasible in a neighborhood of x^*. Therefore both examples have a simple, strictly convex relaxation that has the same solution point as the original one. If that constraint were added to the problem, in the case of the first example the second-order sufficient condition would hold for at least one choice of multipliers.

The approach that seems to work robustly for both examples is the one provided by LOQO. However, given their initial scope, MINOS and SNOPT have also shown fairly robust behavior, since, as opposed to interior-point approaches, they are designed to work for relatively high tolerances. DONLP2, FilterSQP and LANCELOT drift far away of the region of interest for (89), although FilterSQP has proven very stable to the ill-conditioning of the second problem. The availability of second-order derivative information has resulted in substantially better stability of the end results. Finally, the L_∞ algorithm analyzed in this paper has confirmed the theoretical results, although it cannot compete for ill-conditioned problems, as expected from its similarity to the steepest descent algorithm of unconstrained optimization. The results of some algorithms on the convexified problem seem to indicate that there would be a real computational benefit in identifying the multipliers for which second-order sufficient conditions hold.