Computational Optimization and Applications (2023) 85:973–992
https://doi.org/10.1007/s10589-023-00476-1

Convergence rate estimates for penalty methods revisited

A. F. Izmailov[1] **· M. V. Solodov**[2]

Received: 25 September 2022 / Accepted: 11 March 2023 / Published online: 28 March 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
For the classical quadratic penalty, it is known that the distance from the solution of the penalty subproblem to the solution of the original problem is at worst inversely proportional to the value of the penalty parameter under the linear independence constraint qualification, strict complementarity, and the second-order sufficient optimality conditions. Moreover, using solutions of the penalty subproblem, one can obtain certain useful Lagrange multipliers estimates whose distance to the optimal ones is also at least inversely proportional to the value of the parameter. We show that the same properties hold more generally, namely, under the (weaker) strict Mangasarian–Fromovitz constraint qualification and second-order sufficiency (and without strict complementarity). Moreover, under the linear independence constraint qualification and strong second-order sufficiency (also without strict complementarity), we demonstrate local uniqueness and Lipschitz continuity of stationary points of penalty subproblems. In addition, those results follow from the analysis of general power penalty functions, of which quadratic penalty is a special case.

Keywords Penalty function · Quadratic penalty · Convergence rate · Strong regularity · Linear independence constraint qualification · Strict Mangasarian–Fromovitz constraint qualification · Second-order sufficient optimality conditions

Mathematics Subject Classification 90C30 · 65K05

✉ M. V. Solodov
solodov@impa.br

A. F. Izmailov
izmaf@cs.msu.ru

[1] OR Department, VMK Faculty, Lomonosov Moscow State University, MSU, Uchebniy Korpus 2, Leninskiye Gory, Moscow, Russia 119991

[2] IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil

1 Introduction

Penalty functions belong to classical tools of theoretical and numerical optimization, dating back at least to [9]. Penalty methods are discussed in most Optimization textbooks; for example, [11], [3, Section 2.1], [12, Section 12.1], [21, Sections 8.2.5, 9.4.3], [8, Chapter 14], [20, Section 17.1], [25, Section 6.2].

To start with, consider the general constrained optimization problem

$$\text{minimize } f(x) \ \text{ subject to } \ x \in D, \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $D \subset \mathbb{R}^n$. Recall that an exterior penalty for the set $D \subset \mathbb{R}^n$ is any function $\psi : \mathbb{R}^n \to \mathbb{R}$ such that

$$\psi(x) = 0 \quad \forall\, x \in D, \quad \psi(x) > 0 \quad \forall\, x \in \mathbb{R}^n \setminus D.$$

For a chosen penalty, one then defines the family of penalty functions $\varphi_c : \mathbb{R}^n \to \mathbb{R}$,

$$\varphi_c(x) = f(x) + c\psi(x), \tag{1.2}$$

where $c > 0$ is the penalty parameter. The basic penalty method consists in sequentially solving the unconstrained subproblems

$$\text{minimize } \varphi_c(x), \quad x \in \mathbb{R}^n, \tag{1.3}$$

for growing values of c, until a point satisfying some (approximate) stationarity condition for problem (1.1) is obtained.

Under the continuity assumptions on f and ψ, the basic convergence result states that as the penalty parameter c increases to infinity, every accumulation point of (global) solutions of the sequence of penalty subproblems (1.3) is a (global) solution of the original problem (1.1). See, e.g., [12, Theorem 12.1.1], [25, Theorem 6.6], [20, Theorem 17.1]. In Sect. 2 below, we shall also state and use a somewhat more refined result concerning approximating a strict local solution of the original problem by local solutions of penalty subproblems.

At the same time, results about *convergence rate* estimates, characterizing the distance from a solution x_c of the penalty method subproblem (1.3) to a solution $\bar{x}$ of the original problem (1.1), appear to be few in the literature. We discuss this type of results next.

Consider now the equality- and inequality-constrained problem

$$\text{minimize } f(x) \ \text{ subject to } \ h(x) = 0, \ g(x) \leq 0, \tag{1.4}$$

with a smooth objective function $f : \mathbb{R}^n \to \mathbb{R}$ and smooth constraint mappings $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ (the specific smoothness assumptions will be imposed in the sequel, as needed). Consider further the specific quadratic penalty, given by

$$\psi(x) = \|h(x)\|_2^2 + \|\max\{0, \ g(x)\}\|_2^2. \tag{1.5}$$

For equality-constrained optimization problems (no inequality constraints), convergence rate estimates for the corresponding quadratic penalty method were obtained in [2, 22]; see also [12, Theorem 12.1.2] and [21, Theorem 8.2.7]. A reference to a formal statement extending the results to problems with inequality constraints, that the authors are aware of, is [15, Theorem 4.7.3]. However, the understanding of such extensions had certainly appeared earlier in other literature, as (under strict complementarity) the proof is via the standard technique of using squared slack variables to reformulate inequality constraints as equalities; see, e.g., [2, Section 5]. The analysis of convergence rates of the quadratic penalty method relies on the following assumptions: the linear independence constraint qualification (implying the uniqueness of the Lagrange multiplier $(\bar{\lambda},\ \bar{\mu})$ associated to the solution $\bar{x}$ in question), the strict complementarity condition for this Lagrange multiplier, and the second-order sufficient optimality condition. Under these assumptions, it was shown that

$$x_c - \bar{x} = O(1/c) \tag{1.6}$$

as $c \to \infty$.

However, this set of assumptions above is rather restrictive. In what follows, we show that the estimate (1.6) still holds for the quadratic penalty method even if strict complementarity is omitted, provided the strong second-order sufficient optimality condition holds (despite the word "strong", the latter is a weaker assumption than the combination of strict complementarity and the usual second-order sufficient optimality condition). We also show that in this case, stationary points of the penalty subproblems are locally unique and Lipschitz-continuous with respect to $1/c$. Moreover, the strict complementarity can be omitted even in the case when the linear independence constraint qualification is relaxed to the weaker strict Mangasarian–Fromovitz constraint qualification, and under the usual second-order sufficient optimality condition. Although the local uniqueness of the solution of a penalty subproblem generally need not hold in this case (or at least, it does not follow from the analysis presented in this work). In addition, those results are obtained from a more general analysis of power-penalty functions, of which quadratic penalties is one special case.

The next issue is the useful Lagrange multiplier estimates that can be obtained from solving the penalty subproblem (1.3), and their distance to the optimal multipliers. It is well known that if we define

$$\lambda_c = 2ch(x_c), \quad \mu_c = 2c\max\{0,\ g(x_c)\}, \tag{1.7}$$

where x_c solves (1.3) with the quadratic penalty and the max-operation is applied component-wise, then these quantities serve as estimates of Lagrange multipliers associated to a solution of (1.4). Specifically, as $c \to +\infty$, under the Mangasarian–Fromovitz constraint qualification, these estimates are bounded and every accumulation point is a Lagrange multiplier associated to a solution $\bar{x}$ of (1.4) (assuming $\bar{x}$ is an accumulation point of the primal trajectory x_c). We refer to [25, Theorem 6.7] for the general case. For equality-constrained problems see also, e.g., [20, Theorem 17.2]. As for convergence rates for the multiplier estimates given by (1.7), for equality-constrained problems they again date back to [2, 22]; see also [12, Theorem 12.1.2]

and [21, Theorem 8.2.7]. For problems with inequality constraints, see [15, Theorem 4.7.3]. Specifically, under the same conditions as for the primal estimate (1.6), namely, the linear independence constraint qualification, the strict complementarity condition, and the second-order sufficient optimality condition, it holds that

$$2ch(x_c) - \bar{\lambda} = O(1/c), \quad 2c \max\{0,\ g(x_c)\} - \bar{\mu} = O(1/c), \tag{1.8}$$

as $c \to +\infty$, where $(\bar{\lambda},\ \bar{\mu})$ is the unique Lagrange multiplier associated to $\bar{x}$.

As for the primal estimate (1.6), again we show that the estimates (1.8) are still valid under the weaker assumptions: the strict Mangasarian–Fromovitz constraint qualification and second-order sufficiency.

We note that even though smooth exterior penalty methods (like quadratic penalties) are rarely employed nowadays by computational experts to solve a problem directly, they may still be useful for some purposes. In particular, the property (1.8) gives some information about how to obtain good multiplier estimates λ_c and μ_c by solving one (or perhaps a few) smooth unconstrained optimization problems (1.3). These good multiplier estimates can then be used as a starting point in other primal-dual algorithms; see, e.g., some discussions in [8, Chapter 14]. For example, good multiplier starting points are desirable in the augmented Lagrangian methods [2, 4], where they are needed for fast convergence of the primal-dual iterates [10].

Concluding our literature overview of convergence rates of the distance to solutions in penalty methods, we note that the linear independence constraint qualification and strict complementarity condition were avoided also in the analysis of primal convergence estimates presented in [1]. In that reference, the only assumptions are various sufficient optimality conditions, and in particular, the set of Lagrange multipliers associated to $\bar{x}$ need not be a singleton. However, in [1] only primal rates are considered, and the estimate obtained for the quadratic penalty method under the second-order sufficient optimality condition has the form

$$x_c - \bar{x} = O\left(\frac{1}{c^{1/2}}\right).$$

This estimate is weaker than (1.6), and [1, Example 5] demonstrates that it cannot be improved even under the Mangasarian–Fromovitz constraint qualification, even if there exist Lagrange multipliers satisfying the strict complementarity condition, and the second-order sufficient optimality condition holds with some universal (i.e., not depending on critical directions) multiplier possessing this property. Our approach follows the spirit of [1], in that it consists of introducing an auxiliary subproblem related to the penalty method subproblem, in such a way that the former can be interpreted as a right-hand side perturbation of the constraints of the original problem, and x_c is a solution of that auxiliary problem. Then the auxiliary problem is tackled by some tools of sensitivity theory for optimization problems. In this sense, the main difference with [1] consists precisely in the appropriate selection and use of sensitivity tools that allow to obtain new results ensuring that primal (1.6) and dual (1.8) estimates hold under weaker than previous assumptions. We also show that under these weaker assumptions, the estimates (1.6) and (1.8) cannot be improved.

Finally, we mention that there exist other kind of convergence rate results for penalty methods, assessing the difference of the objective function values at x_c and $\bar{x}$ (rather than the distance form x_c to $\bar{x}$, as in the current work). This approach employs error bounds for constraints (estimates of the distance to the feasible set in terms of constraints violations); see, e.g., [14] and [15, Theorem 4.7.5]. The use of error bounds is also closely related to the so-called exact penalization principle; see [7, 18, 19], and also [17, Section 2.1], [5, Proposition 3.111, Theorem 3.112].

The rest of the paper is organized as follows. Section 2 contains some preliminary information and necessary tools from sensitivity theory for optimization problems. These tools are further employed in Sect. 3 to obtain the main result on primal convergence rates of power penalty methods under SMFCQ and SOSC. In Sect. 4, under the appropriate additional assumptions, we derive further improvements of the result from Sect. 3. Specifically, we characterize the quality of dual approximations, and demonstrate local uniqueness and Lipschitz continuity of stationary points of the quadratic penalty subproblems under LICQ and SSOSC, the latter by making use of the celebrated Robinson's theorem on strongly regular solutions of generalized equations. Section 5 provides some examples illustrating the results obtained, and demonstrating sharpness of the estimates and the need of some of the assumptions used in the paper. Finally, Sect. 6 summarizes contributions of the paper and outlines some open questions.

Some words about our notation. For a vector $z \in \mathbb{R}^m$ and an index set $J \subset \{1, \ldots, m\}$, by z_J we mean the vector of components of z indexed by J. Then, $z_{\setminus J}$ stands for the vector of components of z indexed by $\{1, \ldots, m\} \setminus J$. We write $\|z\|$ for any norm, when the specific choice of the norm is of no importance. In other cases, the notation $\|z\|_\infty$ and $\|z\|_q$ for $q \geq 1$ is standard. In the expression z^q with any real exponent q, the power is applied component-wise. By $B(\bar{x}, \delta) = \{x \in \mathbb{R}^n \mid \|x - \bar{x}\| \leq \delta\}$ we denote the closed ball of radius $\delta \geq 0$, centered at $\bar{x} \in \mathbb{R}^n$. Finally, by $N_Q(\cdot)$ we denote the standard normal cone multifunction to a convex set Q.

2 Preliminaries

As already mentioned, when the penalty parameter tends to infinity, global solutions of subproblems (1.3) accumulate to global solutions of the original problem (1.1). The following convergence result (e.g., [15, Theorem 4.7.1]) concerns approximating a strict local solution of (1.1) by some local solutions of (1.3).

Let f be continuous near a strict local solution $\bar{x}$ of problem (1.1), i.e., there exists $\delta > 0$ such that $\bar{x}$ is the unique global solution of the restricted problem

$$\text{minimize } f(x) \quad \text{subject to } \quad x \in D \cap B(\bar{x}, \delta).$$

Let ψ be an exterior penalty for D, continuous near $\bar{x}$. Then for any choice of a global solution x_c of problem

$$\text{minimize } \varphi_c(x) \quad \text{subject to } \quad x \in B(\bar{x}, \delta), \tag{2.1}$$

it holds that

$$x_c \to \bar{x} \ \text{ as } \ c \to +\infty. \tag{2.2}$$

In particular, for all c large enough, the point x_c is a local solution of problem (1.3). In other words, every strict local solution of problem (1.1) is approximated by some local solutions of the penalty subproblem (1.3) as the penalty parameter tends to infinity.

The subsequent developments are concerned with the specific form of constraints defining the feasible set D, namely, with those in (1.4). Let $L : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ be the Lagrangian of problem (1.4), i.e.,

$$L(x,\ \lambda,\ \mu) = f(x) + \langle \lambda,\ h(x) \rangle + \langle \mu,\ g(x) \rangle.$$

Then the Lagrange multipliers associated with a point $\bar{x} \in \mathbb{R}^n$, feasible in (1.4), are defined as elements of the set

$$\mathcal{M}(\bar{x}) = \left\{ (\lambda,\ \mu) \in \mathbb{R}^l \times \mathbb{R}^m_+ \;\middle|\; \frac{\partial L}{\partial x}(\bar{x},\ \lambda,\ \mu) = 0,\ \ \mu_{\setminus A} = 0 \right\},$$

where

$$A = A(\bar{x}) = \{ i \in \{1,\ \ldots,\ m\} \mid g_i(\bar{x}) = 0 \}$$

is the set of indices of inequality constraints active at $\bar{x}$. A feasible point $\bar{x}$ is called stationary for problem (1.4) if $\mathcal{M}(\bar{x}) \neq \emptyset$.

The linear independence constraint qualification (LICQ) at $\bar{x}$ consists of saying that

$$\begin{pmatrix} h'(\bar{x}) \\ g'_A(\bar{x}) \end{pmatrix} \text{ has full row rank.}$$

If LICQ holds at a local solution $\bar{x}$ of problem (1.4), then $\bar{x}$ is a stationary point, and moreover, if LICQ holds at a stationary point $\bar{x}$ of problem (1.4), then $\mathcal{M}(\bar{x})$ is a singleton. The latter property (uniqueness of the Lagrange multiplier associated to $\bar{x}$) is known as the strict Mangasarian–Fromovitz constraint qualification (SMFCQ). Generally, SMFCQ is a weaker condition than LICQ, but stronger than the Mangasarian–Fromovitz constraint qualification (MFCQ) that is stated as follows:

$$\operatorname{rank} h'(\bar{x}) = l$$

and

$$\exists\, \bar{\xi} \in \mathbb{R}^n \text{ such that } h'(\bar{x})\bar{\xi} = 0,\ \ g'_A(\bar{x})\bar{\xi} < 0.$$

At a stationary point $\bar{x}$ of problem (1.4), MFCQ is equivalent to saying that $\mathcal{M}(\bar{x})$ is bounded (the fact first pointed out in [13]), which is evidently implied by SMFCQ.

The second-order sufficient optimality condition (SOSC) at $\bar{x}$ for $(\bar{\lambda},\ \bar{\mu}) \in \mathcal{M}(\bar{x})$ is the following property:

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\ \bar{\lambda},\ \bar{\mu})\xi,\ \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\},$$

where

$$C(\bar{x}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_A(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\ \xi\rangle \leq 0\}$$

is the critical cone of problem (1.4) at $\bar{x}$. The strong second-order sufficient condition (SSOSC) has the form

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\ \bar{\lambda},\ \bar{\mu})\xi,\ \xi \right\rangle > 0 \quad \forall\, \xi \in C_+(\bar{x},\ \bar{\mu}) \setminus \{0\},$$

where

$$\begin{aligned} C_+(\bar{x},\ \bar{\mu}) &= \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+}(\bar{x})\xi = 0\}, \\ A_+ &= A_+(\bar{x},\ \bar{\mu}) = \{i \in A \mid \bar{\mu}_i > 0\}. \end{aligned}$$

Recall that for any $(\bar{\lambda},\ \bar{\mu}) \in \mathcal{M}(\bar{x})$, the critical cone can be written in the form

$$C(\bar{x}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A_+}(\bar{x})\xi = 0,\ g'_{A\setminus A_+}(\bar{x})\xi \leq 0\}.$$

This shows, in particular, that under the strict complementarity condition, i.e., when $\bar{\mu}_A > 0$, it holds that $C_+(\bar{x},\ \bar{\mu}) = C(\bar{x})$. Hence, under strict complementarity, SSOSC is the same as SOSC. In general, as $C(\bar{x}) \subset C_+(\bar{x},\ \bar{\mu})$, SSOSC is a more restrictive assumption than SOSC. However, despite the word "strong" in SSOSC, it is weaker than the combination of strict complementarity with SOSC.

Define the residual mapping of the constraints in (1.4): $\Psi : \mathbb{R}^n \to \mathbb{R}^l \times \mathbb{R}^m$,

$$\Psi(x) = (h(x),\ \max\{0,\ g(x)\}). \tag{2.3}$$

Then the so-called power penalty has the form

$$\psi(x) = (\rho(\Psi(x)))^q, \tag{2.4}$$

where $q > 0$ is a fixed exponent, and a function $\rho : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}_+$ satisfies the following condition:

(i) $\rho(0,\ 0) = 0,\ \rho(y,\ z) > 0\ \ \forall\ (y,\ z) \in (\mathbb{R}^l \times \mathbb{R}^m)\setminus\{(0,\ 0)\}$, and ρ is continuous near $(0,\ 0)$.

In this case, function ψ defined by (2.3)–(2.4) is a penalty for the feasible set of problem (1.4). The analysis below employs the following further assumptions on ρ:

(ii) ρ is monotone in the following sense: if $z^1,\ z^2 \in \mathbb{R}^m$ satisfy $0 \leq z^1 \leq z^2$, then $\rho(y,\ z^1) \leq \rho(y,\ z^2)\ \forall\, y \in \mathbb{R}^l$.

(iii) ρ majorizes norm, i.e.,

$$(y,\ z) = O(\rho(y,\ z))$$

holds as $(y,\ z) \in \mathbb{R}^l \times \mathbb{R}^m$ tends to $(0,\ 0)$.

The two typical choices of ρ in (2.4) are

$$\rho(y,\ z) = \|(y,\ z)\|_\infty = \max\{\|y\|_\infty,\ \|z\|_\infty\},$$

and

$$\rho(y,\ z) = \left(\sum_{j=1}^{l} |y_j|^q + \sum_{i=1}^{m} |z_i|^q\right)^{1/q}.$$

If $q \geq 1$, then the right-hand side of the latter equality is $\|(y,\ z)\|_q$, but one can easily see that both these choices of ρ satisfy (i)–(iii) for any value of $q > 0$. With these choices, (2.4) gives the penalties

$$\psi(x) = (\max\{|h_1(x)|,\ \ldots\ |h_l(x)|,\ 0,\ g_1(x),\ \ldots,\ g_m(x)\})^q$$

(with 0 included for the case when $l = 0$, i.e., there are no equality constraints) and

$$\psi(x) = \sum_{j=1}^{l} |h_j(x)|^q + \sum_{i=1}^{m} (\max\{0,\ g_i(x)\})^q, \tag{2.5}$$

respectively. In particular, if $q = 2$, then (2.5) gives the quadratic penalty (1.5).

The following observation will be used below. Since $\bar{x}$ is feasible in both problems (1.4) and (2.1), from (1.2), (2.3)–(2.4), and (i), one can readily see that $\Psi(\bar{x}) = 0$, and for any global solution x_c of problem (2.1) it holds that

$$f(x_c) + c(\rho(\Psi(x_c)))^q \leq f(\bar{x}). \tag{2.6}$$

We complete this section by recalling some results from sensitivity theory for optimization problems, which will play the central role in our analysis below. For every pair $(y,\ z) \in \mathbb{R}^l \times \mathbb{R}^m$, consider the following optimization problem, obtained by the right-hand side perturbations of the constraints in (1.4):

$$\text{minimize } f(x) \ \text{ subject to } \ h(x) = y,\ g(x) \leq z. \tag{2.7}$$

Proposition 2.1 below follows, e.g., from [24, Theorem 3.2 and Corollary 4.3]. Its assumptions do imply that the point $\bar{x}$ in question is a strict local solution of problem

(1.4): if $\delta > 0$ is chosen small enough, then $\bar{x}$ is the unique global solution of the restricted problem

$$\text{minimize } f(x) \text{ subject to } h(x) = 0,\ g(x) \leq 0,\ x \in B(\bar{x},\ \delta). \tag{2.8}$$

Consider now the corresponding restricted version of the perturbed problem (2.7):

$$\text{minimize } f(x) \text{ subject to } h(x) = y,\ g(x) \leq z,\ x \in B(\bar{x},\ \delta), \tag{2.9}$$

and let $S_\delta(y,\ z)$ stand for the set of (global) solutions of this problem.

Proposition 2.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$, and $g : \mathbb{R}^n \to \mathbb{R}^m$ be twice differentiable at $\bar{x} \in \mathbb{R}^n$, and let h and g be continuously differentiable near $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (1.4), let SMFCQ hold, and let SOSC hold for the associated Lagrange multiplier $(\bar{\lambda},\ \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$.*

Then for any $\delta > 0$ small enough, it holds that

$$\sup_{x \in S_\delta(y,\ z)} \|x - \bar{x}\| = O(\|(y,\ z)\|) \tag{2.10}$$

as $(y,\ z) \to (0,\ 0)$. Moreover, for every $(y,\ z)$ close enough to $(0,\ 0)$, every $x \in S_\delta(y,\ z)$ is a stationary point of problem (2.7), and for the set $\mathcal{M}_{y,\ z}(x) \neq \emptyset$ of associated Lagrange multipliers, it holds that

$$\sup_{\substack{x \in S_\delta(y,\ z),\\ (\lambda,\ \mu) \in \mathcal{M}_{y,\ z}(x)}} \|(\lambda,\ \mu) - (\bar{\lambda},\ \bar{\mu})\| = O(\|(y,\ z)\|) \tag{2.11}$$

as $(y,\ z) \to (0,\ 0)$.

Recall finally that, as demonstrated in [23] (see also, e.g., [5, Proposition 5.38], [16, Proposition 1.28]), if SMFCQ and SOSC in Proposition 2.1 are replaced by the stronger combination of assumptions LICQ and SSOSC, then $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$ is a strongly regular solution of the generalized equation corresponding to the primal-dual first-order optimality conditions for problem (1.4). Moreover, according to [6] (see also, e.g., [16, Proposition 1.28]) the stronger conditions LICQ and SSOSC are necessary for strong regularity of the $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$ when $\bar{x}$ is a local solution of (1.4).

3 General primal convergence rate estimate

We start with the case when SMFCQ and SOSC are assumed to hold, forming the weakest combination of assumptions used in this paper. The analysis will rely on the following considerations.

For every value of the penalty parameter $c > 0$, let x_c be a global solution of problem (2.1). Recall that if $\delta > 0$ is small enough, this x_c necessarily satisfies (2.2). Consider the auxiliary optimization problem

$$\text{minimize } f(x) \text{ subject to } h(x) = h(x_c),\ g(x) \leq \max\{0,\ g(x_c)\},\ x \in B, \tag{3.1}$$

for now with an abstract set $B \subset \mathbb{R}^n$. With x_c replaced by $\bar{x}$, and with $B = B(\bar{x},\ \delta)$, this problem transforms into (2.8), and therefore, (3.1) can be regarded as a result of the right-hand side perturbation of constraints in (2.8), i.e., as (2.9) with $(y, z) = \Psi(x_c)$, where Ψ is defined in (2.3). Moreover, from (2.2) it follows that this $(y, z) \to (0, 0)$ as $c \to +\infty$.

The next result [1] relates the rate of convergence analysis of penalty methods to sensitivity analysis of the solution $\bar{x}$ of problem (2.8) with respect to the perturbation specified above. We include its short proof for the sake of clarity.

Lemma 3.1 *For any $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$, and $g : \mathbb{R}^n \to \mathbb{R}^m$, and any $c \geq 0$, let φ_c be defined according to (1.2) and (2.3)–(2.4), with a function $\rho : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}_+$ satisfying (ii).*

Then for any set $B \subset \mathbb{R}^n$, any global solution x_c of problem

$$\text{minimize } \varphi_c(x) \text{ subject to } x \in B, \tag{3.2}$$

is a global solution of problem (3.1).

Proof Take any $\widetilde{x}$ feasible in (3.1). Then it is also feasible in (3.2), and by (1.2), (2.3)–(2.4), we obtain that

$$\begin{aligned} f(x_c) + c(\rho(h(x_c),\ \max\{0,\ g(x_c)\}))^q &= \varphi_c(x_c) \\ &\leq \varphi_c(\widetilde{x}) \\ &= f(\widetilde{x}) + c(\rho(h(\widetilde{x}),\ \max\{0,\ g(\widetilde{x})\}))^q \\ &\leq f(\widetilde{x}) + c(\rho(h(x_c),\ \max\{0,\ g(x_c)\}))^q, \end{aligned}$$

where the last inequality is by the monotonicity of ρ (property (ii)) and the feasibility of $\widetilde{x}$ in (3.1). The obtained inequality readily yields $f(x_c) \leq f(\widetilde{x})$. □

We are now in a position to establish the primal convergence rate estimates.

Theorem 3.1 *Under the assumptions of Proposition 2.1, for any $c > 0$, let φ_c be defined according to (1.2) and (2.3)–(2.4), with a function $\rho : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}_+$ satisfying (i)–(iii).*

Then for any $\delta > 0$ small enough, and any global solution x_c of problem (2.1) with $c > 0$, the following assertions are valid:

- *If $q > 1$, then the estimates*

$$x_c - \bar{x} = O\left(\frac{1}{c^{1/(q-1)}}\right) \tag{3.3}$$

and

$$\Psi(x_c) = O\left(\frac{1}{c^{1/(q-1)}}\right). \tag{3.4}$$

hold as $c \to +\infty$.

- *If $q \in (0,\ 1]$, then $x_c = \bar{x}$ for all $c > 0$ large enough.*

Proof Fix $\delta > 0$ as in Sect. 2, i.e., such that $\bar{x}$ is the unique global solution of problem (2.8). Let x_c be any global solution of problem (2.1). From Lemma 3.1 it then follows that x_c is a global solution of (3.1) with $B = B(\bar{x},\ \delta)$. Taking into account (2.2), from estimate (2.10) in Proposition 2.1 we then derive (after further reducing $\delta > 0$ if necessary) that

$$x_c - \bar{x} = O(\|\Psi(x_c)\|) \tag{3.5}$$

as $c \to +\infty$.

Next, from (iii), (2.6), and (3.5), we obtain that

$$\begin{aligned}\Psi(x_c) &= O(\rho(\Psi(x_c))) \\ &= O\left(\left(\frac{f(\bar{x}) - f(x_c)}{c}\right)^{1/q}\right) \\ &= O\left(\left(\frac{\|x_c - \bar{x}\|}{c}\right)^{1/q}\right) \\ &= O\left(\left(\frac{\|\Psi(x_c)\|}{c}\right)^{1/q}\right).\end{aligned} \tag{3.6}$$

For $q > 1$, this implies the estimate (3.4), and combining the latter with (3.5) we obtain the claimed (3.3).

Finally, if $q \in (0,\ 1]$, then (3.6) implies that for all c large enough it holds that $\Psi(x_c) = 0$, and hence, according to (3.5), $x_c = \bar{x}$. □

Note that SMFCQ in Theorem 3.1 cannot be replaced by the weaker MFCQ. As mentioned in Sect. 1, this can be demonstrated by [1, Example 5].

The second assertion of Theorem 3.1 means that under its assumptions, the power penalty is in a sense exact for the exponent values $q \in (0,\ 1]$.

4 Further results under further assumptions

We proceed with analyzing convergence rates for the multiplier estimates, which for the quadratic penalty are given by (1.7). We shall consider the more general case when $\rho(\cdot) = \|\cdot\|_q$ with $q > 1$ or, in other words, when the penalty is defined by (2.5).

Lemma 4.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$, and $g : \mathbb{R}^n \to \mathbb{R}^m$ be given. For any $c > 0$, let φ_c be defined according to (1.2) and (2.5) with $q > 1$.*

Then any stationary point x_c of the penalty subproblem (1.3), i.e., a point satisfying $\varphi_c'(x_c) = 0$, where f, h, and g are differentiable at x_c, is a stationary point of problem (3.1) with $B = \mathbb{R}^n$. Furthermore, the associated to x_c in problem (3.1) with $B = \mathbb{R}^n$ Lagrange multiplier $(\lambda_c,\ \mu_c)$ is given by

$$\lambda_c = qc\widehat{h}(x_c), \quad \mu_c = qc(\max\{0,\ g(x_c)\})^{q-1}, \tag{4.1}$$

where the components of $\widehat{h}(x_c) \in \mathbb{R}^l$ *are defined by*

$$\widehat{h}_j(x_c) = \begin{cases} (h_j(x_c))^{q-1} & \text{if } h_j(x_c) \geq 0, \\ -(-h_j(x_c))^{q-1} & \text{if } h_j(x_c) < 0. \end{cases} \tag{4.2}$$

Moreover, if $q = 2$*, then* x_c *is a stationary point of the penalty subproblem (1.3) if and only if* $(x_c,\ \lambda_c,\ \mu_c)$*, with* λ_c *and* μ_c *defined in (1.7), is a solution of the generalized equation*

$$\Phi(\sigma,\ u) + N_Q(u) \ni 0 \tag{4.3}$$

in variable $u = (x,\ \lambda,\ \mu)$*, where* $\sigma = 1/(2c)$*,* $\Phi : \mathbb{R} \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+) \to \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+$*,*

$$\Phi(\sigma,\ u) = \left(\frac{\partial L}{\partial x}(x,\ \lambda,\ \mu),\ h(x) - \sigma\lambda,\ -g(x) + \sigma\mu\right), \tag{4.4}$$

and $Q = \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+$*.*

Proof Observe that the penalty function φ_c, defined according to (1.2) and (2.5) with $q > 1$, is differentiable at any $x \in \mathbb{R}^n$ in the domain of differentiability of f, h, and g, and

$$\varphi_c'(x) = f'(x) + 2qc((h'(x))^\top \widehat{h}(x) + (g'(x))^\top (\max\{0,\ g(x)\})^{q-1}),$$

with $\widehat{h}(x)$ defined according to (4.2). Therefore, x_c is a stationary point of the penalty subproblem (1.3), i.e., $\varphi_c'(x_c) = 0$, if and only if

$$\frac{\partial L}{\partial x}(x_c,\ \lambda_c,\ \mu_c) = 0, \tag{4.5}$$

where λ_c and μ_c are given by (4.1)–(4.2).

Furthermore, it can be easily checked that the definition of μ_c in (4.1) yields

$$\min\left\{\left(\frac{\mu_c}{qc}\right)^{1/(q-1)},\ -g(x_c) + \left(\frac{\mu_c}{qc}\right)^{1/(q-1)}\right\} = 0,$$

which is evidently equivalent to

$$\min\left\{\mu_c,\ -g(x_c) + \left(\frac{\mu_c}{qc}\right)^{1/(q-1)}\right\} = 0. \tag{4.6}$$

This implies that $\mu_c \geq 0$, and $(\mu_c)_i = 0$ if $g_i(x_c) < 0$, and so $g_i(x_c) < \max\{0,\ g_i(x_c)\}$. Together with (4.5), this yields that x_c is a stationary point of problem (3.1) with $B = \mathbb{R}^n$, and $(\lambda_c,\ \mu_c)$ is an associated Lagrange multiplier.

Furthermore, assuming that $q = 2$, we have from (4.1)–(4.2) that λ_c and μ_c are defined as in (1.7), and (4.5)–(4.6) mean precisely that $(x_c,\ \lambda_c,\ \mu_c)$ is a solution of

the generalized equation (4.3) with the ingredients specified in the statement of the lemma. □

This lemma suggests, in particular, that one may expect $(\lambda_c,\ \mu_c)$ defined in (4.1)–(4.2) to be a relevant approximation of $(\bar{\lambda},\ \bar{\mu})$, obtained by using a solution x_c of the penalty subproblem, close enough to $\bar{x}$. This expectation is confirmed by the following result.

Proposition 4.1 *Under the assumptions of Proposition 2.1, for any $c > 0$, let φ_c be defined according to (1.2) and (2.5) with $q > 1$.*

Then for any $\delta > 0$ small enough, and any global solution x_c of problem (2.1) with $c > 0$, in addition to (3.3), the estimates

$$\lambda_c - \bar{\lambda} = O\left(\frac{1}{c^{1/(q-1)}}\right), \quad \mu_c - \bar{\mu} = O\left(\frac{1}{c^{1/(q-1)}}\right) \tag{4.7}$$

hold as $c \to +\infty$, for λ_c and μ_c defined according to (4.1)–(4.2).

Proof For any fixed $\delta > 0$, from (2.2) it follows that x_c is a stationary point of problem (1.3) provided $c > 0$ is large enough. From Lemma 4.1 we then have that for $(\lambda_c,\ \mu_c)$ defined according to (4.1)–(4.2), it holds that $(\lambda_c,\ \mu_c) \in \mathcal{M}_{\Psi(x_c)}(x_c)$. Therefore, from estimate (2.11) in Proposition 2.1 we derive (after further reducing $\delta > 0$ if necessary) that

$$\lambda_c - \bar{\lambda} = O(\|\Psi(x_c)\|), \quad \mu_c - \bar{\mu} = O(\|\Psi(x_c)\|)$$

as $c \to +\infty$. Employing now Theorem 3.1, and in particular, the estimate (3.4), we arrive at the needed conclusion. □

We now proceed with the case when the penalty is quadratic, and instead of SMFCQ and SOSC, the stronger LICQ and SSOSC are assumed to hold. As discussed in Sect. 2, this implies that $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$ is a strongly regular solution of the generalized equation corresponding to the primal-dual first-order optimality conditions for problem (1.4). Observe that this generalized equation is obtained from (4.3) considered in Lemma 4.1 by setting $\sigma = 0$.

Theorem 4.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$, and $g : \mathbb{R}^n \to \mathbb{R}^m$ be twice continuously differentiable near $\bar{x} \in \mathbb{R}^n$. Let $\bar{x}$ be a stationary point of problem (1.4), assume that LICQ holds at $\bar{x}$, and that SSOSC holds for the associated Lagrange multiplier $(\bar{\lambda},\ \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$. For any $c > 0$, let φ_c be defined according to (1.2) and (1.5).*

Then for every $c > 0$ large enough, the penalty subproblem (1.3) has near $\bar{x}$ the unique stationary point x_c, this point is a local solution of (1.3), and the estimates (1.6) and (1.8) hold as $c \to +\infty$, for λ_c and μ_c defined according to (1.7). Moreover, x_c, λ_c and μ_c are Lipschitz-continuous, considered as functions of $1/c$.

Proof Consider the parametric generalized equation (4.3) with Φ defined by (4.4), and with $N_Q(\cdot)$ being the normal cone multifunction to the set $Q = \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+$. As discussed above, $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$ is a strongly regular solution of this generalized equation,

considered with the parameter σ equal to 0. Our smoothness assumptions guarantee continuous differentiability of Φ near $(0,\ (\bar{x},\ \bar{\lambda},\ \bar{\mu}))$, and allow to apply Robinson's theorem on strongly regular solutions of generalized equations [23]. In particular, it holds that for every σ close enough to 0, (4.3) has near $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$ the unique solution $(x(\sigma),\ \lambda(\sigma),\ \mu(\sigma))$, and the dependence of this solution on σ is Lipschitz-continuous near 0.

Fix $\delta > 0$ as in Sect. 2, i.e., such that $\bar{x}$ is the unique global solution of problem (2.8). Let x_c be any global solution of problem (2.1). Taking into account (2.2), this implies that x_c is local solution of problem (1.3), and hence, a stationary point of this problem provided $c > 0$ is large enough.

We will now show that for any stationary point x_c of problem (1.3), close enough to $\bar{x}$ (including the one obtained as a global solution of (2.1)), it necessarily holds that $x_c = x(\sigma)$ for all $c > 0$ large enough. We argue by contradiction: suppose there exist a sequence of positive reals $\{c_k\}$ and a sequence $\{x^k\} \subset \mathbb{R}^n$ such that $c_k \to +\infty$, $\{x^k\} \to \bar{x}$ as $k \to \infty$, and for all k it holds that $\varphi'_{c_k}(x^k) = 0$, $x^k \neq x(\sigma^k)$, where $\sigma_k = 1/(2c_k) \to 0$ as $k \to \infty$. For every k, using (1.7), set $(\lambda^k,\ \mu^k) = (\lambda_{c_k},\ \mu_{c_k})$. According to Lemma 4.1, for every k it holds that $(x^k,\ \lambda^k,\ \mu^k)$ is a solution of (4.3) with $\sigma = \sigma_k$, which can be expressed in the form

$$\frac{\partial L}{\partial x}(x^k,\ \lambda^k,\ \mu^k) = 0, \quad \min\{\mu^k,\ -g(x^k) + \sigma_k \mu^k\} = 0 \tag{4.8}$$

(see (4.5), and (4.6) where we take $q = 2$).

If the sequence $\{(\lambda^k,\ \mu^k)\}$ were to be bounded, we can assume (passing onto a subsequence if necessary) that it converges to some $(\widetilde{\lambda},\ \widetilde{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$. Then passing in (4.8) onto the limit as $k \to \infty$, we obtain that

$$\frac{\partial L}{\partial x}(\bar{x},\ \widetilde{\lambda},\ \widetilde{\mu}) = 0, \quad \min\{\widetilde{\mu},\ -g(\bar{x})\} = 0.$$

This implies that $(\widetilde{\lambda},\ \widetilde{\mu}) \in \mathcal{M}(\bar{x})$, where under LICQ we have that $\mathcal{M}(\bar{x}) = \{(\bar{\lambda},\ \bar{\mu})\}$. Therefore, $(\widetilde{\lambda},\ \widetilde{\mu}) = (\bar{\lambda},\ \bar{\mu})$, and hence, $\{(x^k,\ \lambda^k,\ \mu^k)\} \to (\bar{x},\ \bar{\lambda},\ \bar{\mu})$. But this contradicts the assumption that $x^k \neq x(\sigma^k)$ for all k, since for k large enough, $(x(\sigma^k),\ \lambda(\sigma^k),\ \mu(\sigma^k))$ is the only solution of (4.3) with $\sigma = \sigma_k$ near $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$.

Suppose now that the sequence $\{(\lambda^k,\ \mu^k)\}$ is unbounded. Then we can assume (passing onto a subsequence if necessary) that $(\lambda^k,\ \mu^k) \neq 0$ for all k, and the sequence $\{(\lambda^k,\ \mu^k)/\|(\lambda^k,\ \mu^k)\|\}$ converges to some nonzero $(\eta,\ \zeta) \in \mathbb{R}^l \times \mathbb{R}^m$. Observe that according to (1.7), $\sigma_k \mu^k = \max\{0,\ g(x^k)\}$ for all k, and feasibility of $\bar{x}$ in problem (1.4) then implies that $\{\sigma_k \mu^k\} \to 0$. Dividing both parts of the first equality in (4.8) by $\|(\lambda^k,\ \mu^k)\|$, we obtain

$$\frac{f'(x^k)}{\|(\lambda^k,\ \mu^k)\|} + (h'(x^k))^\top \frac{\lambda^k}{\|(\lambda^k,\ \mu^k)\|} + (g'(x^k))^\top \frac{\mu^k}{\|(\lambda^k,\ \mu^k)\|} = 0.$$

Furthermore, replacing the second equality in (4.8) by the equivalent one

$$\min\left\{\frac{\mu^k}{\|(\lambda^k,\ \mu^k)\|},\ -g(x^k)+\sigma_k\mu^k\right\}=0,$$

and passing onto the limit in the last two relations above as $k\to\infty$, we obtain that

$$(h'(\bar{x}))^\top\eta+(g'(\bar{x}))^\top\zeta=0,\quad \min\{\zeta,\ -g(\bar{x})\}=0.$$

The existence of a nonzero $(\eta,\ \zeta)$ satisfying these equalities contradicts (the dual form of) MFCQ (see, e.g., [16, Remark 1.1]), and hence contradicts the stronger LICQ.

The argument above demonstrates that if $c>0$ is large enough then, on the one hand, $x(\sigma)$ with $\sigma=1/(2c)$ is a local solution of problem (1.3), while on the other hand, in some neighborhood of $\bar{x}$ there are no other stationary points for this problem, and in particular, $x(\sigma)$ coincides with the global solution x_c of problem (2.8).

Finally, the estimates (1.6) and (1.8) now follow from Theorem 3.1 and Proposition 4.1, applied with $q=2$. Moreover, since the estimates in (1.8) imply that λ_c and μ_c get arbitrarily close to $\bar{\lambda}$ and $\bar{\mu}$, for $c>0$ large enough Lemma 4.1 implies that $\lambda_c=\lambda(\sigma)$ and $\mu_c=\mu(\sigma)$. In particular, x_c, λ_c and μ_c are Lipschitz-continuous in $\sigma=1/c$. □

Remark 4.1 Apart from application of Robinson's theorem on strongly regular solutions of generalized equations [23], the only key ingredient of the proof above consists of showing that if for every k it holds that $(x^k,\ \lambda^k,\ \mu^k)$ is a solution of (4.3) with $\sigma=\sigma_k$, and $\{x^k\}\to\bar{x}$, then necessarily $\{(\lambda^k,\ \mu^k)\}\to(\bar{\lambda},\ \bar{\mu})$. Observe that derivation of this property in the proof above is actually not making use of LICQ, but rather of the weaker SMFCQ only. This feature can be useful for further possible developments and improvements concerning uniqueness of stationary points of the penalty subproblems.

At the same time, as pointed out by one of the referees, actually using LICQ the needed property can be derived in a simpler and more direct manner. Indeed, since $\{x^k\}\to\bar{x}$, for all k large enough and for any $i\in\{1,\ \ldots,\ m\}\setminus A$ it holds that $g_i(x^k)<0$. Hence, by the second equality in (4.8), we have that $\mu_i^k=0$. Therefore, the first equality in (4.8) implies that

$$f'(x^k)+(h'(x^k))^\top\lambda^k+(g'_A(x^k))^\top\mu_A^k=0,$$

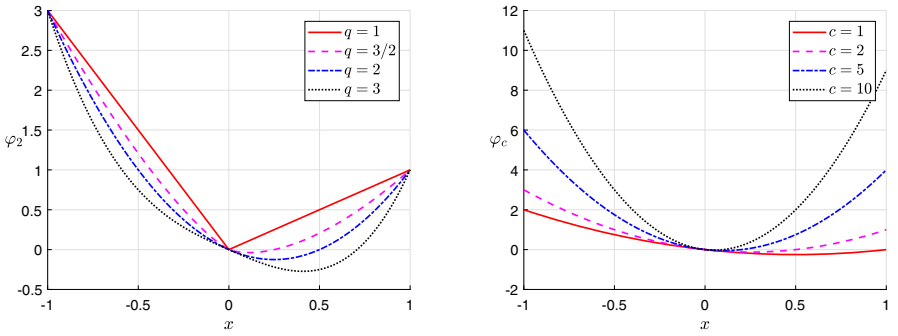
where, by LICQ, the columns of the matrices $(h'(x^k))^\top$ and $(g'_A(x^k))^\top$ form a uniformly linearly independent system for all k large enough. Since

$$f'(\bar{x})+(h'(\bar{x}))^\top\bar{\lambda}+(g'_A(\bar{x}))^\top\bar{\mu}_A=0$$

and $\bar{\mu}_i=0$ for all $i\in\{1,\ \ldots,\ m\}\setminus A$ as well, these observations imply that

$$(\lambda^k-\bar{\lambda},\ \mu^k-\bar{\mu})=O(\|f'(x^k)-f'(\bar{x})\|)$$

as $k\to\infty$. It follows that $\{(\lambda^k,\ \mu^k)\}\to(\bar{\lambda},\ \bar{\mu})$.



(a) $c = 2$ (b) $q = 2$

Fig. 1 Graphs of φ_c in Example 5.1

Apart from the fact that Theorem 4.1 deals only with the case of $q = 2$, the differences between Theorems 3.1 and 4.1 are as follows. The assumptions of Theorem 4.1 are stronger: they include LICQ and SSOSC (instead of SMFCQ and SOSC as in Theorem 3.1), and under these assumptions, the penalty subproblem (1.3) is claimed to have the unique stationary point. In Theorem 3.1, no uniqueness is claimed, and the estimate or the exactness property hold for global solutions of the restricted penalty subproblem (2.1).

5 Illustrating Examples

We complete the paper with some simple examples demonstrating that the estimates obtained herein are sharp, and that some of the imposed assumptions cannot be relaxed.

Example 5.1 Let $n = l = 1, m = 0, f(x) = -x, h(x) = x$. The unique feasible point $\bar{x} = 0$ of (1.4) is the global solution and the unique stationary point of this problem, with the unique associated Lagrange multiplier $\bar{\lambda} = 1$. It holds that $C(\bar{x}) = \{0\}$, and both LICQ and SSOSC are satisfied.

For every $c > 0$, let φ_c be defined according to (1.2) and (2.5). If $q \in (0, 1]$, the penalty subproblem (1.3) with $c > 1$ has the unique global solution $x_c = 0$, which is also its unique local solution. If $q > 1$, the unique stationary point of (1.3) is $x_c = 1/(qc)^{1/(q-1)}$, showing that the primal estimate (3.3) cannot be improved. Note that (4.1)–(4.2) yield $\lambda_c = 1$.

Figure 1a shows the graphs of φ_c with $c = 2$ for different values of q, while Fig. 1b shows the graphs with $q = 2$ for different values of c.

Furthermore, if we modify f by setting $f(x) = -x + |x|^q$ for a fixed $q > 1$, then the unique stationary point of (1.3) is $x_c = 1/(c + 1)^{1/(q-1)}$, and (4.1)–(4.2) yield $\lambda_c = c/(c+1)$. Then $\lambda_c - 1 = -1/(c+1)$, demonstrating that the dual estimate in (4.7) cannot be improved when $q = 2$ (sharpness of (4.7) for any $q > 1$ is demonstrated below).

The modified function $f(x) = -x + |x|^q$ is not twice differentiable at 0 when $q \in (1, 2)$, while twice differentiability is among the assumptions of Theorem 3.1,

Proposition 4.1, and Theorem 4.1. To avoid this imperfection (with respect to this example), one can replace f by a quadratic function $f(x) = -x + x^2$, though derivation of the needed properties in this case requires some manipulations. For $q > 1$ we have

$$\varphi_c'(x) = -1 + 2x + qc \begin{cases} x^{q-1} & \text{if } x \geq 0, \\ -(-x)^{q-1} & \text{if } x < 0, \end{cases}$$

and the equation $\varphi_c'(x) = 0$ may have solutions only when $x \geq 0$, in which case it takes the form $-1 + 2x + qcx^{q-1} = 0$. The left-hand side of this equation is negative for $x = 0$, and positive for $x \geq 1/2$, and therefore, this equation has a solution x_c. Since $q > 1$, it evidently holds that $x_c \to 0$ as $c \to +\infty$, and hence, $x_c^{q-1} = (1 - 2x_c)/(qc) = 1/(qc) + o(1/(qc))$, implying that

$$x_c = \frac{1}{(qc)^{1/(q-1)}} + o\left(\frac{1}{(qc)^{1/(q-1)}}\right)$$

as $c \to +\infty$. This demonstrates that the primal estimate (3.3) cannot be improved. Furthermore,

$$x_c^{q-1} = \frac{1}{qc} - \frac{2x_c}{qc} = \frac{1}{qc} - \frac{2}{(qc)^{q/(q-1)}} + o\left(\frac{1}{(qc)^{q/(q-1)}}\right),$$

and from (4.1)–(4.2) we then have

$$\lambda_c = qcx_c^{q-1} = 1 - \frac{2}{(qc)^{1/(q-1)}} + o\left(\frac{1}{(qc)^{1/(q-1)}}\right),$$

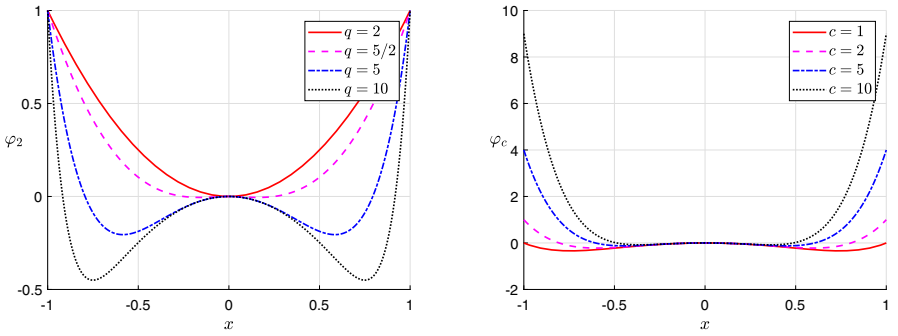
as $c \to +\infty$. This demonstrates that the dual estimate in (4.7) cannot be improved.

Example 5.2 Let $n = l = 1$, $m = 0$, $f(x) = -x^2$, $h(x) = x$. As in Example 5.1, the unique feasible point $\bar{x} = 0$ of (1.4) is the global solution and the unique stationary point of this problem, with the unique associated Lagrange multiplier $\bar{\lambda} = 0$, and both LICQ and SSOSC being satisfied again.

For every $c > 0$, let φ_c be defined according to (1.2) and (2.5). If $q > 2$, then along with a stationary point at 0, the penalty subproblem (1.3) has two other stationary points $x_c = \pm(2/(qc))^{1/(q-2)}$, and these points are global solutions of this subproblem. This shows that local uniqueness of stationary points established in Theorem 4.1 cannot be extended to the case of a power penalty (2.5) with $q > 2$.

Violation of local uniqueness can be seen from Fig. 2, which shows for this example the graphs of φ_c, in the same way as Fig. 1 for Example 5.1.

The next two examples illustrate the results of this work in the absence of strict complementarity. In Example 5.3, Theorem 4.1 is applicable.



(a) $c = 2$ (b) $q = 2.5$

Fig. 2 Graphs of φ_c in Example 5.2

Example 5.3 Let $n = 2, l = 0, m = 2, f(x) = x_1 + (x_1 + x_2)^2/2, g(x) = (-x_1, -x_2)$. The unique global solution and the unique stationary point of the corresponding problem (1.4) is $\bar{x} = 0$, with the unique associated Lagrange multiplier $\bar{\mu} = (1, 0)$. Strict complementarity is violated, but both LICQ and SSOSC are satisfied.

For every $c > 0$, the quadratic penalty subproblem has the unique stationary point $x_c = (-1/(2c), 1/(2c))$, and $2c \max\{0, g(x_c)\} = \bar{\mu}$.

Our final example shows a problem that can be tackled by Theorem 3.1 and Proposition 4.1, but not by Theorem 4.1.

Example 5.4 Consider the problem from Example 5.3, but with one additional constraint: let $m = 3$, $g(x) = (-x_1, -x_2, -x_1 - x_2)$. This modification does not change the solution: it is still $\bar{x} = 0$, with the unique associated Lagrange multiplier $\bar{\mu} = (1, 0, 0)$. Therefore, strict complementarity is violated, and LICQ is violated as well, but SMFCQ and SSOSC (hence SOSC) are satisfied.

For every $c > 0$, the quadratic penalty subproblem has the same unique stationary point $x_c = (-1/(2c), 1/(2c))$ as in Example 5.3, and $2c \max\{0, g(x_c)\} = \bar{\mu}$.

6 Conclusions

We have established the distance-to-solution estimates for power penalty methods under the assumptions weaker than previously used in the literature. Specifically, we assume the strict Mangasarian–Fromovitz constraint qualification and second-order sufficiency, with no strict complementarity assumption. Apart from primal estimates, we also derive the dual estimates for the natural related approximations of the Lagrange multiplier. In addition, we show local uniqueness and Lipschitz-continuity of the stationary points of the quadratic penalty subproblem under the linear independence constraint qualification and the strong second-order sufficient optimality condition. One interesting direction of further development of these results might be concerned with considering "lower-level" constraints not included in the penalty function. Another important question concerns allowing for controllable inexactness

when solving the penalty subproblems. Finally, we mention that we are not aware of an example demonstrating the lack of local uniqueness of solutions for the quadratic penalty subproblem under the strict Mangasarian–Fromovitz constraint qualification and second-order sufficiency, so this also remains an open question. The discussion in Remark 4.1 sheds some light on possible reasons of the corresponding difficulties.

Acknowledgements The authors thank the three referees for very helpful comments and suggestions that led us to re-organize and improve the original version of the paper. Research of the second author was supported in part by CNPq Grant 303913/2019-3, by FAPERJ Grant E-26/200.347/2023, and by PRONEX–Optimization.

Data Availibility Statement Data sharing not applicable to this article as no data sets were generated or analyzed during the current study.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest of any kind.

References

1. Avakov, E.R., Arutyunov, A.V., Izmailov, A.F.: On convergence rate estimates for power penalty methods. Comput. Math. Math. Phys. **44**, 1684–1695 (2004)
2. Bertsekas, D.P.: On penalty and multiplier methods for constrained minimization. SIAM J. Control. Optim. **14**, 216–235 (1976)
3. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York (1982)
4. Birgin, E.G., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. Fundamentals of Algorithms, vol. 10. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2014)
5. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
6. Bonnans, J.F., Sulem, A.: Pseudopower expansion of solutions of generalized equations and constrained optimization. Math. Program. **70**, 123–148 (1995)
7. Burke, J.V.: An exact penalization viewpoint of constrained optimization. SIAM J. Control. Optim. **29**, 968–998 (1991)
8. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust-Region Methods. SIAM, Philadelphia (2000)
9. Courant, R.: Variational methods for the solution of problems of equilibrium and vibration. Bull. Am. Math. Soc. **49**, 1–23 (1943)
10. Fernández, D., Solodov, M.V.: Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. SIAM J. Optim. **22**, 384–407 (2012)
11. Fiacco, A.V., McCormick, G.: Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York (1968)
12. Fletcher, R.: Practical Methods of Optimization, 2nd edn. Wiley, New York (1987)
13. Gauvin, J.: A necessary and sufficient regularity conditions to have bounded multipliers in nonconvex programming. Math. Program. **12**, 136–138 (1977)
14. Izmailov, A.F., Solodov, M.V.: Error bounds for 2-regular mappings with Lipschitzian derivatives and their applications. Math. Program. **89**, 413–435 (2001)
15. Izmailov, A.F., Solodov, M.V.: Numerical Methods for Optimization. Fizmatlit, Moscow (2003). (In Russain)
16. Izmailov, A.F., Solodov, M.V.: Newton-type Methods for Optimization and Variational Problems. Springer, Cham (2014)
17. Luo, Z.-Q., Pang, J.-S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)

18. Luo, Z.-Q., Pang, J.-S., Ralph, D., Wu, S.-Q.: Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints. Math. Program. **75**, 19–76 (1996)
19. Mangasarian, O.L.: Sufficiency of exact penalty minimization. SIAM J. Control. Optim. **23**, 30–37 (1985)
20. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, New York (2006)
21. Polyak, B.T.: Introduction to Optimization. Optimization Software Inc., Publications Division, New York (1987)
22. Polyak, B.T., Tret'yakov, N.V.: The method of penalty estimates for conditional extremum problems. USSR Comput. Math. Math. Phys. **13**, 42–58 (1973)
23. Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. **5**, 43–62 (1980)
24. Robinson, S.M.: Generalized equations and their solutions, Part II: applications to nonlinear programming. Math. Program. Study **10**, 200–221 (1982)
25. Ruszczynski, A.: Nonlinear Optimization. Princeton University Press, Princeton, NJ (2006)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.