A modified exact smooth penalty function for nonlinear constrained optimization

Liu Bingzhuang[*] and Zhao Wenling

[*]Correspondence: lbzlyj@126.com
School of Science, Shandong
University of Technology, Zibo,
Shandong 255049, P.R. China

Abstract

In this paper, a modified simple penalty function is proposed for a constrained nonlinear programming problem by augmenting the dimension of the program with a variable that controls the weight of the penalty terms. This penalty function enjoys improved smoothness. Under mild conditions, it can be proved to be exact in the sense that local minimizers of the original constrained problem are precisely the local minimizers of the associated penalty problem.

MSC: 47H20; 35K55; 90C30

Keywords: constrained minimization; exact penalty function; nonlinear programming problem

1 Introduction

Merit function has always taken an important role in optimization problem. It is traditionally constructed to solve nonlinear programs by augmenting the objective function or a corresponding Lagrange function some penalty or barrier terms with respect of the constraints. Then it can be optimized by some unconstrained or bounded constrained optimization softwares or sequential quadratic programming (SQP) techniques. No matter what kind of techniques are involved, the merit function always depends on a small parameter ε or large parameter $\rho = \varepsilon^{-1}$. As $\varepsilon \to 0$, the minimizer of a merit function such as a barrier function or the quadratic penalty function, converges to a minimizer of the original problem. By using some exact penalty function such as l_1 penalty function (see [1, 9, 20–22]), the minimizer of the corresponding penalty problem must be a minimizer of the original problem when ε is sufficiently small. There are some nonsmooth penalty functions for nonsmooth optimization problems, such as the exact penalty function using the distance function for the nonsmooth variational inequality problem in Hilbert spaces [18] and the one in [19].

The traditional exact penalty functions [8] are always nonsmooth. When it is used as a merit function to accept a new iterate in an SQP method, it may cause the Maratos effect [13]. On the other hand, a traditional smooth penalty function like the quadratic penalty function cannot be an exact one. So we must compute a sequence of minimization subproblem as $\varepsilon \to 0$. At that time, ill-conditioning may occur when the penalty parameter is too large or small, which also brings difficulty of computation. In [3] and [6], some kinds of augmented Lagrangian penalty functions have been proposed with improved exactness under strong conditions. In [11], exact penalty functions *via* regularized gap function for variational inequalities have also been given. All these functions enjoy some smoothness,

but at the very beginning, to use this smoothness we need second-order or third-order derivative information of the problem function that is difficult to estimate in practice. Besides, all the above kinds of penalty functions (see [2–4, 7, 16] for summary) may be unbounded below even when the constrained problem is bounded, which may make it difficult to locate a minimizer.

In the paper [10], a new penalty function is proposed for the constrained optimization problem. By augmenting the dimension of the program with an additional variable ε that controls the weight of the penalty terms, this new penalty function enjoys properties of smoothness and exactness, and remains bounded below under reasonable conditions. Its important new idea is that the penalty function is considered as a function of variable x and the additional variable ε simultaneously. Under proper assumptions, the minimizer (x^*, ε_*) of the merit function satisfies $\varepsilon_* = 0$, and x^* is a minimizer of the original problem. However, the penalty function given in [10] is not smooth in a small neighborhood of $(x^*, 0)$, where the minimizer of the original constrained problem lies. In this paper, we give a penalty function which enjoys the properties of the penalty function given in [10] and has improved smoothness.

The rest of this paper is organized as follows. In Section 2, a penalty function is introduced for a smooth nonlinear optimization problem with equality constraints and bounded constraints. The smoothness of this penalty function is discussed, as well as other properties, including being bounded below under mild assumptions. Section 3 shows the exactness of our penalty function in the sense that under certain conditions, local minimizer of our penalty function has the form (x^*, ε_*) with $\varepsilon_* = 0$ and x^* is a local minimizer of the original problem, and a converse result holds.

Notation Throughout this paper, we use the Euclidean norm $\|x\| = \sqrt{\sum x_k^2}$. The subvector of x indexed by the indices in J is denoted by x_J. We denote sets of the form

$$[\underline{x}, \overline{x}] := x \in R^n | \{\underline{x} \le x \le \overline{x}\},$$

where the lower bound $\underline{x} \in (R \cup \{-\infty\})^n$ and the upper bound $\overline{x} \in (R \cup \{\infty\})^n$ are vectors containing proper or infinite bounds on the components of x and $[\underline{x}, \overline{x}]$ is referred to an n-dimensional box.

2 New penalty function

We consider the smooth nonlinear optimization problem with equality constraints and bound constraints:

$$(P) \quad \begin{aligned} &\min f(x) \\ &\text{s.t.} \quad x \in [u, v], \quad F(x) = 0, \end{aligned} \tag{2.1}$$

where $[u, v]$ is a box in $\Re^n$ with nonempty interior, $f : D \to \Re$ and $F : D \to \Re^m$ are continuously differentiable in an open set D containing $[u, v]$ and $m < n$. We fix $w \in \Re^m$ and consider the equivalent problem:

$$(\overline{P}) \quad \begin{aligned} &\min f(x) \\ &\text{s.t.} \quad F_j(x) = \varepsilon^\gamma w_j, \quad j = 1, \dots, m, \\ &\qquad x \in [u, v], \varepsilon = 0, \end{aligned} \tag{2.2}$$

where $\gamma > 0$.

Let $\overline{\varepsilon} > 0$ be an upper bound of the parameter ε. Then the corresponding penalty function f_σ on $D \times [0, \overline{\varepsilon}]$ for $(\overline{P})$ is given as follows:

$$
f_\sigma(x, \varepsilon) = \begin{cases} f(x) & \text{if } \varepsilon = \triangle(x, \varepsilon) = 0, \\ f(x) + \frac{1}{\varepsilon^\alpha} \frac{\triangle(x, \varepsilon)}{1 - q\triangle(x, \varepsilon)} + \sigma \varepsilon^\beta & \text{if } \varepsilon > 0, \triangle(x, \varepsilon) < q^{-1}, \\ +\infty & \text{otherwise} \end{cases} \tag{2.3}
$$

with the constraint violation measure

$$
\triangle(x, \varepsilon) := \left\| F(x) - \varepsilon^\gamma w \right\|^2,
$$

where, in addition, $\gamma > \alpha \geq 2\beta \geq 2$ and $q > 0$, are all fixed numbers, $\sigma > 0$ is a penalty parameter and $\| \cdot \|$ is a Euclidean norm, with $\|x\| = \sqrt{x^T x}$ for any vector x.

Obviously, $\varepsilon = \triangle(x, \varepsilon) = 0$ if and only if $\varepsilon = 0$, $F_j(x) = 0$, $j = 1, \ldots, m$. The corresponding penalty problem then reads

$$
(P_\sigma) \quad \begin{aligned} & \min f_\sigma(x, \varepsilon) \\ & \text{s.t.} \quad (x, \varepsilon) \in [u, v] \times [0, \overline{\varepsilon}]. \end{aligned} \tag{2.4}
$$

The main difference between (2.3) and the penalty function given in [10] is that in (2.3), $\beta(\varepsilon) = \varepsilon^\beta$, which does not have the property that $\beta'(\varepsilon) \to +\infty$, as $\varepsilon \to 0^+$.

It is easy to see that $f_\sigma(x, \varepsilon)$ is continuously differentiable with respect to (x, ε) on

$$
D_q := \left\{ (x, \varepsilon) \in D \times (0, \overline{\varepsilon}] \,\middle|\, 0 \leq \triangle(x, \varepsilon) < \frac{1}{q} \right\}.
$$

2.1 Boundedness of the penalty function

If $F(x) = 0$, $(x, \varepsilon) \in D_q$, then

$$
f_\sigma(x, \varepsilon) = f(x) + \frac{\varepsilon^{2\gamma - \alpha} \|w\|^2}{1 - q\varepsilon^{2\gamma} \|w\|^2} + \sigma \varepsilon^\beta \geq f_\sigma(x, 0) = f(x). \tag{2.5}
$$

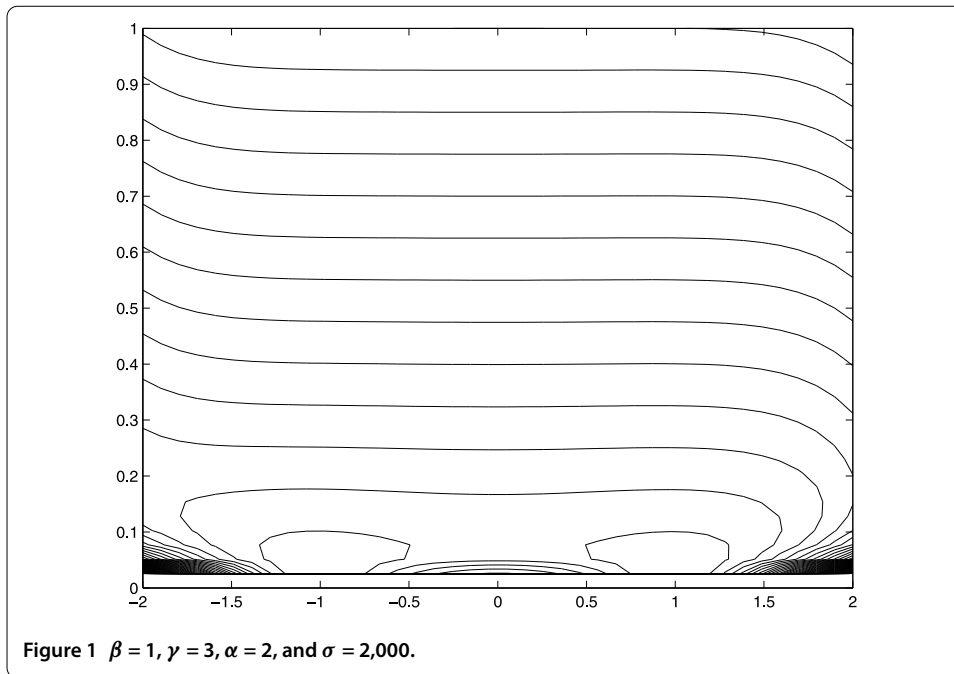Therefore, $f_\sigma(x, \varepsilon)$ is bounded below on the set

$$
D' = \left\{ x \in [u, v] \,\middle|\, \left\| F(x) \right\| \leq q^{-1/2} + \overline{\varepsilon}^\gamma \|w\| \right\},
$$

whenever $f(x)$ is bounded below on the set D'. This is a reasonable condition since it usually holds when f is bounded below on the feasible set, $\overline{\varepsilon}$ is small enough, and q is large enough.

The denominator $1 - q\triangle(x, \varepsilon)$ is included since it forces the level sets of f_σ to remain in the set $\{(x, \varepsilon) \in \Re^{n+1} | \triangle(x, \varepsilon) < q^{-1}\}$, hence in some sense does not go far away from the feasible set of (P).

Now we see a simple example:

$$
\begin{aligned} & \min x^7 \\ & \text{s.t.} \quad x^2 - 1 = 0, \\ & \quad\quad x \in \Re. \end{aligned}
$$

Figure 1 $\beta = 1$, $\gamma = 3$, $\alpha = 2$, and $\sigma = 2{,}000$.

It has a bounded feasible domain, a global minimizer at $x^* = -1$ with $f(x^*) = -1$, and a local minimizer $x = 1$. The traditional quadratic penalty function for this problem

$$P(x) = x^7 + \frac{1}{\varepsilon}\left(x^2 - 1\right)^2,$$

is unbounded below for all penalty parameters $\varepsilon > 0$ since, *e.g.*, $p(x) \to -\infty$ for $x = -s$, $s \to +\infty$. It is also the case for traditional penalty functions, including multiplier penalty functions that use an additional term $+\lambda(x^2 - 1)$. On the other hand, our new penalty function is bounded below. Set $w = 1$, it reads

$$f_\sigma(x, \varepsilon) = \begin{cases} x^7 & \text{if } \varepsilon = r = 0, \\ x^7 + \frac{1}{\varepsilon^\alpha}\frac{r^2}{1-qr^2} + \sigma\varepsilon^\beta & \text{if } \varepsilon > 0, |r| < q^{-1/2}, \\ +\infty & \text{otherwise}, \end{cases}$$

where $r = 1 + \varepsilon^\gamma - x^2$. Since $f_\sigma(x, \varepsilon) = +\infty$, if $|x| \geq \sqrt{q^{-1/2} + 1 + \varepsilon^\gamma}$, it is obvious that our penalty function is bounded below. See Figure 1 for the display of the contour of the penalty function on this example.

3 Exactness of the penalty function

In this section, we show that our penalty function is exact in the sense that under certain conditions, local minimizer of our penalty function has the form (x^*, ε_*) in which $\varepsilon_* = 0$ and x^* is a local minimizer of the original problem and a converse proposition holds.

Firstly, recall the Mangasarian-Fromovitz condition. We say that *the Mangasarian-Fromovitz condition (see [12]) for Problem (P) holds at $x \in [u, v]$ if $F'(x)$ has full rank and there is a vector $p \in \Re^n$ with $F'(x)p = 0$ and*

$$p_i \begin{cases} > 0 & \text{if } x_i = u_i, \\ < 0 & \text{if } x_i = v_i. \end{cases} \tag{3.1}$$

Theorem 3.1 *Assume that the set D' defined in Section 2 is bounded, and the Mangasarian-Fromovitz condition holds at each $x' \in D'$. Let $\gamma > \alpha \geq 2\beta \geq 2$ in (P_σ). If σ is sufficiently large, then there is no Kuhn-Tucker point (x, ε) of (P_σ) with $\varepsilon > 0$.*

Proof Let the Lagrangian function of (P_σ) for $\sigma > 0$ be

$$L(x, \varepsilon, y, z) = f_\sigma(x, \varepsilon) + \sum_{i=1}^{n} y_i(u_i - x_i) + \sum_{i=1}^{n} z_i(x_i - v_i) + y_{n+1}(-\varepsilon) + z_{n+1}(\varepsilon - \overline{\varepsilon}),$$

where $y_i, z_i \in \Re$, $i = 1, \ldots, n+1$ are the Lagrangian multipliers. If (x, ε) is a Kuhn-Tucker point of (P_σ) with $\varepsilon > 0$, then there exist vectors $y, z \in \Re^{n+1}$ such that

$$\nabla_{(x,\varepsilon)} L(x, \varepsilon, y, z) = 0,$$

$$u_i - x_i \leq 0, \qquad x_i - v_i \leq 0, \quad i = 1, \ldots, n,$$

$$-\varepsilon < 0, \qquad \varepsilon - \overline{\varepsilon} \leq 0,$$

$$y_i(u_i - x_i) = 0, \qquad z_i(v_i - x_i) = 0, \qquad z_{n+1}(\varepsilon - \overline{\varepsilon}) = 0, \qquad y_{n+1}\varepsilon = 0,$$

then

$$\nabla_{(x,\varepsilon)} f_\sigma(x, \varepsilon) = y - z,$$

and

$$\inf(y_i, x_i - u_i) = \inf(z_i, v_i - x_i) = 0, \quad i = 1, \ldots, n,$$

$$y_{n+1} = \inf(z_{n+1}, \overline{\varepsilon} - \varepsilon) = 0,$$

where $\nabla_{(x,\varepsilon)} f_\sigma(x, \varepsilon)$ is the gradient of f_σ with respect to (x, ε). The assertion of the theorem is proved by contradiction. □

Assume that there exists a sequence $\{(x^k, \varepsilon_k, \sigma_k)\}$ with $\varepsilon_k > 0$ for all k, $\sigma_k \to +\infty$ as $k \to \infty$, where (x^k, ε_k) is a Kuhn-Tucker point of (P_{σ_k}). We use the abbreviation $\triangle_k := \triangle(x^k, \varepsilon_k)$. The point x^k satisfies

$$\left\| F(x^k) \right\| \leq \triangle_k^{1/2} + \varepsilon_k^\gamma \|w\| \leq q^{-1/2} + \overline{\varepsilon}^\gamma \|w\|,$$

hence, $x^k \in D' = \{x \in [u, v] \mid \|F(x)\| \leq q^{-1/2} + \overline{\varepsilon}^\gamma \|w\|\}$. Since D' is closed and bounded, we may restrict ourselves to a subsequence if necessary and assume that

$$\lim_{k \to \infty} \varepsilon_k = \varepsilon_* \in [0, \overline{\varepsilon}] \quad \text{and} \quad \lim_{k \to \infty} x^k = x^* \in D'.$$

The condition $\frac{\partial}{\partial \varepsilon} f_\sigma(x^k, \varepsilon_k) \leq 0$ yields

$$\alpha q \triangle_k^2 + (2\gamma - \alpha)\varepsilon_k^{2\gamma} \|w\|^2 + 2(\alpha - \gamma)\varepsilon_k^\gamma \sum_{j=1}^{m} F_j(x^k) w_j$$

$$+ \beta \varepsilon_k^{\alpha + \beta} (1 - q\triangle_k)^2 \sigma_k \leq \alpha \sum_{j=1}^{m} F_j^2(x^k) \qquad (3.2)$$

with equality in the case $\varepsilon_k \neq \overline{\varepsilon}$. When $\sigma_k \to \infty$ and because $\alpha \sum_{j=1}^{m} F_j^2(x^k)$ - the right side of (3.2) is a finite number, we have

$$\lim_{k \to \infty} \varepsilon_k = \varepsilon_* = 0 \quad \text{or} \quad \lim_{k \to \infty} \triangle(x^k, \varepsilon_k) = \triangle^* = q^{-1}, \tag{3.3}$$

where $\triangle^* = \triangle(x^*, \varepsilon_*)$. On the other side, the derivative the penalty function f_σ with respect to x is given as

$$\frac{\partial}{\partial x_i} f_\sigma(x^k, \varepsilon_k) = \frac{\partial}{\partial x_i} f(x^k) + \frac{1}{\varepsilon_k^\alpha} \frac{\frac{\partial}{\partial x_i} \triangle(x^k, \varepsilon_k)}{1 - q\triangle(x^k, \varepsilon_k)} + \frac{1}{\varepsilon_k^\alpha} \frac{q\triangle(x^k, \varepsilon_k)\frac{\partial}{\partial x_i}\triangle(x^k, \varepsilon_k)}{(1 - q\triangle(x^k, \varepsilon_k))^2}$$

$$= \frac{\partial}{\partial x_i} f(x^k) + \frac{1}{\varepsilon_k^\alpha} \frac{\frac{\partial}{\partial x_i}\triangle(x^k, \varepsilon_k)}{(1 - q\triangle(x^k, \varepsilon_k))^2} \begin{cases} \geq 0 & \text{if } x_i^k = u_i, \\ = 0 & \text{if } u_i < x_i^k < v_i, \\ \leq 0 & \text{if } x_i^k = v_i \end{cases} \tag{3.4}$$

(3.4) is equivalent to

$$\varepsilon_k^\alpha (1 - q\triangle(x^k, \varepsilon_k))^2 \frac{\partial}{\partial x_i} f(x^k) + \frac{\partial}{\partial x_i} \triangle(x^k, \varepsilon_k)$$

$$= \varepsilon_k^\alpha (1 - q\triangle(x^k, \varepsilon_k))^2 \frac{\partial}{\partial x_i} f(x^k) + 2(F'(x^k)^T (F(x^k) - \varepsilon_k^\gamma w))_i$$

$$\begin{cases} \geq 0 & \text{if } x_i^k = u_i, \\ = 0 & \text{if } u_i < x_i^k < v_i, \\ \leq 0 & \text{if } x_i^k = v_i. \end{cases} \tag{3.5}$$

Let $k \to +\infty$, since $\varepsilon_* = 0$ or $\triangle^* = q^{-1}$, we have

$$(F'(x^*)^T (F(x^*) - \varepsilon_*^\gamma w))_i \begin{cases} \geq 0 & \text{if } x_i^* = u_i, \\ = 0 & \text{if } u_i < x_i^* < v_i, \\ \leq 0 & \text{if } x_i^* = v_i. \end{cases} \tag{3.6}$$

Because $x^* \in D'$, thus Mangasarian-Fromovitz condition holds at x^* and there exists some vector $p \in \Re^n$ such that $F'(x^*)p = 0$, where p satisfies (3.1). Let $I_1 := \{i | x_i^* = u_i\}$, $I_2 := \{i | x_i^* = v_i\}$ and $\overline{\triangle^*} := F(x^*) - \varepsilon_*^\gamma w$. Then by (3.6), we have

$$0 = (F'(x^*)p)^T \overline{\triangle^*} = \sum_{i \in I_1} p_i (F'(x^*)^T \overline{\triangle^*})_i + \sum_{i \in I_2} p_i (F'(x^*)^T \overline{\triangle^*})_i. \tag{3.7}$$

(3.7) and the Mangasarian-Fromovitz condition (3.1) imply $(F'(x^*)^T \overline{\triangle^*})_i = 0$ for $i \in I_1 \cup I_2$. Thus, $F'(x^*)^T \overline{\triangle^*} = 0$. Now the fact that $F'(x^*)$ has full rank yields $\overline{\triangle^*} = 0$, *i.e.*,

$$F(x^*) - \varepsilon_*^\gamma w = 0,$$

and by $\overline{\triangle^*} = \lim_{k \to \infty} (F(x^k) - \varepsilon_k^\gamma w) = 0$, it follows that

$$\lim_{k \to \infty} \triangle(x^k, \varepsilon_k) = \lim_{k \to \infty} \|F(x^k) - \varepsilon_k^\gamma w\|^2 = \triangle^* = 0.$$

By (3.3), we obtain

$$\lim_{k \to \infty} \varepsilon_k = \varepsilon_* = 0.$$

Thus, $\lim_{k \to \infty} F(x^k) = F(x^*) = 0$.

Furthermore, by (3.2), it holds that

$$\frac{q}{\varepsilon_k^{\alpha+\beta}} \triangle^2 (x^k, \varepsilon_k) + \frac{2\frac{\gamma}{\alpha} - 1}{\varepsilon_k^{\alpha+\beta-2\gamma}} \|w\|^2 + 2\left(1 - \frac{\gamma}{\alpha}\right) \varepsilon_k^{\gamma-(\alpha+\beta)} \sum_{i=1}^{m} F_i(x^k) w_i$$
$$+ \frac{\beta}{\alpha} \left(1 - q \triangle (x^k, \varepsilon_k)\right)^2 \sigma_k \le \frac{\sum_{i=1}^{m} F_i^2(x^k)}{\varepsilon_k^{\alpha+\beta}}.$$

Let $k \to \infty$, the last term on the left-hand side tend to $+\infty$. Thus, the vectors $y^k = \frac{F(x^k)}{\varepsilon_k^{\frac{\alpha+\beta}{2}}}$ satisfies $\|y^k\| \to +\infty$. The vectors $z^k = \frac{y^k}{\|y^k\|}$ have norm 1, and (3.5) implies that the numbers μ_i^k $(i = 1, \ldots, n)$, defined by

$$\mu_i^k = \frac{1}{\|y^k\|} \frac{\partial}{\partial x_i} f(x^k) + \frac{1}{\|y^k\|} \frac{2}{\varepsilon_k^{\alpha}(1 - q\triangle(x^k, \varepsilon_k))^2} \left(F'(x^k)^T (F(x_k) - \varepsilon_k^{\gamma} w)\right)_i$$
$$= \frac{1}{\|y^k\|} \frac{\partial}{\partial x_i} f(x^k) + \frac{2}{\varepsilon_k^{\frac{\alpha-\beta}{2}}(1 - q\triangle(x^k, \varepsilon_k))^2} \left(F'(x^k)^T \left(z^k - \frac{\varepsilon_k^{\gamma-\frac{\alpha+\beta}{2}} w}{\|y^k\|}\right)\right)_i,$$

satisfy

$$\mu_i^k \begin{cases} \ge 0 & \text{if } x_i^k = u_i, \\ = 0 & \text{if } u_i < x_i^k < v_i, \\ \le 0 & \text{if } x_i^k = v_i. \end{cases}$$

If we pick a convergent subsequence z^{n_k} with the limit z^* and pass to the limit we obtain

$$\left(F'(x^*) z^*\right)_i \begin{cases} \ge 0 & \text{if } x_i^* = u_i, \\ = 0 & \text{if } u_i < x_i^* < v_i, \\ \le 0 & \text{if } x_i^* = v_i. \end{cases}$$

Now similarly as above, it yields $z^* = 0$, which is a contradiction with $\|z^*\| = 1$. Thus such a sequence $\{(x^k, \varepsilon_k, \sigma_k)\}$ cannot exist, and for sufficiently large $\sigma > 0$, all Kuhn-Tucker points of (P_σ) are of the form $(x, 0)$.

Theorem 3.2 *Assume that (x^*, ε_*) is a local minimizer of minimizer of (P_σ) with finite $f_\sigma(x^*, \varepsilon_*)$, where $\sigma > 0$ is sufficiently large. If the hypotheses of Theorem 3.1 are fulfilled, then x^* is a local minimizer of (P).*

Proof Now let (x^*, ε_*) be a local minimizer of (P_σ) with finite $f_\sigma(x^*, \varepsilon_*)$ and $\sigma > 0$ is sufficiently large. If $\varepsilon_* > 0$, then (x^*, ε_*) must be a Kuhn-Tucker point of (P_σ), which is a contradiction with Theorem 3.1. Therefore, $\varepsilon_* = 0$, and since $f_\sigma(x^*, \varepsilon_*)$ is finite, $\triangle(x^*, \varepsilon_*) = 0$. It

implies that $F(x^*) = 0$, and by (2.5) there is a neighborhood $N(x^*)$ of x^* where $f(x) \geq f(x^*)$ for feasible x. Therefore, x^* is a local minimizer of (P). □

We now show a converse result of Theorem 3.2, which will use the following lemmas.

Lemma 3.1 *Suppose $F(x^*) = 0$, and $F'(x^*)$ has full rank. Then there exist a neighborhood $N_0(x^*)$ of x^* and a constant $\kappa_0 > 0$ such that for each $x \in N_0(x^*)$, and each subset J of $\{1, 2, \ldots, m\}$, there exists a vector $y = y(x) \in N_0(x^*)$ with $F_i(y) = 0$, for $i \in J$ and $F_i(y) = F_i(x)$, $i \in K = \{1, 2, \ldots, m\} \setminus J$, such that*

$$\|x - y\| \leq \kappa_0 \|F_J(x)\|.$$

Proof Since $F(x^*) = 0$ and $F'(x^*)$ has full rank, there exists a matrix $B \in \Re^{(n-m) \times n}$ such that the augmented matrix $\binom{F'(x^*)}{B}$ is nonsingular. By the continuity of $F'(\cdot)$ at x^*, there exists a neighborhood $N_1(x^*) \subset D$ of x^* such that $\binom{F'(x)}{B}$ is nonsingular, for any $x \in N_1(x^*)$. Take for $\mathcal{A}$ the closed convex hull of $\{F'(x) | x \in N_1(x^*)\}$, then for all $A \in \mathcal{A}$, the matrix $\binom{A}{B}$ is nonsingular. We now show that for any $x, y \in N_1(x^*)$, there exists a matrix $A \in \mathcal{A}$ such that

$$F(x) - F(y) = A(x - y). \tag{3.8}$$

In fact, given $x, y \in N_1(x^*)$, it follows from the mean value theorem that

$$F(x) - F(y) = \int_0^1 F'\big(y + s(x - y)\big)(x - y)\,ds$$

$$= A_{x,y}(x - y),$$

where $A_{x,y} = \int_0^1 F'(y + s(x - y))\,ds \in \mathcal{A}$, so (3.8) holds. Set the mapping $H(z) := \binom{F(z)}{B(z - x^*)}$, for $z \in N_1(x^*)$. By the proof in [10, Theorem 4.5], we have that there exists a neighborhood $N_0(x^*) \subset N_1(x^*)$ of x^* such that for each $x \in N_0(x^*)$, and each subset J of $\{1, 2, \ldots, m\}$, there exists a vector $y = y(x) \in N_0(x^*)$ with

$$H(y) = \begin{pmatrix} F(y) \\ B(y - x^*) \end{pmatrix} = \begin{pmatrix} 0 \\ F_K(x) \\ B(x - x^*) \end{pmatrix},$$

so $F_i(y) = 0$, for $i \in J$ and $F_i(y) = F_i(x)$, $i \in K$.

For $x, y \in N_0(x^*)$, we have

$$H(x) - H(y) = \begin{pmatrix} A \\ B \end{pmatrix}(x - y),$$

for some $A \in \mathcal{A}$. On the other side, we have

$$\|H(x) - H(y)\| = \|F_J(x)\|. \tag{3.9}$$

Therefore, combining (3.8) with (3.9), we have

$$\|x - y\| \leq \left\| \begin{pmatrix} A \\ B \end{pmatrix}^{-1} \right\| \|F_J(x)\| \leq \kappa_0 \|F_J(x)\|,$$

where

$$\kappa_0 := \left\| \sup_{A \in \mathcal{A}} \begin{pmatrix} A \\ B \end{pmatrix}^{-1} \right\| < +\infty,$$

and this complete the proof. □

Lemma 3.2 *Assume that x^* is a local minimizer of problem (2.1) with $u_i \leq x_i^* \leq v_i$, $i \in \{1, \ldots, p\}$, where $m \leq p \leq n$. Suppose that $F'(x^*)$ has full rank. Then there exists a constant $\kappa_1 > 0$ such that*

$$f(x) \geq f(x^*) - \kappa_1 \|F(x)\|, \quad \text{for all } x \in N_0(x^*),$$

where $N_0(x^*)$ is defined in Lemma 3.1.

Proof From Lemma 3.1, let $N_0(x^*)$ and κ_0 be the one in Lemma 3.1. Let $x \in N_0(x^*)$, then by Lemma 3.1 with $J = \{1, \ldots, m\}$, there exists an $y = y(x)$ with $F(y) = 0$ and $y_i = x_i^*$, $i = m + 1, \ldots, n$ such that

$$\|x - y\| \leq \kappa_0 \|F(x)\|. \tag{3.10}$$

So y is a feasible point of problem (2.1), and $f(y) \geq f(x^*)$. Since f is continuously differentiable, for any $x, y \in N_0(x^*)$, there exists a vector $\xi \in \Re^n$ such that

$$f(x) - f(y) = \nabla f(\xi)^T (x - y),$$

where ξ lies in the segment between x and y. Set $L := \sup_{z \in N(x^*)} \|\nabla f(z)\|$, we have

$$\begin{aligned}
\left| f(x) - f(y) \right| &\leq \left\| \nabla f(\xi) \right\| \|x - y\| \\
&\leq L \|x - y\| \\
&\leq L \kappa_0 \|F(x)\|,
\end{aligned}$$

where the last inequality holds from (3.10).

Let $\kappa_1 = L \kappa_0$, then

$$f(x) = f(x) - f(y) + f(y) \geq f(x^*) - \kappa_1 \|F(x)\|,$$

which complete the proof. □

Theorem 3.3 *If x^* is a local minimizer of problem (P) with $u_i \leq x_i^* \leq v_i$, $i \in \{1, \ldots, p\}$, where $m \leq p \leq n$, and $F'(x^*)$ has full rank, then for sufficiently large $\sigma > 0$, there are a neighborhood $N(x^*)$ of x^* and a $\varepsilon' \in (0, \overline{\varepsilon}]$ such that*

$$f_\sigma(x, \varepsilon) > f_\sigma(x^*, 0) = f(x^*), \quad \text{for all } (x, \varepsilon) \in N(x^*) \times (0, \varepsilon'].$$

In particular, $(x^*, 0)$ is a local minimizer of f_σ.

Proof Let $N(x^*) \subset N_0(x^*)$ is a neighborhood of x^* such that

$$\sup_{x \in N(x^*)} \left(f(x^*) - f(x) \right) < 1, \tag{3.11}$$

where $N_0(x^*)$ is defined in Lemma 3.1.

For $(x, \varepsilon) \in N(x^*) \times (0, \varepsilon']$, where $\varepsilon' \in (0, \overline{\varepsilon}]$ and $\varepsilon' \leq 1$, we distinguish two cases.

Case 1. $\triangle(x, \varepsilon) = \|F(x) - \varepsilon^\gamma w\|^2 \geq \varepsilon^\alpha$. For this case, we have

$$\begin{aligned}
f_\sigma(x, \varepsilon) &\geq f(x) + 1 + \sigma \varepsilon^\beta \\
&\geq f(x^*) + \sigma \varepsilon^\beta \\
&> f(x^*) = f_\sigma(x^*, 0),
\end{aligned}$$

where the second inequality is by (3.11).

Case 2. $\triangle(x, \varepsilon) < \varepsilon^\alpha$. Then $\|F(x)\| \leq \triangle^{\frac{1}{2}} + \varepsilon^\gamma \|w\| \leq \varepsilon^{\frac{\alpha}{2}} + \varepsilon^\gamma \|w\|$, and

$$\begin{aligned}
f_\sigma(x, \varepsilon) &\geq f(x) + \sigma \varepsilon^\beta \\
&\geq f(x^*) - \kappa_1 \|F(x)\| + \sigma \varepsilon^\beta \\
&\geq f(x^*) - \kappa_1 \left(\varepsilon^{\frac{\alpha}{2}} + \varepsilon^\gamma \|w\| \right) + \sigma \varepsilon^\beta \\
&\geq f(x^*) + \left(\sigma - \kappa_1 \left(1 + \varepsilon^{\gamma - \frac{\alpha}{2}} \right) \|w\| \right) \varepsilon^{\frac{\alpha}{2}}. \tag{3.12}
\end{aligned}$$

The last inequality holds since $\beta \leq \frac{\alpha}{2}$.

Let $\sigma > \kappa_1 (1 + \varepsilon^{\gamma - \frac{\alpha}{2}}) \|w\|$, we get

$$f_\sigma(x, \varepsilon) \geq f(x^*) = f_\sigma(x^*, 0).$$

From Case 1 and Case 2, we obtain the conclusion. □

4 Conclusion remarks

In this paper, a modified exact penalty function for equality constrained nonlinear programming problem is constructed by augmenting a new variable that controls the constraint violation. This function enjoys smoothness, and with very mild conditions it is proved to be an exact penalty function.

Since in practice, a lot of applied problems are nonsmooth, it is a meaningful work to extend the results in this paper to the nonsmooth case. By using the limiting subgradients that is presented in two books written by Mordukhovich [14, 15], as well as Clarke's generalized gradients in [5], we can extend the penalty function with the mentioned good

properties to nonsmooth optimization problems, just as that has been done in [17–19]. That will be our future research direction.

References
1. Antczak, T: Exact penalty functions method for mathematical programming problems involving invex functions. Eur. J. Oper. Res. **198**, 29-36 (2009)
2. Bazaraa, MS, Sherali, HD, Shetty, CM: Nonlinear Optimization Theory and Algorithms, 2nd edn. Wiley, New York (1993)
3. Bertsekas, DP: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York (1982)
4. Boukari, D, Fiacco, AV: Survey of penalty, exact-penalty and multiplier methods from 1968 to 1993. Optimization **32**, 301-334 (1995)
5. Clarke, FH: Optimization and Nonsmooth Analysis. Wiley-Interscience, New York (1983)
6. Di Pillo, G: Exact penalty methods. In: Spedicato, E (ed.) Algorithms for Continuous Optimization, pp. 209-253. Kluwer Academic, Dordrecht (1994)
7. Fletcher, R: Practical Methods of Optimization, 2nd edn. Wiley, New York (1987)
8. Han, SP, Mangasarian, OL: Exact penalty functions in nonlinear programming. Math. Program. **17**, 251-269 (1979)
9. Hoheisel, T, Kanzow, C, Outrata, J: Exact penalty results for mathematical programs with vanishing constraints. Nonlinear Anal. **72**, 2514-2526 (2010)
10. Huyer, W, Neumaier, A: A new exact penalty function. SIAM J. Optim. **13**, 1141-1158 (2003)
11. Li, W, Peng, J: Exact penalty functions for constrained minimization problems via regularized gap function for variational inequality. J. Glob. Optim. **37**, 85-94 (2007)
12. Mangasarian, OL, Fromovitz, S: The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. J. Math. Anal. Appl. **17**, 37-47 (1967)
13. Maratos, N: Exact penalty function algorithms for finite dimensional and control optimization problems. PhD thesis, University of London (1978)
14. Mordukhovich, BS: Variations Analysis and Generalized Differentiation, I: Basic Theory. Grundlehren Series (Fundamental Principles of Mathematical Sciences), vol. 330. Springer, Berlin (2006)
15. Mordukhovich, BS: Variations Analysis and Generalized Differentiation, II: Applications. Grundlehren Series (Fundamental Principles of Mathematical Sciences), vol. 331. Springer, Berlin (2006)
16. Nocedal, J, Wright, SJ: Numerical Optimization. Springer, New York (1999)
17. Soleimani-damaneh, M: The gap function for optimization problems in Banach spaces. Nonlinear Anal. **69**, 716-723 (2008)
18. Soleimani-damaneh, M: Penalization for variational inequalities. Appl. Math. Lett. **22**, 347-350 (2009)
19. Soleimani-damaneh, M: Nonsmooth optimization using Mordukhovich's subdifferential. SIAM J. Control Optim. **48**, 3403-3432 (2010)
20. Zangwill, WI: Nonlinear programming via penalty function. Manag. Sci. **13**, 344-358 (1967)
21. Zaslavski, AJ: A sufficient condition for exact penalty in constrained optimization. SIAM J. Optim. **16**, 250-262 (2005)
22. Zaslavski, AJ: Stability of exact penalty for nonconvex inequality-constrained minimization problems. Taiwan. J. Math. **14**, 1-19 (2010)