A Self-adaptive Penalty Approach for Nonlinear Programming

Pu-yan Nie

1 Introduction

We consider nonlinear programming problem with equality constraints, which is stated as follows:

$$\text{minimize } f(x) \text{ subject to } c(x) = 0,$$

where $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are all continuously differentiable. $m$ and $n$ are all positive integers. For constrained optimization problems, researchers try to treat them as unconstrained problems by penalty function techniques. Penalty function methods, including interior and exterior penalty strategies for solving mathematical programming problems, are thus focused on several scores years ago and play extremely important roles in the optimization community. In penalty techniques, a sequence of unconstrained optimization problems are handled for different values of certain penalty parameter to ensure convergence. The early barrier function methods are proposed for equality constrained problems. They are extended to nonlinear programming with inequality constraints.

About penalty methods, [1] is an excellent monograph for nonlinear programming. Penalty methods are used extensively till now and there is abundant literature about this topic. Some other penalty approaches are recently put forward, for example,
two parameters penalty method [3]. In [5], a new exact penalty technique is put forward. Because penalty strategies may issue in the ill-condition of the problem, penalty approaches are sometime neglected in the history. Some techniques are therefore presented to overcome these bad cases [1]. With the appearance of interior point methods, penalty techniques are reattached importance too. They are extensively utilized, see in [6] for mathematical programming with equilibrium (MPEC) problems with the cited references and in [8] for bilevel programming problems, respectively. In [4], penalty approach for multiple objective function is considered. In [7], exact penalty techniques are employed for semi-infinite min-max problems.

There exist diverse ways to avoid ill-condition of the problem issued from penalty methods, such as augmented multiplier approaches [1], penalty parameter control technique [2], and so forth. The previous works motivate us to adjust both penalty term and penalty parameters according to necessity. Furthermore, in diverse optimization techniques, constraints converging to optimal point may be inconsistent with that of the objective function. We hope to take step to avoid it, which also motivates the new approach. Some new techniques, to adjust penalty parameter and penalty term in each step, will be given in the penalty methods, which is called self-adaptive technique.

Definition 1.1. A penalty approach is a self-adaptive penalty method if and only if the penalty parameter and penalty term are changed with the current information at each step.

There are some advantages for self-adaptive strategies.

(i) At the beginning of an algorithm, one does not know the best penalty term. In a self-adaptive approach, we can make full use of the current information to obtain a “better” penalty function.

(ii) When penalty function changes, threshold value may correspondingly change. The early approaches ignore the current information about penalty parameter. We hope to take advantage of that in this paper.

(iii) In many algorithms, constraints converge to feasible point with the rate out of step with objective function, which will result in the inconsistency of convergence for all variables. The method in this paper may avoid it to a certain degree because of the self-adaptive technique.

(iv) The early penalty approaches are exactly the special cases of self-adaptive penalty family, in which just penalty parameter is adjusted at each step.

This work is organized as follows: a new penalty method for nonlinear programming problems is given in the next section. The properties of the corresponding algorithm are analyzed in Section 3. Finally, some remarks are presented.
2 Self-adaptive penalty methods for nonlinear programming

To tackle (1.1), in the early penalty methods, a sequence of minimization problems, which have the following form, is required to be solved:

$$\min_x f(x) + \alpha_k P(c(x)), \quad (2.1)$$

where $\alpha_k \to \infty$. In practice, the result is pessimistic if the penalty function is not good enough. Moreover, it is very difficult to find a suitable penalty function initially.

A new approach is proposed in this work. To obtain the solution of (1.1), we minimize recursive problems of the following forms:

$$\min_x Q_k(x, \alpha_k) = f(x) + \alpha_k P_k(c(x)). \quad (2.2)$$

In the new approach, penalty parameter and penalty term are changed at each step according some current (and past) information. Accordingly, the new technique is called self-adaptive penalty method. We hope that the self-adaptive penalty approach family not only includes the existing penalty methods, but also explores some new penalty methods. Firstly, we give a choice of self-adaptive penalty strategy. Self-adaptive penalty algorithm is formally presented as follows.

Algorithm 2.1. Step (1). Give the initial value to $u^{(0)}, v^{(0)}, \alpha^{(0)} > 1.0, \bar{\gamma} > 10, \gamma^{(0)}$. Let $P_0(c(x_0)) = \sum_{i=1}^m c_i^{(0)}(x_0), k := 0$.

Step (2). If the optimal solution is obtained, then stop.

Step (3). Solve problem (2.2) to obtain $x^{(k)}$. Set $k := k + 1$ and give $\alpha^{(k)}$ and $\gamma^{(k)}$.

If $\gamma^{(k)} \leq \bar{\gamma}$, let $P_k(c(x_k)) := \sum_{i=1}^m c_i^{(k)}(x_k)$ according to $x^{(k-2)}, x^{(k-1)}, \alpha^{(k-1)}$, and $P_{k-1}$.

Otherwise, find $\bar{\gamma}_k$ and let $P_k(c(x_k)) := \sum_{i=1}^m \exp (|c_i^{(k)}(x_k)|)$. Go to step (2).

A type of the choice of $P_k$ is now presented. In the algorithm, let $\alpha_k := \alpha^{(0)}$ and $\bar{\alpha}_k = \bar{\alpha}_{k-1}$ if $|\gamma^{(k)} - \gamma^{(k-1)}| \geq \delta_0 > 0$, where $\delta_0$ is a constant. $\alpha_k := \bar{\alpha}_{k-1}$ if $|\gamma^{(k)} - \gamma^{(k-1)}| < \delta_0$.

We employ the following formulation throughout:

$$\gamma^{(k)} = \ln \left( \frac{|p^{k-1} - p^{k-2}|}{|f(x^{(k-1)}) - f(x^{(k-2)})|} \right),$$

$$\bar{\gamma}_k = \ln \left( \frac{|p^{k-1} - p^{k-2}|}{|f(x^{(k-1)}) - f(x^{(k-2)})|} \right). \quad (2.3)$$

We point out that the choice of $\gamma^{(k)}$ and $\bar{\gamma}_k$ is not unique. The following analysis is all based on (2.3).
3 Properties of self-adaptive penalty methods

We consider the properties of self-adaptive penalty methods in the similar way to other penalty approaches. If $\alpha_k \to \infty$, the following results are met.

Theorem 3.1. Assume that $f$ and $c_i$ are all continuous for $i = 1, 2, \ldots, m$. Let $\mathcal{F} = \{x | c(x) = 0\}$. For $k = 1, 2, \ldots$, let $x^{(k)}$ be a global minimum of problem (2.2), where $\alpha_k \to \infty$. Then every accumulation point of the sequence $\{x^{(k)}\}$ is a global minimum of (1.1).

Proof. Assume $\bar{x}$ to be a limit point of $\{x^{(k)}\}$. According to the definition of $x^{(k)}$, we have

$$Q_k(x^{(k)}, \alpha_k) \leq Q_k(x, \alpha_k). \quad (3.1)$$

Let $f^*$ be the optimal value of (1.1). Namely,

$$f^* = \inf_{x \in \mathcal{F}} f(x) = \inf_{x \in \mathcal{F}} Q_k(x, \alpha_k). \quad (3.2)$$

Considering the right-hand side of (3.1) over $\mathcal{F}$ and taking the infimum of the right-hand side of (3.1), we obtain

$$Q_k(x^{(k)}, \alpha_k) \leq f^*. \quad (3.3)$$

Taking the superior in the relation above and by using the continuity of $f$ and $c_i$, we have

$$f(\bar{x}) + \limsup_{k \to \infty} \alpha_k P_k(c(x^{(k)})) \leq f^*. \quad (3.4)$$

By virtue of $P_k(c(x^{(k)})) \geq 0$ and $\alpha_k \to \infty$, we have $P_k(c(x^{(k)})) \to 0$ and

$$\|c(\bar{x})\|^2 = 0, \quad (3.5)$$

for otherwise the limit superior in the left-hand side of (3.2) will equal $+\infty$. Since $\mathcal{F}$ is a closed set, we also obtain that $\bar{x} \in \mathcal{F}$. Hence $\bar{x}$ is feasible and

$$f^* \leq f(\bar{x}). \quad (3.6)$$
Combining (3.2)–(3.6), we obtain

\[ f^* + \lim_{k \to \infty} \sup \alpha_k P_k(c(x^{(k)})) \leq f(\bar{x}) + \lim_{k \to \infty} \sup \alpha_k P_k(c(x^{(k)})) \leq f^*. \tag{3.7} \]

We thus have

\[ \lim_{k \to \infty} \alpha_k \|c(x^{(k)})\|^2 = 0, \quad f(\bar{x}) = f^*, \tag{3.8} \]

which is the result and the proof is complete.

There are some interesting properties about \( P \) and \( f \).

**Corollary 3.2.** If \( \alpha_k \to \infty \), then

1. \( \{Q_k(x^{(k)}, \alpha_k)\} \) is nondecreasing,
2. \( \|P_k(c(x^{(k)}))\| \) is nonincreasing,
3. \( f^{(k)} \) is nondecreasing.

**Proof.** Let \( \alpha_k < \alpha_l \). Then from the definition of \( x^{(k)} \) and (2.2) we have

\[ Q_k(x^{(k)}, \alpha_k) \leq Q_l(x^{(l)}, \alpha_k) \leq Q_l(x^{(l)}, \alpha_l) \leq Q_l(x^{(k)}, \alpha_l). \tag{3.9} \]

The first two inequalities give case (1). From the first term and the last term the following relation holds:

\[ (\alpha_l - \alpha_k)(P_k(c(x^{(k)})) - P_l(c(x^{(l)}))) \geq 0. \tag{3.10} \]

Case (2) is accordingly obtained. The first inequality yields (3). The result therefore holds and the proof is complete.

We assume the global minimum of (2.2) in Theorem 3.1, which is difficult to obtain. We hence consider the local minimum of (2.2). We hope that the limit point of the local minimum of (2.2) has good properties. Similarly, we have the following results.

**Theorem 3.3.** Assume that \( f \) and \( c_i \) are all continuous for \( i = 1, 2, \ldots, m \). \( F_l \) is a closed set. \( \alpha_k < \alpha_{k+1} \) and \( \alpha_k \to \infty \). Let \( X^* \) be an isolated set of local minima of problem (1.1), that is, for each point in \( X^* \) and some \( \epsilon > 0 \), the set

\[ X^*_\epsilon = \{x \mid \|x - x^*\| \leq \epsilon, \text{ for some } x^* \in X^*\} \tag{3.11} \]

contains no local minima of point (1.1) other than the points of \( X^* \). Then there exists a sequence \( \{x^{(k)}\} \) converging to a point \( x^* \in X^* \) such that \( x^{(k)} \) is a local minimum of problem
(2.2). Furthermore, if \( X^* \) consists of a single point \( x^* \), there exist a sequence \( \{x^{(k)}\} \) and an integer \( \tilde{k} \geq 0 \) such that \( x^{(k)} \to x^* \) and \( x^{(k)} \) is a local minimum of problem (2.2) for \( k \geq \tilde{k} \).

\[ \square \]

**Proof.** Using the strategy in Theorem 3.1, we define

\[ X_{\epsilon}^* = \{ x \mid \| x - x^* \| \leq \epsilon, \text{ for some } x^* \in X^* \}, \tag{3.12} \]

where \( 0 < \epsilon < \epsilon \). The compactness of \( X_{\epsilon}^* \) implies compactness of \( X_{\epsilon}^* \). And the problem

\[ \text{minimize } Q_k(x, \alpha_k) = f(x) + \alpha_k P_k(c(x)) \text{ subject to } x \in X_{\bar{\epsilon}} \]

has a global minimum \( x^{(k)} \) by Weierstrass' theorem. With the technique similar to Theorem 3.1, the result holds. The details of the proof are omitted. \[ \square \]

All the results above are obtained without differentiation of Kuhn-Tucker regularity assumptions. We now analyze at another angle. In doing this the vector

\[ (\lambda^{(k)}) = \alpha_k \gamma^{(k)} (c_1^{\gamma^{(k)}-1}(x^{(k)}), c_2^{\gamma^{(k)}-1}(x^{(k)}), \ldots, c_m^{\gamma^{(k)}-1}(x^{(k)}))^T \tag{3.14} \]

is defined, which can be regarded as an estimate of Lagrangian multipliers if polynomial penalty term is employed. If exponent penalty function is employed, we have

\[ (\lambda^{(k)}) = \alpha_k \gamma_k (c_1^{\gamma_k-1}(x^{(k)}), c_2^{\gamma_k-1}(x^{(k)}), \ldots, c_m^{\gamma_k-1}(x^{(k)})) \]

Assume that the other Lagrangian multipliers to inequality constraints are obtained accurately. We consider the Lagrangian function of (2.2):

\[ L_k(x^{(k)}, \lambda^{(k)}) = f(x) + \sum_{i=1}^{m} \lambda_i^{(k)} c_i(x^{(k)}). \tag{3.16} \]

And (2.2) terminates at a point satisfying

\[ \| \nabla_x L_k(x^{(k)}, \lambda^{(k)}) \| \leq \epsilon_k, \tag{3.17} \]
Theorem 3.4. Assume that \( f, c_i \in C^1 \). For \( k = 0, 1, \ldots \), let \( x^{(k)} \) satisfy
\[
\| \nabla_x L_k(x^{(k)}, \lambda^{(k)}) \| \leq \epsilon_k, \tag{3.18}
\]
where \( \lambda^{(k)} \) is bounded, \( 0 < \alpha_k < \alpha_{k+1}, \alpha_k \to \infty \), and \( 0 \leq \epsilon_k \), for all \( k \). Assume that a sequence \( \{x^{(k)}\} \) converges to a vector \( x^* \) such that \( \nabla c(x^*) \) has full rank. For some vector \( \lambda^* \), the following formulations hold:
\[
\begin{align*}
\alpha_k \gamma^{(k)}(c_1^{(k)-1}(x^{(k)}), c_2^{(k)-1}(x^{(k)}), \ldots, c_m^{(k)-1}(x^{(k)}))^T &\to \lambda^*, \\
\alpha_k \gamma_k(c_1^{(k)-1}(x^{(k)}))e_1v(x^{(k)}), c_2^{(k)-1}(x^{(k)}))e_2v(x^{(k)}), \ldots, c_m^{(k)-1}(x^{(k)}))e_mv(x^{(k)}))^T &\to \lambda^*, \\
\nabla_x L_k(x^*, \lambda^*) &= 0, \quad c_i(x^*) = 0 \quad \forall 1 \leq i \leq m.
\end{align*}
\tag{3.19}
\]
Proof. From (3.14) we have
\[
\nabla_x L_k(x^{(k)}, \lambda^{(k)}) = \nabla f(x^{(k)}) + \alpha_k \sum_{i=1}^m c_i(x^{(k)}) \nabla c_i(x^{(k)}),
\tag{3.20}
\]
for all \( k \) such that \( \nabla c(x^*) \) has full rank. Since \( \epsilon_k \to 0 \), it follows that
\[
\begin{align*}
\alpha_k \gamma^{(k)}(c_1^{(k)-1}(x^{(k)}), c_2^{(k)-1}(x^{(k)}), \ldots, c_m^{(k)-1}(x^{(k)}))^T &\to \lambda^*, \\
\alpha_k \gamma_k(c_1^{(k)-1}(x^{(k)}))e_1v(x^{(k)}), c_2^{(k)-1}(x^{(k)}))e_2v(x^{(k)}), \ldots, c_m^{(k)-1}(x^{(k)}))e_mv(x^{(k)}))^T &\to \lambda^*, \\
\nabla_x L(x^*, \lambda^*) &= 0.
\end{align*}
\tag{3.21}
\]
Taking the bound of \( \{\lambda^{(k)}\} \) and \( \alpha_k \gamma^{(k)}c^{(k)-1}(x^{(k)}) \to \lambda^* \) into account, we obtain
\[
c_i(x^*) = 0, \quad i = 1, 2, \ldots, m, \tag{3.22}
\]
from \( \alpha_k \to \infty \). Consequently the result is obtained and the proof is complete. \( \square \)

A kind of self-adaptive penalty method is given above and it is analyzed under \( \alpha_k \to \infty \). Just as other penalty methods, there is a disadvantage that the penalty tends to infinity, which may cause numerical difficulty because of ill-condition, which is shown as follows. For convenience, we define
\[
L(x^{(k)}, \lambda^{(k)}, \alpha_k) = f(x) + \lambda^{(k)}c(x^{(k)}) + \alpha_k P(c(x^{(k)})).
\tag{3.23}
\]
Theorem 3.5. Let (3.13), $\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)})$, and $\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}, \alpha_k)$ be semipositive definite for sufficiently large $k$, where $L(x^{(k)}, \lambda^{(k)}, \alpha_k)$ is the Lagrangian function of (3.13). Then

$$\mathcal{K}(\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)})) \leq \mathcal{K}(\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}, \alpha_k)), \quad (3.24)$$

where $\mathcal{K}$ is the condition number.

Proof. Let $\gamma_1$ and $\bar{\gamma}_1$ be the smallest eigenvalues of $\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}, \alpha_k)$ and $\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)})$, respectively. If the polynomial penalty term is employed in some step, we now have

$$\gamma_1 = \min_{C_{\beta_1}(x^{(k)}+d) \leq 0} \frac{d^T \nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}, \alpha_k) d}{d^T d} \leq \min_{\nabla_{x} c(x^{(k)}) d = 0} \frac{d^T \nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}) d}{d^T d} = \bar{\gamma}_1. \quad (3.25)$$

Let $\gamma_n$ and $\bar{\gamma}_n$ be the largest eigenvalues of $\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}, \alpha_k)$ and $\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)})$, respectively. We therefore have

$$\gamma_n = \max_{d} \frac{d^T \nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}, \alpha_k) d}{d^T d} \leq \min_{\nabla_{x} c(x^{(k)}) d = 0} \frac{d^T \nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}) d}{d^T d} = \bar{\gamma}_n. \quad (3.26)$$

A similar result is also obtained if the exponent penalty term is employed. From (3.25) and (3.26) we have

$$\frac{\mathcal{K}(\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}))}{\gamma_1} = \frac{\bar{\gamma}_n}{\gamma_1} = \mathcal{K}(\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}, \alpha_k)), \quad (3.27)$$

which is the result, and the proof is complete.

Obviously, in terms of (3.25) and (3.26), we have

$$\lim_{k \to \infty} \mathcal{K}(\nabla_{xx}^2 L(x^{(k)}, \lambda^{(k)}, \alpha_k)) = \infty, \quad (3.28)$$

if $\alpha_k \to \infty$. The condition number thus becomes progressively worse provided that penalty-type methods are employed and penalty parameter $\alpha_k \to \infty$. Apparently, all the results in this section also hold provided that there exists a subsequence $\alpha_{k_i} \to \infty$. 
4 Concluding remarks

In this work, a self-adaptive penalty method is brought forward. The penalty term and penalty parameters are adjusted at each step. The early penalty approaches are the special cases of the new techniques in this paper with fixed $P_i$. In self-adaptive penalty approaches for general nonlinear programming, it is more flexible. Apparently, augmented Lagrangian multiplier approach is a kind of self-adaptive penalty method. Almost all early penalty approaches are special cases of self-adaptive penalty method.

In the above analysis, some new problems appear, which are listed as follows.

(i) For penalty term, we give an update in this work. It is interesting to find a “better” update rule.

(ii) For penalty parameter, can we give a threshold from the input data? When penalty term changes, how about penalty parameter? Moreover, in augmented Lagrangian methods, it is not necessary to increase penalty parameter if Lagrangian multipliers update.

(iii) Giving a practical algorithm is an important issue. Certainly, there exist some problems to implement the algorithm.

(iv) Are there any other choices about penalty parameter and penalty terms, which may be useful to design more practical algorithms?

Moreover, diverse penalty techniques can also be employed in each step. Namely, more flexible penalty can be utilized and the same results are also obtained. For example, we can use $P_k(c(x)) = \sum_{i=1}^{n} p_i^k(c_i(x))$. In this way, the differences between various constraints can be efficiently handled and the same results as those in Section 3 can also be obtained.

The following example is given to illustrate the self-adaptive penalty methods.

Example 4.1. Consider the following problem with single variable:

$$\min -x^8 \quad \text{s.t.} \quad x^2 = 0. \quad (4.1)$$

With suitable parameter $P(c(x)) = c(x)^{\alpha}$ and $\alpha > 1$, the optimal solution can be immediately obtained. At the beginning, it is very difficult to find the best penalty parameter and suitable penalty term. We can obtain the suitable penalty parameter and the fit penalty term by self-adaptive technique. Moreover, when $P(c(x)) = c(x)^{\alpha}$ and $\alpha > 1$ are employed, the rate of the constraints and the objective function converging to optimization is consistent.

As one of the efficient approaches, penalty methods have some above advantages. Different penalty terms relate to different threshold values. The penalty term is changed...
to get a “good” penalty term with small threshold value and small conditioned number in essence. In summary, importance has been attached to penalty approaches for several score of years because of their practice. The self-adaptive penalty strategies can be extended to general nonlinear programming and mathematical programs with equilibrium constraints.

Acknowledgment

Sincere thanks to the anonymous referee for his beneficial suggestion to improve the presentation.

References


Pu-yan Nie: Department of Mathematics, Jinan University, Guangzhou 510632, China
E-mail address: tpynie@jnu.edu.cn