International
Journal
Of Advanced Research in Engineering & Management (IJAREM)
ISSN: 2456-2033 || PP. 01-06

# A Smooth Function Approximating for an Exact Penalty Function

Xiaorong Zhu[1], Liang Fang[2], Guifeng Zhao[3], Yumei Huang[2]
[1](College of Information Science and Technology Taishan University, Taian, China)
[2](College of Mathematics and Statistics Taishan University, Taian, China)
[3](Tai'an Dongyue Middle School, Taian, China)

**Abstract:** The exact penalty function commonly used in the solution of nonlinear programming problems has a significant disadvantage of non-smoothness, which hinders the use of fast and effective minimization algorithms, and may cause some numerical instability. Therefore, this paper consider a method to smoothly approximate exact penalty function A, and gives the error estimation among smoothing penalty problem, the problem of nonsmooth penalty and the optimal objective function value of the original problem. On the basis of the given smoothing function, an algorithm for calculating the approximate optimal solution of the problem is proposed and the convergence of the algorithm is given.
**Keywords:** nonlinear programming problem, exact penalty function, convergence

## I. THE INTRODUCTION

Consider the nonlinear constrained optimization problem:

$$\min f(x)$$
$$(P) \quad s.t. \quad g_j(x) \le 0,\ j=1,2,\cdots,m$$

Where $f,\ g_j : R^n \to R,\ j=1,2,\cdots,m$ is a continuously differentiable function.

Let $F_0$ be the set of feasible solutions, i.e $F_0=\{x\in R^n \mid g_j(x)\le 0,\ j=1,2,\cdots,m\}$

We assume that $F_0$ is non empty, and a value function that is often used to solve such problems is the $l_1$ exact penalty function. It was first proposed by Zangwill[1] and has the following form:

$$F_1(x,\rho)=f(x)+\rho\sum_{j=1}^{m}\max\{g_j(x),0\} \qquad (1.1)$$

Where $\rho>0$ penalty parameter. The constrained optimization problem can be transformed into an unconstrained optimization problem by $l_1$ exact penalty function:

$$\min F_1(x,\rho)$$
$$(p_\rho) \quad s.t. \quad x\in R^n$$

Sometimes ,we only need to obtain an approximate solution of problem $(P)$ ,so smoothing approximation of the exact penalty function becomes one of the methods to solve this problem. Such methods have appeared in the literature such as Zang [2], Beatal and Teboulle [3]. It has been used by Auslender, Cominetti and Haddou [4] to study convex programming problems and linear programming problems, and Gonzaga and Castillo [5] to study nonlinear programming problems. Similarly, Chen and Mangasarian [6,7], Z. Q. Meng[8] proposed two different smoothing penalty functions for the bottom expression.

$$F_2(x,\rho)=f(x)+\rho\sum_{i=1}^{m}\sqrt{\max\{g_j(x),0\}} \qquad (1.2)$$

In Pinar and Zenios [9],a linear quadratic smoothing approximation penalty function is presented for convex programming problems, but such functions do not have quadratic continuous differentiability. We consider another smooth method to approximate the exact penalty function $F_1(x,\rho)$. The approximation function given by us for $F_1(x,\rho)$ is quadratic continuous differentiable, so we can combine Newton type method to solve the nonlinear constrained optimization problem.

In the second part, a method of smooth approximation of the exact penalty function (1.1) of $l_1$ is proposed and its error analysis is given. In the third part, the algorithm for calculating the approximate optimal solution of problem is presented based on the given smoothing function, and the convergence of the algorithm is given.

## II. A SMOOTHING PENALTY FUNCTION

Given $\varepsilon > 0$, define a function $P_\varepsilon(t)$:

$$P_\varepsilon(t)=\begin{cases}0, & if \quad t\le 0\\ \dfrac{t^3}{6\varepsilon^2}, & if \quad 0\le t<\varepsilon\\ t+\dfrac{\varepsilon^2}{2t}-\dfrac{4}{3}\varepsilon, & if \quad t\ge\varepsilon\end{cases}$$

If $P(t)=t^+=\max\{t,0\}$. It's easy to prove $\lim_{\varepsilon\to 0} P_\varepsilon(t)=P(t)$

In addition, $\forall\varepsilon>0$, $P_\varepsilon(t)$ are twice continuously differentiable with respect to $t$. Actually, we can get

$$P'_\varepsilon(t)=\begin{cases}0, & if \quad t\le 0\\ \dfrac{t^2}{2\varepsilon^2}, & if \quad 0\le t<\varepsilon\\ 1-\dfrac{\varepsilon^2}{2t^2}, & if \quad t\ge\varepsilon\end{cases}$$

$$P''_\varepsilon(t)=\begin{cases}0, & if \quad t\le 0\\ \dfrac{t^2}{\varepsilon^2}, & if \quad 0\le t<\varepsilon\\ \dfrac{\varepsilon^2}{t^3}, & if \quad t\ge\varepsilon\end{cases}$$

Notice that $P(g_j(x))=\max\{g_j(x),0\},\ j=1,2,\cdots,m$. Let's think about the penalty function for $(P)$

$$F(x,\rho,\varepsilon)=f(x)+\rho\sum_{j=1}^{m}P_\varepsilon(g_j(x)) \tag{2.1}$$

Where $\rho>0$ is the penalty parameter,

$$(p_{I_\rho}) \qquad \min_{x\in R^n} F(x,\rho,\varepsilon)$$

Since $\lim_{\varepsilon\to 0}F(x,\rho,\varepsilon)=F_1(x,\rho)$, for any $\rho$, let's first study the relationship between $(p_\rho)$ and $(p_{I_\rho})$.

**Lemma 2.1** For any given $x\in R^n$, and $x\in R^n$, $0\le F_1(x,\rho)-F(x,\rho,\varepsilon)\le\frac{4}{3}m\rho\varepsilon$.

Proof: according to the definition of $P_\varepsilon(t)$, we can get $0\le P(t)-P_\varepsilon(t)\le\frac{4}{3}\varepsilon$ Then for any $x\in R^n$, $0\le P(g_j(x))-P_\varepsilon(g_j(x))\le\frac{4}{3}\varepsilon,\ j=1,2,\cdots,m$, So

$$0\le\sum_{j=1}^{m}P(g_j(x))-\sum_{j=1}^{m}P_\varepsilon(g_j(x))\le\frac{4}{3}m\varepsilon$$

So, $0 \le F_1(x,\rho) - F(x,\rho,\varepsilon) \le \frac{4}{3} m\rho\varepsilon$

A direct consequence of Lemma 2.1 is as follows:

**Theorem 2.1** Let $\{\varepsilon_j\} \to 0$ be a sequence of positive numbers and assume that $x^j$ is a solution to $\min_{x \in R^n} F(x,\rho,\varepsilon_j)$. For some $\rho > 0$, let $\bar{x}$ be a convergence point of sequence $\{x^j\}$, then $\bar{x}$ is an optimal solution to $\min_{x \in R^n} F_1(x,\rho)$.

**Theorem 2.2** Let $x^*$ be an optimal solution of $(p_\rho)$ and $\bar{x} \in R^n$ be an optimal solution of $(p_{I_\rho})$, then

$$0 \le F_1(x^*,\rho) - F(\bar{x},\rho,\varepsilon) \le \frac{4}{3} m\rho\varepsilon$$

Definition 2.1 $x_\varepsilon \in R^n$ is called a feasible solution of $\varepsilon -$ or a solution of $\varepsilon -$ if $g_j(x_\varepsilon) \le \varepsilon, j = 1,2,\cdots,m$. Under this definition, the following results are obtained.

**Theorem 2.3** Let $x^*$ be an optimal solution of $(p_\rho)$ and $\bar{x} \in R^n$ be an optimal solution of $(p_{I_\rho})$. Further, if $x^*$ is feasible for $(P)$ and $\bar{x}$ is feasible for $(P)$ $\varepsilon -$, then $0 \le f(\bar{x}) - f(x^*) \le \frac{3}{2} m\rho\varepsilon$.

Proof: Since $\bar{x}$ is feasible for $(P)$ $\varepsilon -$, then

$$\sum_{j=1}^{m} P_\varepsilon(g_j(\bar{x})) = \sum_{(j:0 \le g_j(\bar{x}) < \varepsilon)} \frac{g_j(\bar{x})^3}{6\varepsilon^2} + \sum_{(j:g_j(\bar{x}) = \varepsilon)} (g_j(\bar{x}) + \frac{\varepsilon^2}{2g_j(\bar{x})} - \frac{4}{3}\varepsilon)$$

$$\le \sum_{(j:0 \le g_j(\bar{x}) < \varepsilon)} \frac{\varepsilon}{6} + \sum_{(j:g_j(\bar{x}) = \varepsilon)} \frac{\varepsilon}{6}$$

$$\le \frac{1}{6} m\varepsilon \tag{2.2}$$

Since $x^*$ is an optimal solution to $(P)$, we have $\sum_{j=1}^{m} P(g_j(x^*)) = 0$.

$$0 \le f(x^*) + \rho \sum_{j=1}^{m} P(g_j(x^*)) - (f(\bar{x}) + \rho \sum_{j=1}^{m} P_\varepsilon(g_j(\bar{x}))) \le \frac{4}{3} m\rho\varepsilon$$

$$\rho \sum_{j=1}^{m} P(g_j(x^*)) \le f(\bar{x}) - f(x^*) \le \rho \sum_{j=1}^{m} P_\varepsilon(g_j(\bar{x})) + \frac{4}{3} m\rho\varepsilon$$

Therefore, $0 \le f(\bar{x}) - f(x^*) \le \frac{3}{2} m\rho\varepsilon$ can be obtained from (2.2)

Theorem 2.1 and Theorem 2.2 prove that an approximate solution of $(p_{I_\rho})$ is also an approximate solution of $(p_\rho)$,When the error $\varepsilon$ is sufficiently small, according to Theorem 2.3,if the approximate optimal solution $\varepsilon -$ of $(p_{I_\rho})$ is feasible, an approximate optimal solution of $(p_{I_\rho})$ is also an approximate optimal solution of $(P)$.

## III. APPROXIMATION ALGORITHM

For $x \in R^n$, define

$$J^0(x) = \{ j \mid g_j(x) = 0, j = 1, 2, \cdots, m \},$$
$$J_\varepsilon^-(x) = \{ j \mid g_j(x) < \varepsilon, j = 1, 2, \cdots, m \},$$
$$J_\varepsilon^+(x) = \{ j \mid g_j(x) \geq \varepsilon, j = 1, 2, \cdots, m \}. \tag{3.1}$$

Consider the following algorithm:

**Algorithm A**

Step 1 Given $x^0, \varepsilon > 0, \varepsilon_0 > 0$ and $\rho_0 > 0$, let $k = 0$ go to step 2;

Step 2 Calculate $x^k \in \arg\min_{x \in R^n} F(x, \rho_k, \varepsilon_k)$;

Step 3 If $x^k$ for $(P)$ $\varepsilon -$, then stop, we get the approximate optimal solution $x^k$ of $(P)$, or $\rho_{k+1} = 2\rho_k, \varepsilon_{k+1} = \frac{1}{2}\varepsilon_k$, and $k = k+1$, turn to step 2.

Note: For this algorithm, if $k \to \infty$, then the sequence $\{\varepsilon_k\}$ approaches $0$ and $\{\rho_k\}$ approaches $+\infty$.If this algorithm does not terminate in finite steps, the following results can be obtained under appropriate conditions:

**Theorem 3.1** Suppose that $f(x)$ is enforced on $R^n$, that is, $\lim_{\|x\| \to +\infty} f(x) = +\infty$. Let $\{x^k\}$ be the sequence generated by algorithm A. Assuming $\{F(x^k, \rho_k, \varepsilon_k)\}$ is bounded, then $\{x^k\}$ is bounded and any limit point $x^*$ of $\{x^k\}$ is feasible, and there exists $\lambda \geq 0$, and $\mu_j \geq 0, j = 1, 2, \cdots, m$, such that

$$\lambda \nabla f(x^*) + \sum_{j \in J^0(x^*)} \mu_j \nabla g_j(x^*) = 0 \tag{3.2}$$

**Proof:** First, let's prove that $\{x^k\}$ is bounded. By assumption, there exist real numbers $L$ that makes $L > F(x^k, \rho_k, \varepsilon_k), k = 0, 1, 2, \cdots$

Suppose $\{x^k\}$ is unbounded. Let's assume that $\| x^k \| \to \infty$, when $k \to \infty$, then $L > f(x^k), k = 0, 1, 2, \cdots$ contradicts with the assumption that $f(x)$ on $R^n$ is enforced.

Let's prove that any limit point of $\{x^k\}$ is a member of $F_0$, without loss of generality, assuming $\lim_{k \to \infty} x^k = x^*$. Assuming $x^* \notin F_0$, then there exists $j \in \{1, \cdots, m\}$ which makes $P(g_j(x^*)) > \alpha > 0$ or $g_j(x^*) > \alpha > 0$, and notice that $g_j$, $j = 1, \cdots, m$ $F(x^k, \rho_k, \varepsilon_k)(k = 1, 2, \cdots)$ are continuous.

$$F(x^k, \rho_k, \varepsilon_k) = f(x^k) + \sum_{j \in J_\varepsilon^-(x^k)} \frac{g_j^+(x^k)^3}{6\varepsilon_k^2} + \sum_{j \in J_\varepsilon^+(x^k)} (g_j(x^k) + \frac{\varepsilon_k^2}{2g_j^+(x^k)} - \frac{4}{3}\varepsilon_k) \tag{3.3}$$

If $k \to \infty$, then for $k$, the set $\{1, \cdots, m\}$ is nonempty. Since $\{1, \cdots, m\}$ is a finite set, there exists $j_0 \in \{1, \cdots, m\}$ and a subset $K \subset N$ such that for any sufficiently large $k \in K, g_{j_0}(x^k) > \alpha$. According to (3.3), $F(x^k, \rho_k, \varepsilon_k) \to +\infty$ is a bounded contradiction to assumption $\{F(x^k, \rho_k, \varepsilon_k)\}$.

Now prove (3.2) was set up, according to the step 2, know $\nabla F(x^k, \rho_k, \varepsilon_k) = 0$, i.e.,

$$\nabla f(x^k) + \rho_k \sum_{j \in J_{\varepsilon_k}^-(x^k)} \frac{g_j^+(x^k)^2}{2\varepsilon_k^2} \nabla g_j(x^k) + \rho_k \sum_{j \in J_{\varepsilon_k}^+(x^k)} (1 - \frac{\varepsilon_k^2}{2g_j(x^k)^2}) \nabla g_j(x^k) = 0 \tag{3.4}$$

For $k = 1, 2, \cdots$, $\gamma_k = 1 + \sum_{j \in J_{\varepsilon_k}^-(x^k)} \rho_k \frac{g_j^+(x^k)^2}{2\varepsilon_k^2} + \sum_{j \in J_{\varepsilon_k}^+(x^k)} \rho_k (1 - \frac{\varepsilon_k^2}{2g_j(x^k)^2})$ ,so $\gamma_k > 0$.

From（3.4）， we can get

$$\frac{1}{\gamma_k}\nabla f(x^k)+\sum_{j\in J^-_{\varepsilon_k}(x^k)}\frac{\rho_k g_j^+(x^k)^2}{2\gamma_k\varepsilon_k^2}\nabla g_j(x^k)+\sum_{j\in J^+_{\varepsilon_k}(x^k)}\frac{\rho_k(1-\frac{\varepsilon_k^2}{2g_j(x^k)^2})}{\gamma_k}\nabla g_j(x^k)=0$$

Assuming

$$\lambda^k=\frac{1}{\gamma_k},\mu_j^k=\frac{\rho_k g_j^+(x^k)^2}{2\gamma_k\varepsilon_k^2},\ j\in J^-_{\varepsilon_k}(x^k),$$

$$\mu_j^k=\frac{\rho_k(1-\frac{\varepsilon_k^2}{2g_j(x^k)^2})}{\gamma_k},\ j\in J^+_{\varepsilon_k}(x^k),$$

$$\mu_j^k=0,\ j\in\{1,\cdots,m\}\setminus(J^-_{\varepsilon_k}(x^k)\cup J^+_{\varepsilon_k}(x^k)),$$

then

$$\lambda^k+\sum_{j=1}^m\mu_j^k=1,\ \mu_j^k\ge 0,\ j\in\{1,\cdots,m\},k=0,1,2,\cdots$$

When $k\to\infty$, we have $\lambda^k\to\lambda\ge 0,\mu_j^k\to\mu_j\ge 0,\forall j\in\{1,\cdots,m\}$, and

$$\lambda\nabla f(x^*)+\sum_{j=1}^m\mu_j^k\nabla g_j(x^*)=0,\ \lambda+\sum_{j=1}^m\mu_j=1$$

For $j\in J^-(x^*)$, we have $\mu_j^k\to 0$. So $\mu_j=0$, $\forall j\in J^-(x^*)$. Therefore, (3.2) holds.

Theorem 3.1 implies that the sequence produced by algorithm $x^*$ can converge to the KKT point of $(P)$ under certain conditions. In fact, if the Mangasarian-Fromovitz constraint holds at $x^*$ in Theorem 3.1, then $x^*$ is the KKT point of problem $(P)$.

## IV. CONCLUSION

The convergence rate of algorithm A depends on the speed of the algorithm applied to the unconstrained optimization problem in Step 2. Since the function $F(x,\rho_k,\varepsilon_k)$ is twice differentiable when $f,g_j,j=1,2,\cdots,m$ is twice continuously differentiable, we can obtain a faster convergence rate.

## ACKNOWLEDGEMENTS

This research was supported by the Shandong Provincial Natural Science Foundation, China(ZR2020MF035)

## REFERENCES

[1]. W. I. Zangwill, Nonlinear programming via penalty function [J], *Management Science*,1967 Vol.13, pp.334-358.
[2]. I. Zang, A smoothing-out technique for min-max optimization [J], *Mathematical Programming*, 1980, Vol.19, pp.61-77.
[3]. A. Bental and M. Teboulle, A smoothing technique for non-differentiable optimization problems [J], *Lecture Notes in Mathematics, Springer Verlag Berlin*, 1989, Vol.1405, pp.1-11.
[4]. A. Auslender, R. Cominetti and M. Haddou, Asymptotic analysis for penalty and barrier method in convex and linear programming [J], *Mathematics of Operational Research*,1997,Vol 22, pp.43-62.
[5]. C. C. Gonzaga and R. A. Castillo, A nonlinear programming algorithm based on noncoercive Penalty functions [J], *Mathematical Programming*, 2003, Vol.96, pp.87-101.
[6]. C. Chen and O. L. Mangasarian, Smoothing methods for convex inequalities and linear complementarity problems [J], *Mathematical Programming*, 1995, Vol.71, pp.51-69.
[7]. C. Chen and O. L. Mangasarian, A class of smoothing functions for nonlinear and mixed complementarity problems [J], *Computational Optimization and Applications*, 1996, Vol.5, pp.97-138.

[8]. Z. Q. Meng, C. Y. Dang,and X. Q. Yang, On the smoothing of the square-root exact penalty function for inequality constrained optimization [J], *Computational Optimization and Applications*, 2006, Vol.35, pp.375-398.

[9]. M. C. Pinar and S. A. Zenios, On smoothing exact penalty functions for convex constrained Optimization [J], *SIAM Journal on Optimization*, 1994, Vol.4, pp.486-511.