2012 年 3 月 运 筹 学 学 报 第 16 卷 第 1 期
March, 2012 Operations Research Transactions Vol.16 No.1

A New Penalty Function Based on Non-coercive Penalty Functions*

SHANG Youlin[1†] LIU Muhua[1] LI Pu[1]

Abstract For the differentiable nonlinear programming problem, this paper proposes a new penalty function form of the approached exact penalty function, presents with the gradual approximation algorithm and evolutionary algorithm, and proves that if the sequences of the approximation algorithm exist accumulation point, it certainly is the optimal solution of original problem. In the weak assumptions, we prove that the minimum sequences from the algorithm is bounded, and its accumulation points are the optimal solution of the original problem and get that in the Mangasarian-Fromovitz qualification condition, through limited iterations the minimum point is the feasible point.

Keywords exact penalty function, the feasible point, optimal solution, nonlinear programming

Chinese Library Classification O221.2

2010 Mathematics Subject Classification 90C33

一种新的逼近精确罚函数的罚函数及性质

尚有林 [1†] 刘牧华 [1] 李 璞 [1]

摘要 针对可微非线性规划问题提出了一个新的逼近精确罚函数的罚函数形式，给出了近似逼近算法与渐进算法，并证明了近似算法所得序列若有聚点，则必为原问题最优解. 在较弱的假设条件下，证明了算法所得的极小点列有界，且其聚点均为原问题的最优解，并得到在 Mangasarian-Fromovitz 约束条件下，经过有限次迭代所得的极小点为可行点.

关键词 精确罚函数，可行点，最优解，非线性规划

中图分类号 O221.2

数学分类号 90C33

收稿日期： 2011 年 3 月 30 日.

* The work is partially supported by The National Natural Science Foundation of China (No.10971053, 10771162), and The National Natural Science Foundation of Henan (No. 094300510050).

1. School of Mathematics & Statistics, Henan University of Science & Technology, Luoyang 471003, China; 河南科技大学数学与统计学院，河南洛阳， 471003.

† 通讯作者 Corresponding author

0 Introduction

Consider the nonlinear programming problem (P):

$$\begin{cases} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leqslant 0, i = 1, \cdots, m, x \in R^n, \end{cases} \tag{1}$$

where $f_i(x) : R^n \to R$ for $i = 1, \cdots, m$ are continuously differentiable functions.

We can assume there is lower bound for $f_0(x)$, or it can be replaced by $\mathrm{e}^{f_0(x)}$. The l_1 exact penalty function of problem (P) is:

$$f_\beta(x) = f_0(x) + \beta \sum_{i=1}^{m} (f_i(x))^+, \tag{2}$$

where $\beta > 0$ acts as the penalty parameter, $(f_i(x))^+ = \max\{0, f_i(x)\}, i = 1, \cdots, m$.

For the convex problem, Zangwill[1] found: for the l_1 exact penalty function, if the minimum point existed for a certain $\beta_0 > 0$, then for any $\beta > \beta_0$, minimum point exist for l_1 exact penalty function. Moreover, under the weak assumption, whenβ is sufficiently large, then the minimum point for $f_\beta(\cdot)$ is possible and also can be considered as the optimal. The most obvious defect of l_1 exact penalty function lies in its in differentiability, which has affect its application in effect minimization algorithm. Therefore, the differentiable approximation study for exact penalty function has aroused the scholars' interest[2−7]. In paper [2], Auslender has studied the differentiable approximation under convex and non-convex condition respectively, which has been extended to weaker assumption condition in [3-4], in addition, he has worked out the corresponding conclusion for positive semi-definite programming.

In paper [10], Gonzaga and Castilo has provided an algorithm including parameters γ and β, and minimizes at each iteration a penalized function with the shape

$$f_{\beta,\gamma}(x) = f_0(x) + \beta\gamma \sum_{i=1}^{m} \theta\left(\frac{f_i(x)}{\gamma}\right), \tag{3}$$

where $\theta(\cdot)$ is smooth approximation of the exact penalty, they proved that $f_{\beta,\gamma}(x)$ is a smooth approximation for l_1 exact penalty function, and the two parameters play the different roles: the decreasing of γ can increase accuracy of the same rate, the increased β is the power of the penalty, and $\beta \geqslant 1$, The main conclusion is: under the Mangasarian-Fromovitz constraint qualification, all the iteration are feasible for the β sufficiently large and γ sufficiently small, namely the minimum point obtained from the iteration is feasible, which will extend the feasibility of l_1 exact penalty function obtained by Zangwill[1] to the condition of uncertain penalty function.

This paper provides a new form of penalty function approached l_1 exact penalty function, and obtains the conclusion of the paper [10] under weaker assumption condition which has also been extended.

Different from the paper [10], for the penalty function form of the new approximate l_1 exact penalty function, this paper provides a new approximate progressive algorithm.

It is also proved in this paper that if there is accumulation point existed in the sequence obtained by approximate algorithm, it shall be the optimum solution, and if the sequence is boundless, then a sufficient condition shall be provided for the optimum value when the sequence converged. In addition, it is also proved that all the iteration are feasible for the β sufficiently large and γ sufficiently small satisfying Mangasarian-Fromovitz constrain qualification, meanwhile, the feasibility of l_1 exact penalty function is also extended to the inexact penalty function. An example is also provided at last.

1 Problem and penalty approach

Consider the differentiable nonlinear programming problem (P). For $\varepsilon \geqslant 0$ define the relaxed feasible set

$$\Omega_\varepsilon = \{x \in R^n | f_i(x) \leqslant \varepsilon, i = 1, \cdots, m\}.$$

Then Ω_0 is the feasible set for (P). We shall use the following hypotheses:

(H1) For some $\varepsilon \geqslant 0$, Ω_ε is bounded.

(H2) All optimal solutions of problem (P) satisfy the Mangasarian-Fromovitz qualification condition, namely, for any optimal solution x^*,there exist $h \in R^n$ and $\rho > 0$ such that

$$\nabla f_i(x^*)^{\mathrm{T}} h < -\rho$$

for all $i \in I(x^*) \equiv \{i = 1, 2, \cdots, m | f_i(x^*) = 0\}$.

A point $x \in R^n$ will be called interior if $f_i(x)$ for $i = 1, 2, \cdots, m$. Note that the Mangasarian-Fromovitz qualification condition implies that points exist.

A form of differentiable penalty function approached l_1 exact penalty function is provided as follow:

$$\theta(t) = \begin{cases} \dfrac{\mathrm{e}^t}{\mathrm{e}}, & t \leqslant 1, \\ t, & t > 1. \end{cases} \tag{4}$$

Thus the first-order derivative：

$$\theta^{'}(t) = \begin{cases} \dfrac{\mathrm{e}^t}{\mathrm{e}}, & t \leqslant 1, \\ t, & t > 1. \end{cases} \tag{5}$$

The function $\theta(t)$ satisfies the following properties：

Property 1 $\theta(t)$ is convex, differentiable and has $\theta^{'}(0) > 0$.

Property 2 $\lim\limits_{t \to -\infty} \theta(t) = 0$.

Property 3 (Recession function) $\theta^{'}_\infty(1) := \lim\limits_{t \to +\infty} \dfrac{\theta(t)}{t} = 1$;

Notice that property 2 implies：

$$\theta^{'}_\infty(-1) := \lim_{t \to +\infty} \frac{\theta(-t)}{t} = 0. \tag{6}$$

$$\lim_{t\to-\infty}\theta^{'}(t)=\lim_{t\to-\infty}\frac{\mathrm{e}^{t}}{\mathrm{e}}=0. \tag{7}$$

Consider a function $\theta(t)$ defined as above, and a parameter $\gamma\in(0,1]$, we consider the function：

$$\theta_{\gamma}(t)=\gamma\theta(\frac{t}{\gamma})=\begin{cases}\dfrac{\gamma\mathrm{e}^{\frac{t}{\gamma}}}{\mathrm{e}}, & t\leqslant 1,\\ t, & t>1.\end{cases} \tag{8}$$

For any $\gamma\in(0,1]$, $\theta_{\gamma}(t)$ satisfies the properties 1-3. Besides this, we have：

Property 4 For $t>0$, $\gamma\in(0,1]$, $\theta^{'}_{\gamma}(t)=\theta^{'}(\frac{t}{\gamma})\geqslant\theta^{'}(t)$,

Proof If $0<t\leqslant 1$,

$$\theta^{'}_{\gamma}=\theta^{'}(\frac{t}{\gamma})=\frac{\mathrm{e}^{\frac{t}{\gamma}}}{\mathrm{e}}\geqslant\frac{\mathrm{e}^{t}}{\mathrm{e}}=\theta^{'}(t),$$

If $t\geqslant 1,\theta^{'}_{\gamma}=\theta^{'}(t)=1$.

This completes the proof.

Property 5 For any $t\leqslant 0$ or $t>1$, $\theta_{\gamma}(t)$ converges pointwise to t^{+} when $\gamma\to 0^{+}$. If $0<t\leqslant 1,\ \theta_{\gamma}(t)=+\infty$.

Proof For t fixed,

$$\lim_{\gamma\to 0^{+}}\theta_{\gamma}(t)=\lim_{\gamma\to 0^{+}}\gamma\theta(\frac{t}{\gamma}).$$

If $t<0$, then clearly

$$\lim_{\gamma\to 0^{+}}\theta_{\gamma}(t)=\lim_{\gamma\to 0^{+}}\gamma\theta(\frac{t}{\gamma})=\lim_{\gamma\to 0^{+}}\gamma\frac{e^{\frac{t}{\gamma}}}{\mathrm{e}}=0=t^{+}.$$

If $0<t\leqslant 1$,

$$\lim_{\gamma\to 0^{+}}\theta_{\gamma}(t)=\lim_{\gamma\to 0^{+}}\gamma\theta(\frac{t}{\gamma})=\lim_{\gamma\to 0^{+}}\gamma\frac{\mathrm{e}^{\frac{t}{\gamma}}}{\mathrm{e}}=\lim_{\gamma\to 0^{+}}\frac{\mathrm{e}^{\frac{t}{\gamma}}}{\frac{\mathrm{e}}{\gamma}}=+\infty.$$

If $t>1$,

$$\lim_{\gamma\to 0^{+}}\theta_{\gamma}(t)=\lim_{\gamma\to 0^{+}}\gamma\theta(\frac{t}{\gamma})=t=t^{+}.$$

If $t=0$, then

$$\lim_{\gamma\to 0^{+}}\gamma\theta(\frac{t}{\gamma})=0=t^{+}.$$

This completes the proof.

Property 6 For $\gamma\in(0,1]$, $\theta_{\gamma}(t)\geqslant\theta(t)-\theta(0)$ when $t>0$.

Proof By the property 4, we have $\theta^{'}_{\gamma}(t)\geqslant\theta^{'}(t)$, both sides take the integral, then:

$$\int_{0}^{t}\theta^{'}_{\gamma}(x)\mathrm{d}x\geqslant\int_{0}^{t}\theta^{'}(x)\mathrm{d}x,$$

thus,

$$\theta_{\gamma}(t)-\theta_{\gamma}(0)\geqslant\theta(t)-\theta(0),$$

for $\theta_{\gamma}(0)=\gamma\theta(0)=\dfrac{\gamma}{\mathrm{e}}>0$. So

$$\theta_{\gamma}(t)\geqslant\theta(t)-\theta(0).$$

2 Smoothed penalty functions

The penalized function is now constructed using $\theta_\gamma(t)$:

$$f_{\beta,\gamma}(x) = f_0(x) + \beta \sum_{i=1}^{m} \theta_\gamma(f_i(x)). \tag{9}$$

We have two parameters, γ and β. The weight β changes the inclination of the penalty $\beta\theta_\gamma(\cdot)$. The parameter γ controls the precision of the smoothing.

The algorithm will play with γ and β as follows: γ decreases at all iterations (by say, $\gamma := \frac{\gamma}{2}$). At infeasible points β increases (by say, $\beta := 2\beta$); at feasible points β is not changed. The product $\beta\gamma$ never increase, which allows us to use the following property:

Property 7 If $\beta\gamma$ is bounded, then for any interior point x,

$$\lim_{\gamma\to 0^+} \beta \sum_{i=1}^{m} \theta_\gamma(f_i(x)) = 0.$$

Proof

$$\lim_{\gamma\to 0^+} \beta \sum_{i=1}^{m} \theta_\gamma(f_i(x)) = \lim_{\gamma\to 0^+} \beta\gamma \sum_{i=1}^{m} \theta\left(\frac{f_i(x)}{\gamma}\right).$$

For an interior point x, $f_i(x) < 0$ for $i \in \{1, 2, \cdots, m\}$. Hence

$$\lim_{\gamma\to 0^+} \theta\left(\frac{f_i(x)}{\gamma}\right) = 0,$$

by property 2. The result follows from the boundedness of $\beta\gamma$.

3 The algorithm

At this point we assume that for any $\beta \geqslant 1$, $\gamma \in (0, 1]$, $f_{\beta,\gamma}(\cdot)$ has minimizers.

The algorithm is given as follows:

step1: Let $\beta_0 = 1$, $\gamma_0 = 1$, $k := 0$.

step2: For $k = 0, 1, \cdots$, minimization:

$$x^k \in \arg\min_{x\in R^n} \left\{ f_0(x) + \beta_k \sum_{i=1}^{m} \theta_{\gamma_k}(f_i(x)) \right\}. \tag{10}$$

step3: $\gamma_{k+1} := \frac{\gamma_k}{2}$, if x^k is a feasible point, then $\beta_{k+1} := \beta_k$; else $\beta_{k+1} := 2\beta_k$.

step4: $k := k + 1$, turn to step2.

The following facts are trivially true for sequences generated by this algorithm:
For any $t \leqslant 0$ or $t > 1$, $\gamma_k \to 0$ and hence by property 5, $\theta_{\gamma_k} \to t^+$ pointwise. $\beta_k\gamma_k \leqslant 1$ and by property 7,

$$\beta_k\gamma_k \sum_{i=1}^{m} \theta\left(\frac{f_i(\tilde{x})}{\gamma_k}\right) \to 0$$

for any interior point $\tilde{x}$.

If $\{\tilde{x}\}$ has an infeasible infinite subsequence then $\beta_k \to \infty$.

For the above algorithm, we have the following assumption:

H3 For some $\beta \geqslant 1$, $\gamma \in (0,1]$, $\arg\min\limits_{x\in R^n}(x) \neq \emptyset$.

We also assume that the minimization algorithm can decide that a problem has no minimizer. Note that if the penalized problem has no minimizer, then there are points arbitrarily far from the feasible set where $f_{\beta,\gamma}(\cdot)$ is smaller than at any given feasible point. For the minimization will be well for β sufficiently large and γ sufficiently small under this assumption (H3), we shall change the minimization step in the algorithm to replace the step2.

4 Minimization

Either compute $x^k \in \arg\min\limits_{x\in R^n}\left\{f_0(x) + \beta_k \sum\limits_{i=1}^{m} \theta_{\gamma_k}(f_i(x))\right\}$, or decide that there exists $x^k \bar{\in} \Omega_\varepsilon$ such that $f_{\beta_k,\gamma_k}(x^k) \leqslant f_{\beta_k,\gamma_k}(\tilde{x})$, where $\tilde{x}$ is a feasible point.

The minimization step that we given above is for a feasible point $\tilde{x}$, generalizing the step which be given in the paper [10].

Thus we can proof that the sequence from the algorithm converges to the optimal solution of the problem (P), without need of Mangasarian-Fromovitz constraint qualification. However the paper [10] gets this conclusion must be satisfying the Mangasarian-Fromovitz constraint qualification.

From now on, assume that$\{x^k\}$, $\{\gamma_k\}$, $\{\beta_k\}$ are sequences generated by the algorithm, keeping in mind that by hypothesis the first iteration is successful.

Lemma 1 Assuming that the hypothesis H1 and H3 establishment, then the sequence $\{x^k\}$ is bounded, and all its accumulation points are feasible solutions.

Proof In the first iteration, $\beta_0 = \gamma_0 = 1$, for $k \in N = \{1, 2, \cdots\}$,

$$f_0(x^k) + \sum_{i=1}^{m} \theta(f_i(x^k)) \geqslant f_0(x^0) + \sum_{i=1}^{m} \theta(f_i(x^0)). \tag{11}$$

By the property 6, $\theta(t) \leqslant \theta_{\gamma_k}(t) + \theta(0)$, hence for $k \in N$, $i = 1, 2, \cdots, m$,

$$\theta_{\gamma_k}(f_i(x^k)) \geqslant \theta(f_i(x^k)) - \theta(0),$$

and it follows that

$$\beta_k \sum_{i=1}^{m} \theta_{\gamma_k}(f_i(x^k)) \geqslant \beta_k \sum_{i=1}^{m} \theta(f_i(x^k)) - m\theta(0). \tag{12}$$

Adding (11) and (12), so

$$f_0(x^k)+\beta_k \sum_{i=1}^{m} \theta_{\gamma_k}(f_i(x^k)) \geqslant \beta_k \sum_{i=1}^{m} \theta(f_i(x^k)) - \sum_{i=1}^{m} \theta(f_i(x^k)) + f_0(x^0) + \sum_{i=1}^{m} \theta(f_i(x^0)) - m\theta(0). \tag{13}$$

Now consider the interior point $\tilde{x}$. By definition of x^k for $k \in K$,

$$f_0(x^k) + \beta_k \sum_{i=1}^{m} \theta_{\gamma_k}(f_i(x^k)) \leqslant f_0(\tilde{x}) + \beta_k \sum_{i=1}^{m} \theta_{\gamma_k}(f_i(\tilde{x})). \tag{14}$$

Subtracting (13) from (14) and collecting the constant terms,

$$f_0(x^k) + \beta_k \sum_{i=1}^{m} \theta_{\gamma_k}(f_i(\tilde{x})) \geqslant \beta_k \sum_{i=1}^{m} \theta(f_i(x^k)) - \sum_{i=1}^{m} \theta(f_i(x^k)) + \alpha,$$

where $\alpha = f_0(x^0) + \sum\limits_{i=1}^{m} \theta(f_i(x^0)) - m\theta(0)$.

Now, we proof: $\forall \varepsilon > 0$, there exists a k_ε, when $k > k_\varepsilon$, we have

$$x_k \in \Omega_\varepsilon. \tag{15}$$

Assume by contradiction that there exists $\varepsilon_0 > 0$, $i_0 = \{1, 2, \cdots, m\}$ and an infinite subsequence $K \subseteq N$, such that:

$$f_{i_0}(x^k) \geqslant \varepsilon_0, \quad \forall \quad k \in K, \tag{16}$$

hence, for ever k, $\{1 \leqslant i \leqslant m | f_{i_0}(x^k) \geqslant \varepsilon_0\} \neq \emptyset$.

By (14) and (15), property 1 and the properties of convex function, for $k \in K$,

$$\begin{aligned}
& f_0(\tilde{x}) + \beta_k \sum_{i=1}^{m} \theta_{\gamma_k}(f_i(\tilde{x})) \\
\geqslant \;& \beta_k \sum_{f_i(x^k) \geqslant \varepsilon_0} \theta(f_i(x^k)) - \sum_{f_i(x^k) \geqslant \varepsilon_0} \theta(f_i(x^k)) + \beta_k \sum_{f_i(x^k) \leqslant \varepsilon_0} \theta(f_i(x^k)) \\
& - \sum_{f_i(x^k) \leqslant \varepsilon_0} \theta(f_i(x^k)) + \alpha \\
\geqslant \;& (\beta_k - 1) \sum_{f_i(x^k) \geqslant \varepsilon_0} \theta'(f_i(x^k)) f_i(x^k) + \beta_k \sum_{f_i(x^k) \leqslant \varepsilon_0} \theta(f_i(x^k)) - m\theta(\varepsilon_0) + \alpha \\
\geqslant \;& (\beta_k - 1) \sum_{f_i(x^k) \geqslant \varepsilon_0} \theta'(\varepsilon_0)\varepsilon_0 - m\theta(\varepsilon_0) + \alpha \\
\geqslant \;& (\beta_k - 1)\theta'(\varepsilon_0)\varepsilon_0 - m\theta(\varepsilon_0) + \alpha.
\end{aligned}$$

The first inequality above is got by (14), the second inequality is by the properties of convex function, the third inequality is by $\beta_k \theta(f_i(x^k)) \geqslant 0$ and the increasing of $\theta'(\cdot)$, by the property 1 we can get $\theta'(\varepsilon_0) > 0$, so the fourth inequality is got by (16). Taking limits for the both side of the formula above, using the property 7, right-hand side tend to finite value $f_0(\tilde{x})$, but by $\beta_k \to +\infty$, we know that the right-hand side tend to infinity, this is a contradiction. So we get(15).

Using the hypothesis H1, we obtain the$\{x^k\}$ is bounded.

By (15), $\forall\varepsilon > 0$, for large value of k,

$$f_i(x_k) \leqslant \varepsilon.$$

Assuming $\bar{x}$ is any accumulation point of $\{x^k\}$, taking limits for k of the formula above, thus

$$f_i(\bar{x}) \leqslant \varepsilon,$$

$$f_i(\bar{x}) \leqslant 0,$$

when $\varepsilon \to 0$. So we proof that any accumulation point of $\{x^k\}$ is a feasible solution of (P).

Theorem 1 Assuming that the hypothesis H1 and H3 establishment, then any accumulation point of $\{x^k\}$ is an optimal solution of (P).

Proof Let $\bar{x}$ be an accumulation point of $\{x^k\}$, and let $k \in K$ be such that $x^k \to \bar{x}$. By lemma 1, $\bar{x}$ is feasible, for large value of k, then

$$x_k \in \Omega_\varepsilon. \tag{17}$$

Let $\tilde{x}$ be a feasible point, by the definition of x_k and (17), for large value of $k \in K$,

$$f_0(x^k) \leqslant f_0(\tilde{x}) + \beta_k \sum_{i=1}^{m} \theta_{\gamma_k}(f_i(\tilde{x})) - \beta_k \sum_{i=1}^{m} \theta_{\gamma_k}(f_i(x^k)). \tag{18}$$

With the definition, we can get $\theta_{\gamma_k}(f_i(x^k)) \geqslant 0$, by (18), we have

$$f_0(x^k) \leqslant f_0(\tilde{x}) + \beta_k \sum_{i=1}^{m} \theta_{\gamma_k}(f_i(\tilde{x})). \tag{19}$$

Taking limits for $k \to +\infty$, $\quad \forall k \in K$, using the property 7, we have $\beta_k\theta_{\gamma_k}(f_i(\tilde{x})) \to 0$.

Taking limits for the both side of the (19) formula, for any feasible point$\tilde{x}$,

$$f(\bar{x}) \leqslant f(\tilde{x}),$$

hence, any accumulation point of $\{x^k\}$ is an optimal solution of (P).

Given a point $x \in R^n$, define the sets

$$I(x) = \{i = 1, 2, \cdots, m | f_i(x) = 0\}, \quad I^+(x) = \{i = 1, 2, \cdots, m | f_i(x) \geqslant 0\}.$$

Lemma 2 Assuming that the hypothesis H1 and H3 establishment, if $x^k \to x^*$, then for sufficiently large k, $I^+(x^k) \subset I(x^*)$.

Proof Assume that $x^k \to x^*$, and for any $i \in \{1, 2, \cdots, m\}$. We must prove that if $i\bar{\in}I(x^*)$, then for large k, $i\bar{\in}I^+(x^*)$. If $i\bar{\in}I(x^*)$, then $f_i(x^*) < 0$, due to the continuity of $f_i(\cdot)$, for large k, $f_i(x^k) < 0$, or equivalently, $i\bar{\in}I^+(x^*)$, completing the proof.

Theorem 2 Assuming that the hypothesis H1 and H3 establishment, let $\{x^k\}$ be a sequence generated by the algorithm. Then there exists $\bar{k} \in N$, such that for $k \geqslant \bar{k}$, x^k is a feasible point.

Proof Assume by contradiction that there exists an infinite set K_1 such that for $k \in K_1$, x^k is an infeasible point. Then by construction, $\beta_k \to +\infty$, by lemma 1, $\{x^k\}$ is bounded, and hence $\{x^k\}_{k \in K_1}$ has an accumulation point, which is optimal by theorem 1.

Let $K_2 \subset K_1$, be such that $k \in K_2$, $x^k \to x^*$.

Associating with each $k \in N$, the set $I^+(x^k)$ defined above, the following three facts are true for large $k \in K_2$:

(a) $I^+(x^k) \subset I(x^*)$, by lemma 2. If $I(x^*) = \emptyset$, then x^k is feasible, contradicting our assumption. Assume then that $I(x^*) \neq \emptyset$.

(b) For $i \in I(x^*)$,

$$f_i^{'}(x^k, h) = \nabla f_i(x^k)^{\mathrm{T}} h \leqslant -\frac{\rho}{2},$$

where $h \in R^n, \rho > 0$ are given by the qualification condition (H2). This follows from the continuity of the gradients and the fact that $x^k \to x^*$, $k \in K_2$.

(c) For $i \bar{\in} I(x^*)$, $f_i(x^k) \leqslant -\mu < 0$, where $\mu > 0$ such that

$$-\mu > \max\{f_i(x^*) | i \bar{\in} I(x^*)\}.$$

Define the set $K_3 \subset K_2$ as the set of iteration indices where these three facts are true.

Consider the directional derivatives along h, by the definition of x^k, $f^{'}_{\beta_k,\gamma_k}(x^k, h) = 0$. And hence, using the fact that $\theta^{'}_{\gamma}(t) = \theta^{'}_{\gamma}(\frac{t}{\gamma})$,

$$f_0^{'}(x^k, h) + \beta_k \sum_{i=1}^{m} \theta^{'} \left(\frac{f_i(x^k)}{\gamma_k} \right) f_i^{'}(x^k, h) = 0. \tag{20}$$

Dividing by β_k and spitting the indices in the summation,

$$\frac{f_0^{'}(x^k, h)}{\beta_k} + \sum_{i \in I(x^*)} \theta^{'} \left(\frac{f_i(x^k)}{\gamma_k} \right) f_i^{'}(x^k, h) = - \sum_{i \bar{\in} I(x^*)} \theta^{'} \left(\frac{f_i(x^k)}{\gamma_k} \right) f_i^{'}(x^k, h). \tag{21}$$

Taking limits for $k \to +\infty (k \in K)$, we have:

Left-hand side of (21):

$$\frac{f_0^{'}(x^k, h)}{\beta_k} \to 0,$$

because $\beta_k \to +\infty$.

By (13),

$$\frac{\beta_k f_i(x^k)}{\gamma_k} \leqslant \frac{-\mu \beta_k}{\gamma_k},$$

for $i \bar{\in} I(x^*)$, hence,

$$\frac{\beta_k f_i(x^k)}{\gamma_k} \to -\infty.$$

And

$$\theta^{'}(\frac{f_i(x^k)}{\gamma_k}) \to 0,$$

due to (7).

We conclude that the left-hand side converges to 0 in K_3.

Consider the right-hand side of (21): For $i \in I(x^*)$,

$$f_i^{'}(x^k, h) = \nabla f_i(x^k)^T h \leqslant -\frac{\rho}{2},$$

by (b). Hence,

$$-\sum_{i \in I(x^*)} \theta^{'}\left(\frac{f_i(x^k)}{\gamma_k}\right) f_i^{'}(x^k, h) \geqslant \frac{\rho}{2} \sum_{i \in I(x^*)} \theta^{'}\left(\frac{f_i(x^k)}{\gamma_k}\right) \geqslant \frac{\rho}{2} \sum_{i \in I^+(x^*)} \theta^{'}\left(\frac{f_i(x^k)}{\gamma_k}\right) \geqslant \frac{\rho}{2}\theta^{'}(0)|I^+(x^*)|.$$

Because $I^+(x^*) \subset I(x^*)$ by (a), and for $i \in I^+(x^*)$, $f_i(x^*) \geqslant 0$ and $\theta^{'}(\cdot)$ is increasing.

Since x^k is an infeasible point for $k \in K_3$, $I^+(x^*) \geqslant 1$, and hence the right-hand side of (21) has a positive lower bound, contradicting (21) and completing the proof.

The above mentioned Theorem 2 indicates that limited iteration is not feasible under the Mangasarian-Fromovitz constraint qualification.

Therefore, we try our best to demonstrate that all iteration are feasible for β sufficiently large and γ sufficiently small under the Mangasarian-Fromovitz constraint qualification. Thus, the feasibility of l_1 exact penalty function obtained by Zangwill[1] is extended to the situation of inexact penalty function. To ensure the establishment of theorem, it is required that the minimal point obtained in each iteration shall be entire minimal point which is feasible for convex problem but difficult for non-convex problem, therefore, we are thinking about if the point obtained by iteration is local minimal point, the similar conclusion can be achieved or not, this issue needs further study.

Example Now we show a simple one-dimensional example:

$$\begin{cases} \min & -2x, \\ \text{s.t.} & (x+0.25)^2 - \dfrac{1}{8} \leqslant 0, \\ & x \leqslant 0. \end{cases} \tag{22}$$

That can be transformed into the questions follow:

$$\begin{cases} \min & \mathrm{e}^{-2x}, \\ \text{s.t.} & (x+0.25)^2 - \dfrac{1}{8} \leqslant 0, \\ & x \leqslant 0. \end{cases} \tag{23}$$

We use the penalty function given by $\theta(t)$, and obtain the results of the problem (23), the optimal solution is 0, and the optimal value is 1, obviously, the optimal solution of the problem (23) is same to the problem (22). Then we obtain the optimal solution of the problem (22) is 0.

References

[1] Zangwill W I. Non-linear programming via penalty functions [J]. *Manage Sci*, 1967, **13**: 344-358.

[2] Auslender A. Penalty and barrier methods: a unified framework [J]. *Optim*, 1999, **10**: 211-230.

[3] Auslender A, Cominetti R, Haddou M. Asymptotic analysis for penalty and barrier methods in convex and linear programming [J]. *Math Operations Res*, 1997, **22**: 43-62.

[4] Ben-Tal A, Teboulle M. A smoothing technique for non-differentiable optimazation problems [C]. *Lecture Notes in Mathematics*, Berlin, West-Germany: Springer Verlag, 1989, 1405: 1-11.

[5] Chen C, Mangasarian Q L. Smoothing methods for convex inequalities and linear complementary problems [J]. *Math Programming*, 1995, **71**: 1-112.

[6] Chen C, Mangasarian Q L. A class of smoothing functions for non-linear and mixed complementary problems [J]. *Comput Optim Appl*, 1996, **5**: 97-138.

[7] Pinar M, Zenios S. On smoothing exact penalty functions for convex constrained optimization [J]. *SIAM J Optim*, 1994, **4**: 486-511.

[8] Zang I. A smoothing-out technique for min-max optimization [J]. *Math Prog,* 1980, **19**: 61-77.

[9] Sun X L, Li D. Asymptotic strong duality for bounded integer programming: a logarithmic-exponential dual formulation [J]. *Mathematics of Operations Research*, 2000, **25**(4): 625-644.

[10] Gonzaga C C, Castillo R A. A nonlinear programming algorithm based on non-coercive penalty function [J]. *Math Programming* (Ser. A), 2003, **96**: 87-101.