正確なペナルティ関数の理論における現実性の追求

柳下 翔太郎　リスク解析戦略研究センター 特任助教

（本研究は日本大学 伊藤勝氏との共同研究です）

1　制約付き最適化問題とペナルティ関数

Ωを$\mathbb{R}^n$の部分集合とし，以下のような最適化問題を考える：

$$\underset{x\in\Omega}{\text{minimize}} \quad f(x) \tag{1}$$

- $\mathbb{R}^n$はn次元ベクトルの集合を表します
- 制約集合Ωは実行可能領域ともいい，ベクトルに満たしてほしい性質を表します
- 例えば，$\Omega = \{x \in \mathbb{R}^n \mid x$の各要素は非負$\} = [0,\infty)$，$\Omega = \{x \in \mathbb{R}^n \mid x$の非ゼロ要素数は$K$以下$\}$ = 基数制約などが用いられます
- 最適化問題が与えられたとき，その最適解を解析的に求めるのは一般には困難であるので，何らかのアルゴリズムを用いて近似的に解くことを考えます
- 制約付き最適化問題（$\Omega \neq \mathbb{R}^n$）は無制約最適化問題（$\Omega = \mathbb{R}^n$）よりも扱いにくいことが知られています
- そこで，制約付き最適化問題の代わりにペナルティ関数を用いた無制約最適化問題を考えます

$P : \mathbb{R}^n \to [0,\infty)$が

$$P(x) = 0 \iff x \in \Omega$$

を満たすとき，PをΩに対する**ペナルティ関数**であるという

- 例えば，$\max\{-x,0\}$や$\max\{-x,0\}^2$は$\Omega = [0,\infty)$のペナルティ関数です（図1）
- $T_K^{(1)}(x) := \min_{\substack{\Lambda\subset\{1,\dots,n\}\\|\Lambda|=n-K}} \sum_{i\in\Lambda}|x_i|$や$T_K^{(2)}(x) := \min_{\substack{\Lambda\subset\{1,\dots,n\}\\|\Lambda|=n-K}} \sum_{i\in\Lambda}x_i^2$は基数制約のペナルティ関数です

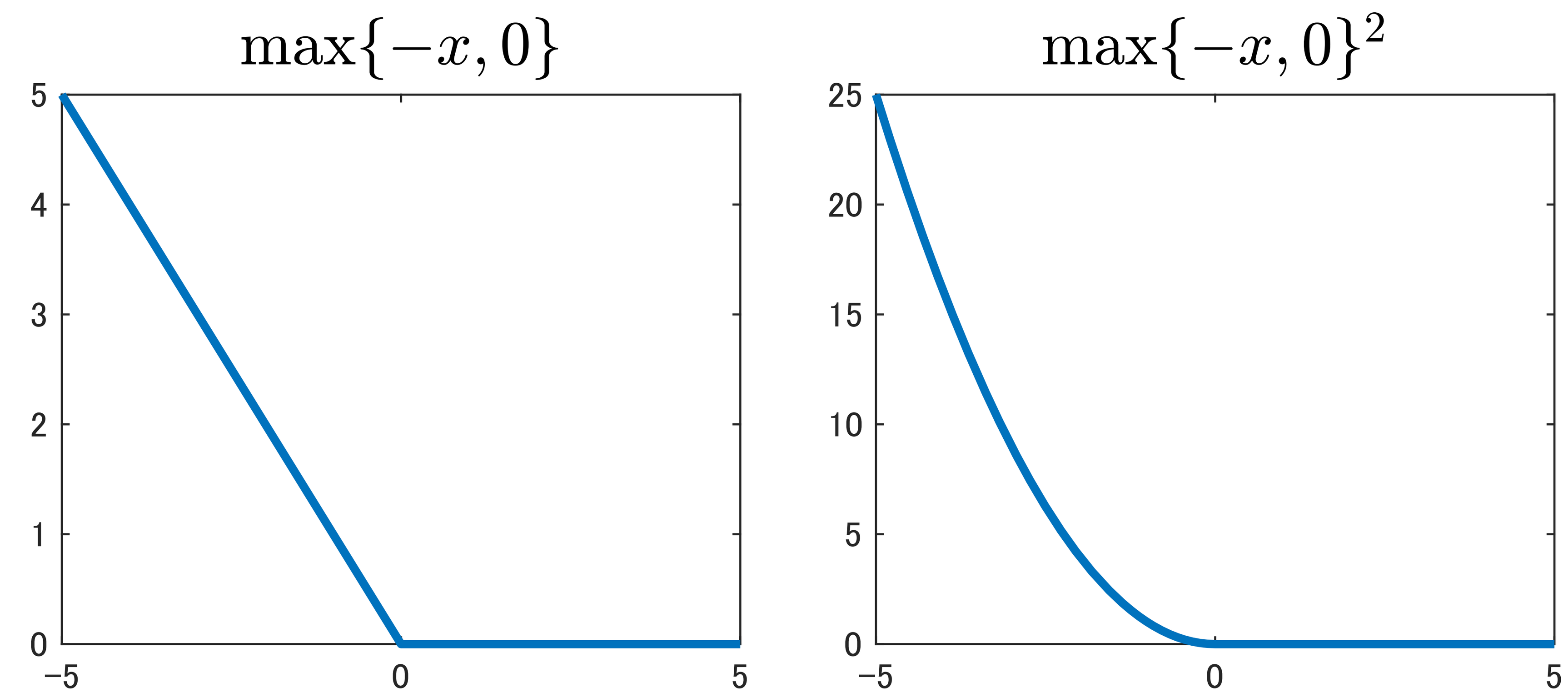

図 1: $\Omega = [0,\infty)$のペナルティ関数

$\gamma > 0$として，以下のペナルティフォームを考える：

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad F_\gamma(x) := f(x) + \gamma P(x) \tag{2}$$

- ペナルティフォーム(2)は元の制約付き最適化問題(1)とは異なります．(2)を解くことで(1)を解いたことになるのでしょうか？
- 典型的なものとして，fとPに対する適当な仮定のもと，γをある閾値より大きくとると，両者の最適解の集合が一致することが知られています
- しかし，非凸最適化問題においては最適解を手に入れることは非常に困難であるため，現実的にはあまり意味のない結果です
- そこで，(2)の停留点が(1)の停留点になるための十分条件が調べられています（e.g., Demyanov et al., 1998）
- 停留点はアルゴリズムによって手に入るので最適解に関する結果よりも現実的ですが，あくまで極限で手に入るものであるので，現実的かという観点でいえばまだ不十分です
- 最適解や停留点に対する上記のような性質をペナルティ関数の**正確性**といいます
- 正確性の理論は，1960年代後半から（Eremin, 1967, Zangwill, 1967, Pietrzykowski, 1969）現在まで盛んに研究されています
- 本研究では，近似停留点に対して正確性を示しました
- 近似停留点はアルゴリズムの有限回の反復で手に入るため，**十分に現実的な結果**であるといえます
- さらにその結果を応用することによって，ペナルティ関数法の効率性の理論解析である**反復計算量解析が可能に**なりました

2　近似停留点における正確性

$\varepsilon > 0$として，最適化問題$\min_{x\in X} F(x)$に対して

$$\underline{d}F(x^*;v) \geq -\varepsilon\|v\|, \quad \forall v \in \mathcal{F}(x^*;X)$$

を満たす点$x^* \in X$を**ε-停留点**とよぶ

- $\underline{d}F(x;v) := \liminf_{\eta\searrow 0} \dfrac{F(x+\eta v) - F(x)}{\eta}$かつ$\mathcal{F}(x;X)$は実行可能錐，$\varepsilon$-停留点の集合を$\mathrm{dStat}_X^\varepsilon(F)$と表す
- P：方向微分可能，$dP(x,v) = \lim_{\eta\searrow 0}\frac{P(x+\eta v)-P(x)}{\eta} > -\infty$，$\forall x \notin \Omega$, $\exists v \in \mathbb{R}^n$ $s.t.$ $dP(x;v) < 0$を仮定
- (2)が正確 $\overset{def}{\iff}$ $\exists\gamma^* \geq 0$ $s.t.$ $\forall\gamma > \gamma^*$, $\mathrm{dStat}_{\mathbb{R}^n}^\varepsilon(F_\gamma) \subset \mathrm{dStat}_\Omega^\varepsilon(f)$ $\iff$ $\exists\gamma^* \geq 0$ $s.t.$ $\forall\gamma > \gamma^*$, $\forall x \in \mathrm{dStat}_{\mathbb{R}^n}^\varepsilon(F_\gamma) \setminus \Omega$, $\exists v \in \mathbb{R}^n$ $s.t.$

$$\underline{d}F_\gamma(x;v) < -\varepsilon\|v\|$$

定義 1. $\exists\gamma^* \geq 0$ $s.t.$ $\forall\gamma > \gamma^*$, $\forall x \in \tilde{\Omega} \setminus \Omega$, $\exists v \in \mathbb{R}^n$ $s.t.$

$$dP(x;v) < 0,\ \underline{d}F_\gamma(x;v) < -\varepsilon\|v\|$$

が成立するとき，(2)が**$\tilde{\Omega}(\subset \mathbb{R}^n)$上で狭義正確**であるという

- $\tilde{\Omega} = \cup_{\gamma>0}\mathrm{dStat}_{\mathbb{R}^n}^\varepsilon(F_\gamma)$とすれば，「(2)が$\tilde{\Omega}$上で狭義正確 $\Longrightarrow$ (2)が正確」
- $\gamma^*(\tilde{\Omega}) := \max\left\{0,\ \sup_{x\in\tilde{\Omega}\setminus\Omega} \inf_{v:dP(x;v)<0} -\frac{\underline{d}f(x;v)+\varepsilon\|v\|}{dP(x;v)}\right\}$

定理 1 (狭義正確性の必要十分条件). (2)が$\tilde{\Omega}$上で狭義正確 $\iff$ $\gamma^*(\tilde{\Omega}) < \infty$

定理 2 (狭義正確性の十分条件). 以下を仮定する：

(I) $\exists\Gamma \geq 0$ $s.t.$ $\sup_{\substack{x\in\tilde{\Omega}\setminus\Omega\\v\in\mathbb{R}^n}} \underline{d}f(x;v) \leq \Gamma\|v\|$

(II) $\exists a > 0$ $s.t.$ $\forall x \notin \Omega$, $\exists v \neq 0$ $s.t.$ $dP(x;v) \leq -a\|v\|$

このとき，$\gamma^*(\tilde{\Omega}) \leq \frac{\Gamma+\varepsilon}{a} < \infty$

- fが**リプシッツ連続**のとき，仮定(I)は$\tilde{\Omega} = \mathbb{R}^n$で成立
- $\tilde{\Omega} = \cup_{\gamma>0}\mathrm{dStat}_{\mathbb{R}^n}^\varepsilon(F_\gamma)$とすると，仮定(I)は**近似停留点における勾配の有界性**
- 仮定(II)は制約の外で**ペナルティ関数が一様に傾いている**ことを意味
- $\max\{-x,0\}$や$T_K^{(1)}(x)$は仮定(II)を満たすが，$\max\{-x,0\}^2$や$T_K^{(2)}(x)$は満たさない（図1）