國立臺灣師範大學理學院數學系(所)

碩士論文

Department of Mathematics
College of Science
National Taiwan Normal University

Master’s Thesis

對於支撐向量機中 Truncated Pinball 損失的平滑化函數

Smoothing Functions of the Truncated Pinball Loss for Support Vector Machines

蔡旻原

Min-Yuan Tsai

指導教授：陳界山博士

Advisor: Jein-Shan Chen, Ph.D.

中華民國 110 年 7 月

July 2021

謝辭

感謝家人們朋友們給我的鼓勵與幫助。感謝我的指導教授:陳界山老師。平常我上台報告的時候，老師給予我許多的回饋，讓我能夠更加進步。在論文撰寫上，老師也給予許多指導，讓我學習到許多新知識。

感謝口試委員杜威仕、柯春旭、朱亮儒和張毓麟教授，給予我許多的建議，使得我能夠在論文的撰寫上，更加完善。也感謝 Harold 和 Chieu 兩位學長，在研究途中給我一些方向、指引，以及論文架構上的修改建議。最後感謝陳界山老師，在我大學的時候，跟我們說過，恆心毅力、挫折容忍力、自學能力是很重要的三個能力。在研究所的三年中，我也確實習得了這三項能力，謝謝界山老師的指導。

感謝數學系提供了許多資源，讓學生們得以專注學習以及做研究。

Acknowledgements

Thanks to my advisor Dr. Jein Shan Chen. When I presented my report, I got a lot of feedback from him. Therefore, I became better. When I wrote my thesis, he gave me a lot of guidance. Therefore, I learned new knowledge. Thanks to the committee meners of my oral exam for expert suggestions on thesis revision.

When I was undergraduate student, Prof. Chen tell us several important abilities: grit and self-study. I kept this in my mind. During the years in my master program, I got them. Hence, I become a better person. Last but not least, thanks to Dr. Harold and Dr. Chieu for helps.

I thanks department of Mathematics, NTNU. I have gotten a lot of resources, then I am able to finish my thesis. I also thanks to my families and my friends. They always encourage me and support me. Hence, I can keep on doing my research.

蔡旻原 謹識於

國立臺灣師範大學 數學所

中華民國 110 年 7 月

摘要

我們的研究目的是探討 truncated pinball 損失函數$P_{\tau,s}(x)$以及它的平滑化函數$\varphi_{\tau,s}(x,\mu)$。我們推導了$P_{\tau,s}(x)$可以被寫為絕對值函數跟仿射函數的和：$-\frac{\tau}{2}|x+s|+\frac{1+\tau}{2}|x|+\frac{\tau s+x}{2}$。再者，我們使用了來自於多篇參考文獻中的絕對值函數$|x|$的平滑化函數$\varphi_{abs}^{k}(x,\mu)$ $(\mathrm{k}=1,2,\ldots,10)$來產出我們關於 truncated pinball 損失函數的平滑化函數$\varphi_{\tau,s}^{k}(x,\mu)$ $(\mathrm{k}=1,2,\ldots,10)$性質的主要結果。因此，我們可以把原先的最佳化問題$\min_{w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}P_{\tau,s}(1-y_i(w^T\Phi(x_i)+b))$中的$P_{\tau,s}(x)$替換成$\varphi_{\tau,s}^{k}(x,\mu)$來得到可微分的最佳化問題。我們得出的結論是當$\mu$趨近於$0^+$的時候我們的可微分的最佳化問題$\min_{w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}\varphi_{\tau,s}^{k}(1-y_i(w^T\Phi(x_i)+b),\mu)$就變回原問題。更進一步地說，尋找可微分的最佳化問題的解將引出原問題的解。
關鍵字：Truncated pinball 損失函數、平滑化函數、可微分的最佳化問題。

Abstract

The objective of our research was to investigate the truncated pinball loss function $P_{\tau,s}(x)$ and its smoothing function $\varphi_{\tau,s}(x,\mu)$. We derived $P_{\tau,s}(x)$ can be rewritten as the sum of absolute value functions and an affine function: $-\frac{\tau}{2}|x+s|+\frac{1+\tau}{2}|x|+\frac{\tau s+x}{2}$. Moreover, we used the results of smoothing functions $\varphi_{abs}^{k}(x,\mu)$ $(\mathrm{k}=1,2,\ldots,10)$ of absolute value function $|x|$ from many references to produce our main results about properties of smoothing functions $\varphi_{\tau,s}^{k}(x,\mu)$ $(\mathrm{k}=1,2,\ldots,10)$ of the truncated pinball loss function $P_{\tau,s}(x)$. Hence, we can replace $P_{\tau,s}(x)$ with $\varphi_{\tau,s}^{k}(x,\mu)$ for the original minimization problem $\min_{w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}P_{\tau,s}(1-y_i(w^T\Phi(x_i)+b))$ to obtain a differentiable minimization problem. We concluded that as μ approaches 0^+ our differentiable minimization problem, $\min_{w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}\varphi_{\tau,s}^{k}(1-y_i(w^T\Phi(x_i)+b),\mu)$, becomes the original one. Furthermore, finding solutions to the differentiable minimization problem will lead to solution to the original one.
Key words: The truncated pinball loss function, Smoothing function, Differentiable optimization problem.

Table of Contents

1 Introduction 1

2 Preliminaries 11

3 Main Results 24

4 Conclusion 38

References 38

List of Tables

Table 1. The density function ρ_i
and its corresponding smoothing function 14

List of Figures

Figure 1. Graph of the “best” hyperplane H 1
Figure 2. Graph of the two distances from H to the closest point in I and II, respectively 1
Figure 3. Graph of the feature map Φ 4
Figure 4. Graph of $L_{hinge}(x)$ 5
Figure 5. Graph of $L_\tau(x)$ with $\tau = 0.6$ 6
Figure 6. Graphs of distances of $d(\cdot)$ and $D(\cdot)$ 7
Figure 7. Two kinds of distances 7
Figure 8. Graph of $P_{\tau,s}(x)$ with $\tau = 0.6,\ s = 1$ 8
Figure 9. Three important loss functions apperaed in [15] 8
Figure 10. Two important sigmoid functions 15
Figure 11. $|x|$ and $\varphi^k_{abs}(x,\mu)$ for $k = 1,2,3,4$ with $\mu = 0.1$ 21
Figure 12. $|x|$ and $\varphi^k_{abs}(x,\mu)$ for $k = 5,6,7,8$ with $\mu = 0.1$ 22
Figure 13. $\varphi^k_{abs}(x,\mu)$ is non-convex for $k = 9,10$ with $\mu = 0.1$ 22
Figure 14. $\varphi^k_{abs}(x,\mu)$ and $\nabla_x\varphi^k_{abs}(x,\mu)$ for $k = 9,10$ with $\mu = 0.1$ 23
Figure 15. $f_1(x) + f_2(x) = P_{\tau,s}(x)$ with $\tau = 0.6,\ s = 1$ 25
Figure 16. Graphs of $P_{\tau,s}(x)$, $f_1(x)$, and $f_2(x)$ with $\tau = 0.6,\ s = 1$ 26
Figure 17. $f_{convex}(x) + f_{concave}(x) = P_{\tau,s}(x)$ with $\tau = 0.6,\ s = 1$ 27
Figure 18. Graphs of $P_{\tau,s}(x)$, $f_{convex}(x)$, and $f_{concave}(x)$ with $\tau = 0.6,\ s = 1$ 28
Figure 19. Graphs of $P_{\tau,s}(x)$ and $\varphi^k_{\tau,s}(x,\mu)$ with $\mu = 0.2,\ \tau = 0.6,\ s = 1$ 29
Figure 20. Graphs of $|x|$ and all ten $\varphi^k_{abs}(x,\mu)$ with $\mu = 0.1$ 34
Figure 21. Conjecture one 35
Figure 22. Conjecture two 35
Figure 23. Ordering $\varphi^k_{\tau,s}(x,\mu)$ for $k = 9,10$ with different μ 36
Figure 24. Conjecture three 37

1 Introduction

Support vector machine (SVM) is a popular tool for binary classification [9, 17, 19, 22]. The basic idea of SVM for binary classification is to divide the samples into two classes by the “best” hyperplane. What is the “best” hyperplane? A simple measure is that the hyperplane has the largest distance to the two classes of samples, see Figure 1.

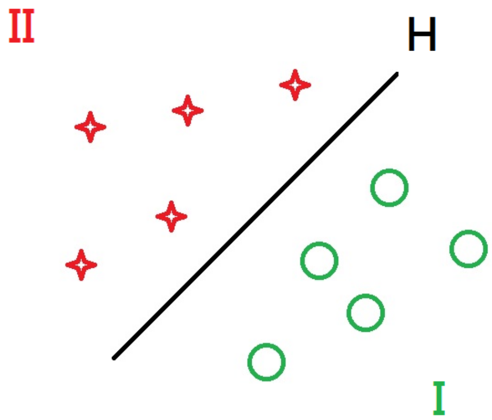


Figure 1: Graph of the “best” hyperplane H.

In addition, the diagram in Figure 2 illustrates the concept of the largest distance. We depict more by an algebraic way. Given a sample set $S = \{(x_i, y_i)|i = 1, 2, ..., l\}$, where $x_i \in \mathbb{R}^n$ and $y_i \in \{-1, 1\}$. We denote two index sets I and II by $\mathrm{I} = \{i \,|\, y_i = +1\}$ and $\mathrm{II} = \{i \,|\, y_i = -1\}$. Then, the hyperplane $H : w^T x + b = 0$ satisfies

$$\begin{cases} w^T x_i + b > 0, & \text{if } i \in \mathrm{I}. \\ w^T x_i + b < 0, & \text{if } i \in \mathrm{II}. \end{cases}$$

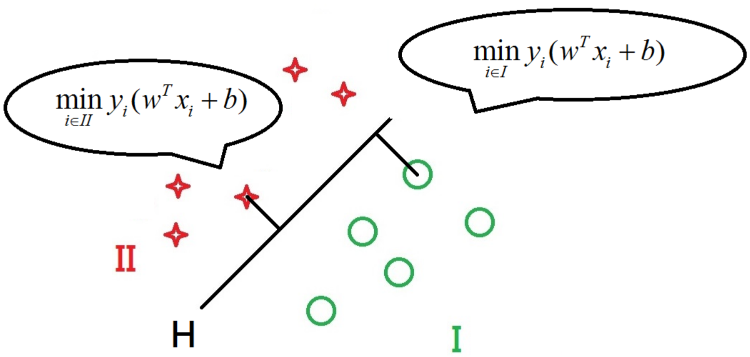


Figure 2: Graph of the two distances from H to the closest point in I and II, respectively.

In fact, the distance $\mathrm{dist}(x_i, H)$ between x_i and $H : w^T x + b = 0$ is

$$\frac{|w^T x_i + b|}{\|w\|} = y_i \frac{w^T x_i + b}{\|w\|} = y_i \left[\left(\frac{w}{\|w\|} \right)^T x_i + \frac{b}{\|w\|} \right].$$

Now, consider w being a unit vector in $\mathbb{R}^n$, we compute two distances (for fixed w with $\|w\| = 1$ and b) as below:

$$\min_{i \in \mathrm{I}} y_i (w^T x_i + b) \tag{1}$$

is the shortest distance between H and x_i for $i \in \mathrm{I}$; whereas

$$\min_{i \in \mathrm{II}} y_i (w^T x_i + b) \tag{2}$$

is the shortest distance between H and x_i for $i \in \mathrm{II}$. Therefore, the maximum of the sum of (1) and (2) consists of the largest distance. Thus, it is natural to look into the optimization problem:

$$\max_{\substack{w \in \mathbb{R}^n \\ \text{with } \|w\|=1, \\ b \in \mathbb{R}}} \left[\min_{i \in \mathrm{I}} y_i (w^T x_i + b) + \min_{i \in \mathrm{II}} y_i (w^T x_i + b) \right]. \tag{3}$$

For convenience, we let $\min_{i \in \mathrm{I}} y_i (w^T x_i + b) = \min_{i \in \mathrm{II}} y_i (w^T x_i + b)$ by a suitable replacement of b. In other words, the two distances from H to two classes are equal. To see this, if $\min_{i \in \mathrm{I}} y_i (w^T x_i + b) \neq \min_{i \in \mathrm{II}} y_i (w^T x_i + b)$, then we can set the value of b to other value. For instance, say

$$\begin{cases} \min_{i \in \mathrm{I}} y_i (w^T x_i + b) = 2. \\ \min_{i \in \mathrm{II}} y_i (w^T x_i + b) = 4. \end{cases}$$

This indicates that the two hyperplanes are $H_1 : w^T x + b = 2$ and $H_2 : w^T x + b = -4$. It is clear that for $i \in \mathrm{I}$, x_i satisfying $w^T x_i + b \geq 2$; and for $i \in \mathrm{II}$, x_i satisfying $w^T x_i + b \leq -4$. Then, replacing b with $(b - 1)$, we acquire two hyperplanes $H_1' : w^T x + b = 3$ and $H_2' : w^T x + b = -3$. For fixed w and b, now H_1' is the closest hyperplane to $\{x_i \,|\, i \in \mathrm{I}\}$, while H_2' is the closest hyperplane to $\{x_i \,|\, i \in \mathrm{II}\}$. As a result, the hyperplane H with w being unit vector satisfies

$$\begin{cases} w^T x_i + b \geq 3, & \text{if } i \in \mathrm{I} \\ w^T x_i + b \leq -3, & \text{if } i \in \mathrm{II} \end{cases}$$

which leads to $\min_{i \in \mathrm{I}} y_i (w^T x_i + b) = 3 = \min_{i \in \mathrm{II}} y_i (w^T x_i + b)$. Hence, we can rewrite (3) as

$$\max_{\substack{w \in \mathbb{R}^n \\ \text{with } \|w\|=1, \\ b \in \mathbb{R}}} \left[\min_i 2(y_i (w^T x_i + b) \right]. \tag{4}$$

Here, $\min_i y_i(w^T x_i + b)$ means the shortest distance between H and x_i for $i = 1, 2, ..., l$.

To tackle the above problem (4), a possible approach is trying to move the constraint condition $\|w\| = 1$ to the objective function. For this purpose, we define two hyperplanes H_{I} and H_{II} as below:

$$H_{\text{I}} : w^T x + b = +1,$$

$$H_{\text{II}} : w^T x + b = -1.$$

Note that the distance between H_{I} and H_{II} is $\frac{1-(-1)}{\|w\|} = \frac{2}{\|w\|}$. Accordingly, we obtain an equivalent optimization problem:

$$\begin{aligned} &\max_{w\in\mathbb{R}^n, b\in\mathbb{R}} \frac{2}{\|w\|} \\ &\quad \text{s.t. } y_i(w^T x_i + b) \geq 1 \end{aligned} \tag{5}$$

Now, the objective function is not differentiable, a relaxation comes by squaring it and taking the reciprocal so that the problem (5) becomes

$$\begin{aligned} &\min_{w\in\mathbb{R}^n, b\in\mathbb{R}} \frac{1}{2}\|w\|^2 \\ &\quad \text{s.t. } y_i(w^T x_i + b) \geq 1 \end{aligned} \tag{6}$$

This is a minimization problem with differentiable objective function. In summary, when the sample set $S = \{(x_i, y_i)|i = 1, 2, ..., l\}$ can be linearly separated by hyperplane H, the above minimization model (6) is often employed to deal with classification.

However, if the sample set S cannot be linearly separated at all, there is an alternative approach, which considers the higher-dimensional space (**feature space**) $\mathbb{R}^m$ $(m > n)$ to deal with this kind of problem. In particular, a mapping Φ: $\mathbb{R}^n \to \mathbb{R}^m$, called **feature map or kernel function** [17], is needed. More specifically, the resulting problem becomes

$$\begin{aligned} &\min_{w\in\mathbb{R}^m, b\in\mathbb{R}} \frac{1}{2}\|w\|^2 \\ &\quad \text{s.t. } y_i(w^T \Phi(x_i) + b) \geq 1 \end{aligned} \tag{7}$$

Figure 3 depicts the main idea behind, we only present an easy example to elaborate it. Assume that data $\{p_i\}_{i=1}^l$ in $\mathbb{R}^2$ are separated by this closed (ellipse) curve $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$ into two classes. By taking the feature map $\Phi((x_1, x_2)) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$, then there exists a hyperplane $\frac{1}{a^2}z_1 + 0z_2 + \frac{1}{b^2}z_3 = 1$ separating $\{\Phi(p_i)\}_{i=1}^l$ into two classes, which indicates $\{\Phi(p_i)\}_{i=1}^l$ can be linearly separated in $\mathbb{R}^3$.

In reality, there may need to handle the problem with feature noise, which is noise on input $\{x_i \,|\, i = 1, 2, \cdots, l\}$. Accordingly, the concept of **soft margin** is introduced to

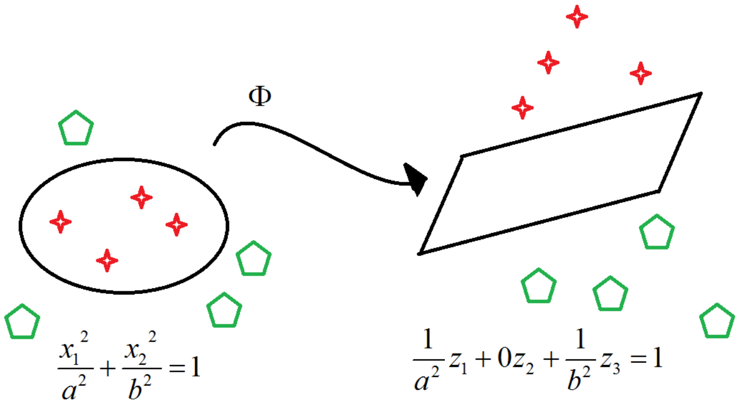


Figure 3: Graph of the feature map Φ

handle feature noise. Soft margin means to relax the boundary of the constraint set of (7), which possess new constraint sets:

$$\begin{cases} y_i(w^T\Phi(x_i)+b) & \geq & 1-\xi_i, \\ \xi_i & \geq & 0. \end{cases}$$

These constraint sets mean that we allow some misclassification of points in the sample set S. Meanwhile, we hope to minimize the sum of errors ξ_i. Thus, we add the error term $\sum\limits_{i=1}^{l}\xi_i$ to cost function $\frac{1}{2}\|w\|^2$. Then, it arises the following problem, which is called **the soft margin support vector machine**:

$$\begin{aligned} \min_{w\in\mathbb{R}^m, b\in\mathbb{R}, \xi\in\mathbb{R}^m} \ & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \\ \text{s.t. } & y_i(w^T\Phi(x_i)+b) \geq 1-\xi_i \\ & \xi_i \geq 0 \end{aligned} \tag{8}$$

The parameter $C > 0$ is called the **penalty parameter** of the error term. In the literature [15], by adopting the so-called loss function, it is known that the soft margin support vector machine is equivalent to loss support vector machine. Indeed, there are a few existing loss functions, and we present three popular one.

The first popular loss function is **the hinge loss function** [15], given by

$$L_{\text{hinge}}(x) = \begin{cases} x & \text{if} \quad x \geq 0, \\ 0 & \text{if} \quad x < 0. \end{cases}$$

whose graph is shown in Figure 4. The corresponding hinge loss support vector machine

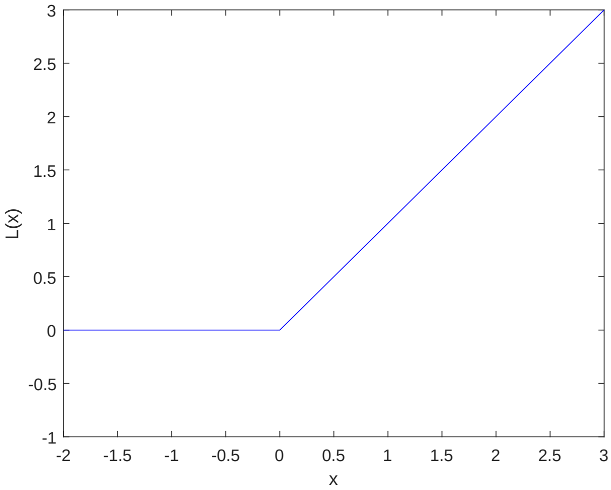


Figure 4: Graph of $L_{\text{hinge}}(x)$

(hinge loss SVM) is the following minimization problem:

$$\min_{w\in\mathbb{R}^m,b\in\mathbb{R}} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} L_{\text{hinge}}(1-y_i(w^T\Phi(x_i)+b)). \tag{9}$$

Apparently, from the definition of the hinge loss function, it is easy to see that this problem (9) is equivalent to above problem (8). To seek the minimum w of the hinge loss SVM (9), differentiating (9) yields

$$w \in C\sum_{i=1}^{l} \left[\partial L_{\text{hinge}}\left(1-y_i(w^T\Phi(x_i)+b)\right) y_i\right] \Phi(x_i),$$

where $\partial L_{\text{hinge}}(x)$ is the sub-differential of $L(x)$ with

$$\partial L_{\text{hinge}}(x) = \begin{cases} 1 & \text{if } x > 0, \\ [0,1] & \text{if } x = 0, \\ 0 & \text{if } x < 0. \end{cases}$$

Note that the coefficient $[\partial L(1-y_i(w^T\Phi(x_i)+b))y_i]$ of $\Phi(x_i)$ is non-zero when $y_i(w^T\Phi(x_i)+b) < 1$. Then, we call such (x_i, y_i) satisfying non-zero coefficient of $\Phi(x_i)$ as **support vector (SV)**. In other words, those SVs of (9) are in the condition $y_i(w^T\Phi(x_i)+b) < 1$ which are satisfied. In [15], it is concluded that the percentage of SVs is relative to the horizontal part of loss function curve. Therefore, it is said that the hinge loss SVM (9) have good sparsity due to the low percentage of SVs [15].

Next, we introduce the second loss function in [15], which is **the pinball loss function**. It is defined by

$$L_\tau(x) = \begin{cases} x & \text{if } x \geq 0, \\ -\tau x & \text{if } x < 0, \end{cases}$$

where $\tau \in [0,1]$; its graph is shown in Figure 5. The corresponding pinball loss SVM

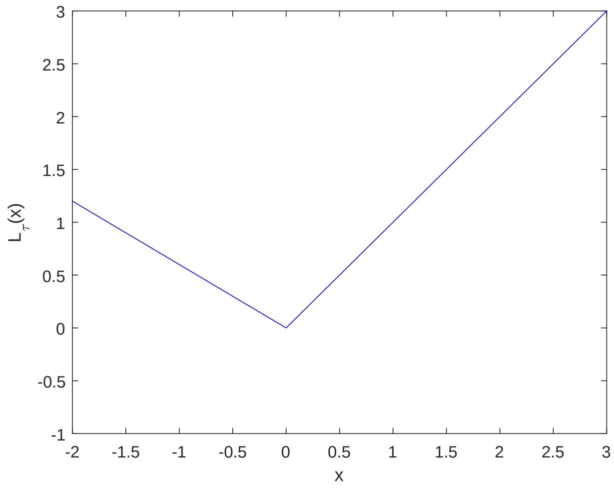


Figure 5: Graph of $L_\tau(x)$ with $\tau = 0.6$

(pin-SVM) from [9] is the minimization problem:

$$\min_{w\in\mathbb{R}^m, b\in\mathbb{R}} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} L_\tau\left(1 - y_i(w^T\Phi(x_i)+b)\right), \tag{10}$$

where $L_\tau(u)$ is the pinball loss function. In [9], it is already verified that the problem (10) is equivalent to the following problem:

$$\begin{aligned} \min_{w\in\mathbb{R}^m, b\in\mathbb{R}, \xi\in\mathbb{R}^m} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i, \\ \text{s.t. } & y_i(w^T\Phi(x_i)+b) \geq 1 - \xi_i \\ & y_i(w^T\Phi(x_i)+b) \leq 1 + \frac{1}{\tau}\xi_i. \end{aligned} \tag{11}$$

We point out that when $\tau = 0$, the second constraint becomes $\xi_i \geq 0$, see [9]. Moreover, the constraints set of (11) can be recast as

$$\begin{cases} w^T\Phi(x_i)+b \leq 1 + \frac{1}{\tau}\xi_i, & \text{if } y_i = +1 \\ w^T\Phi(x_i)+b \geq 1 - \xi_i, & \text{if } y_i = +1 \\ w^T\Phi(x_i)+b \geq -1 + \xi_i, & \text{if } y_i = -1 \\ w^T\Phi(x_i)+b \leq -1 - \frac{1}{\tau}\xi_i, & \text{if } y_i = -1 \end{cases} \tag{12}$$

The diagram in Figure 6 elaborates the visualizations of constraints (12).

There are two distances $d(\cdot)$ and $D(\cdot)$ in Figure 6. The distance $d(\cdot)$ describes $\text{dist}(H'_{\text{I}}, H'_{\text{I},1})$, whereas the distance $D(\cdot)$ represents $\text{dist}(H'_{\text{I}}, H'_{\text{I},2})$. From Figure 6, it

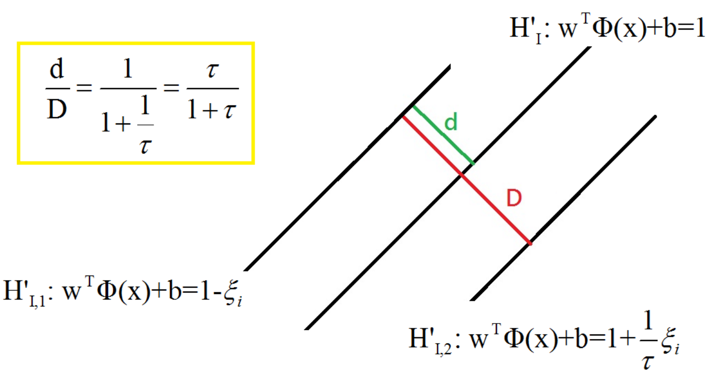
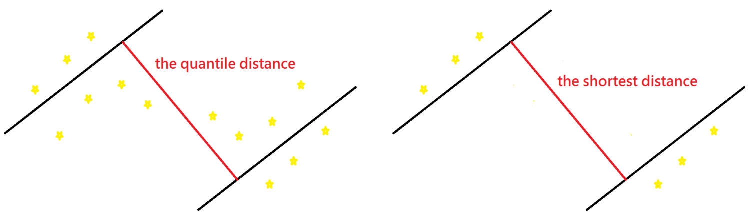


Figure 6: Graphs of distances $d(\cdot)$ and $D(\cdot)$



(a) Quantile distance (b) Shortest distance

Figure 7: Two kinds of distances.

is clear to see the ratio $\frac{d}{D} = \frac{\tau}{1+\tau}$, which is called **lower quantile**. Hence, we also call the distance between H'_{I} and H'_{II} **the quantile distance**.

The quantile distance is more insensitive to feature noise than the shortest distance. Hence, the pinball loss SVM model is more insensitive to feature noise, see [15] for more details.

Now, look back at the pin-SVM problem (10). In order to find its minimum w, by differentiating (10), we acquire the expression of w as below:

$$w \in C \sum_{i=1}^{l} \left[\partial L_\tau \left(1 - y_i(w^T \Phi(x_i) + b)\right) y_i \right] \Phi(x_i), \tag{13}$$

where

$$\partial L_\tau(x) = \begin{cases} 1 & \text{if } \ x > 0, \\ [-\tau, 1] & \text{if } \ x = 0, \\ -\tau & \text{if } \ x < 0. \end{cases}$$

We know that the coefficient $[\partial L_\tau(1 - y_i(w^T \Phi(x_i) + b))y_i]$ of $\Phi(x_i)$ is non-zero when $y_i(w^T \Phi(x_i) + b) \neq 1$. Hence, the pin-SVM problem (10) has high percentage of SVs, that is, the pin-SVM loss the sparsity completely, see [15].

The third loss function was also studied in [15, pp. 201], which is **the truncated pinball loss function**. It is defined by

$$P_{\tau,s}(x) = \begin{cases} \tau s & \text{if } \; x \le -s, \\ -\tau x & \text{if } \; -s < x < 0, \\ x & \text{if } \; x \ge 0, \end{cases} \tag{14}$$

where $\tau \in [0,1]$ and $s > 0$ and its graph is depicted in Figure 8.

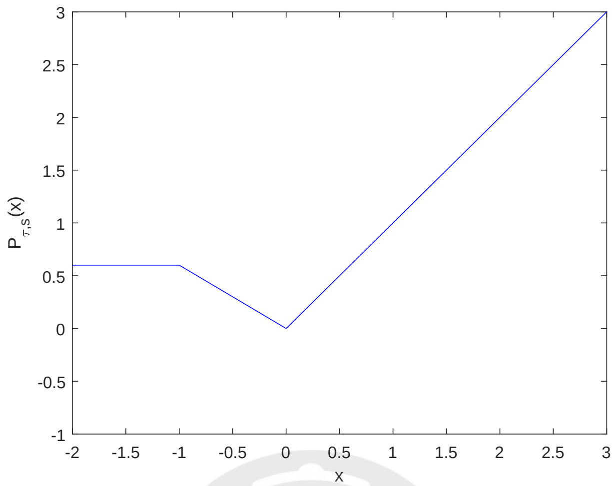
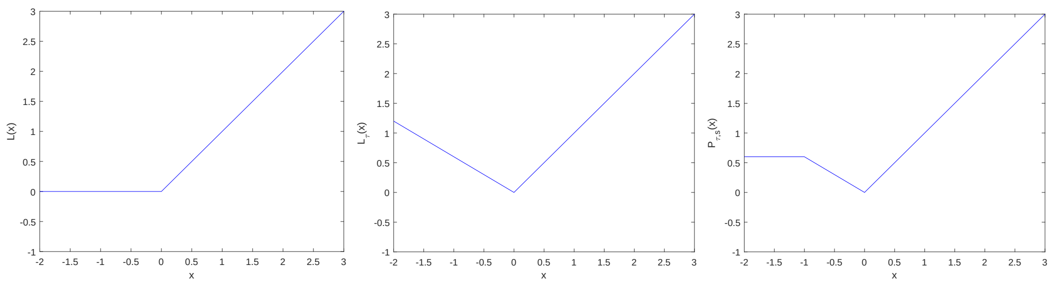


Figure 8: Graph of $P_{\tau,s}(x)$ with $\tau = 0.6, s = 1$.

From Figure 8, we see that this truncated pinball loss function is not differentiable at $x = 0$ and $x = -s$. Moreover, in light of Figure 9, we observe that

$$P_{0,0}(x) = L_{\text{hinge}}(x) \quad \text{and} \quad P_{\tau,+\infty}(x) = L_\tau(x), \tag{15}$$

which indicates that $P_{\tau,s}$ is a generalization of L_{hinge} and L_τ.



(a) L (b) L_τ (c) $P_{\tau,s}$

Figure 9: Three important loss functions appeared in [15]

The corresponding truncated pinball loss SVM ($\overline{\text{pin}}$-SVM) is the minimization problem:

$$\min_{w \in \mathbb{R}^m, b \in \mathbb{R}} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{l} P_{\tau,s}\left(1 - y_i(w^T \Phi(x_i) + b)\right). \tag{16}$$

The sub-differential of $P_{\tau,s}(x)$ is characterized as

$$\partial P_{\tau,s}(x) = \begin{cases} 1 & \text{if } \ x > 0, \\ [-\tau, 1] & \text{if } \ x = 0, \\ -\tau & \text{if } \ -s < x < 0, \\ [-\tau, 0] & \text{if } \ x = -s, \\ 0 & \text{if } \ x < -s. \end{cases} \tag{17}$$

Similarly, looking for w satisfying $w \in C \sum_{i=1}^{l} [\partial P_{\tau,s}(1 - y_i(w^T \Phi(x_i) + b))y_i]\Phi(x_i)$ provides solution candidates. In addition, the coefficient $\left[\partial P_{\tau,s}(1 - y_i(w^T \Phi(x_i) + b))y_i\right]$ of $\Phi(x_i)$ is non-zero when $y_i(w^T \Phi(x_i) + b) < 1 + s$. Hence, the $\overline{\text{pin}}$-SVM (16) is a relatively sparse model with a properly selected parameter s, see [15].

In summary, as mentioned in [15], the hinge loss SVM (9) is sparse, but it is sensitive to feature noise. In contrast to the hinge loss SVM, the pinball loss SVM (10) is insensitive to feature noise. Therefore, $P_{\tau,s}(x)$ was proposed in [15] to combine these two features, see the observation (15). Moreover, from the hinge loss $L_{\text{hinge}}(x)$ to the truncated pinball loss function $P_{\tau,s}(x)$, the sparsity is inherited by controlling τ and s; from the pinball loss $L_\tau(x)$ to $P_{\tau,s}(x)$, the noise insensitivity is inherited by adjusting τ and s.

Because the truncated pinball loss function $P_{\tau,s}(x)$ plays a key role in the minimization problem (16), it is the main motivation to further investigate this function $P_{\tau,s}(x)$ in this thesis. Our goal is to construct smoothing functions of $P_{\tau,s}(x)$ since $P_{\tau,s}(x)$ is not differentiable at $x = -s$ and $x = 0$. Therefore, we intend to study some differentiability properties of that smoothing function. Accordingly, a few algorithms can be employed along with the smoothing functions.

As will be seen in Proposition 3.1, we show that

$$P_{\tau,s}(x) = \left(\frac{1+\tau}{2}|x| \right) + \left(-\frac{\tau}{2}|x+s| + \frac{\tau s + x}{2} \right). \tag{18}$$

Note that the first term of (18) is convex and the second term of (18) is concave. Hence, the truncated pinball loss function $P_{\tau,s}(x)$ can be expressed as the sum of a convex function and a concave function. According to expression (18), we will construct a smoothing function of $P_{\tau,s}$ by taking advantage of the smoothing functions of the absolute value function. We shall denote the smoothing function for the truncated pinball loss function $P_{\tau,s}(x)$ by $\varphi_{\tau,s}(x, \mu)$ where $\mu > 0$ is the smoothing parameter. To this end, some important properties of smoothing function of $|x|$ from [1, 2, 3, 13, 14] will be used to prove our main results.

In fact, many algorithms require differentiability. In this thesis, we construct ten smoothing functions $\varphi^k_{\tau,s}(x,\mu)$ $(k=1,2,\cdots,10)$ of the truncated pinball loss function $P_{\tau,s}(x)$. Then, we just need to deal with the smooth minimization problem:

$$\min_{w\in\mathbb{R}^m, b\in\mathbb{R}} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\varphi^k_{\tau,s}((1-y_i(w^T\Phi(x_i)+b)),\mu), \tag{19}$$

for $k=1,2,\cdots,10$. Moreover, by Proposition 3.3 in Section 3, the minimization problem (19) becomes the minimization problem (16) as μ approaches 0^+. In other words, finding solutions to problem (19) will lead to solution to the target truncated pinball loss SVM (16). This is the main contribution of this thesis.

2 Preliminaries

Definition 2.1. *Given a function f: $\mathbb{R} \to \mathbb{R} \cup \{+\infty\}$. If there exists a function ϕ: $\mathbb{R} \times \mathbb{R}_{++} \to \mathbb{R} \cup \{+\infty\}$ satisfying*

(a) *ϕ is continuously differentiable at $(x, \mu) \in \mathbb{R} \times \mathbb{R}_{++}$,*

(b) $\lim_{\mu \to 0^+} \phi(x, \mu) = f(x)$,

then we say that $\phi(x, \mu)$ is a smoothing function of $f(x)$.

We present ten examples of smoothing functions for the absolute value function $|x|$ from [1, 2, 3, 13, 14] as below:

$$\varphi^1_{\text{abs}}(x, \mu) = \begin{cases} \frac{x^2}{2\mu} & \text{if } \ |x| \leq \mu, \\ |x| - \frac{\mu}{2} & \text{if } \ |x| > \mu. \end{cases} \tag{20}$$

$$\varphi^2_{\text{abs}}(x, \mu) = \mu \ln\left(\cosh(\frac{x}{\mu})\right), \text{where } \cosh(x) = \frac{e^x + e^{-x}}{2}. \tag{21}$$

$$\varphi^3_{\text{abs}}(x, \mu) = \sqrt{x^2 + \mu^2} - \mu. \tag{22}$$

$$\varphi^4_{\text{abs}}(x, \mu) = \mu[\ln(1 + e^{\frac{-x}{\mu}}) + \ln(1 + e^{\frac{x}{\mu}})]. \tag{23}$$

$$\varphi^5_{\text{abs}}(x, \mu) = \begin{cases} x & \text{if } \ x \geq \frac{\mu}{2}, \\ \frac{x^2}{\mu} + \frac{\mu}{4} & \text{if } \ \frac{-\mu}{2} < x < \frac{\mu}{2}, \\ -x & \text{if } \ x \leq \frac{-\mu}{2}. \end{cases} \tag{24}$$

$$\varphi^6_{\text{abs}}(x, \mu) = \sqrt{4\mu^2 + x^2}. \tag{25}$$

$$\varphi^7_{\text{abs}}(x, \mu) = \begin{cases} -\frac{x^4}{8\mu^3} + \frac{3x^2}{4\mu} + \frac{3\mu}{8} & \text{if } \ |x| \leq \mu, \\ |x| & \text{if } \ |x| > \mu. \end{cases} \tag{26}$$

$$\varphi^8_{\text{abs}}(x, \mu) = x\text{erf}\left(\frac{x}{\sqrt{2}\mu}\right) + \sqrt{\frac{2}{\pi}}\mu e^{\frac{-x^2}{2\mu^2}}, \text{ where } \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{27}$$

$$\varphi^9_{\text{abs}}(x, \mu) = x\tanh\left(\frac{x}{\mu}\right), \text{ where } \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{28}$$

$$\varphi^{10}_{\text{abs}}(x, \mu) = x\text{erf}\left(\frac{x}{\mu}\right), \text{ where } \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{29}$$

In general, the above ten smoothing functions of $|x|$ can be constructed via three methods.

(A) : Convex Conjugate Method,

(B) : Convolution Method,

(C) : Using Sigmoid Functions.

More specifically, smoothing functions $\varphi^1_{\rm abs}(x,\mu)$, $\varphi^2_{\rm abs}(x,\mu)$, and $\varphi^3_{\rm abs}(x,\mu)$ are produced by convex conjugate method. Smoothing functions $\varphi^1_{\rm abs}(x,\mu)$, $\varphi^4_{\rm abs}(x,\mu)$, $\varphi^5_{\rm abs}(x,\mu)$, $\varphi^6_{\rm abs}(x,\mu)$, $\varphi^7_{\rm abs}(x,\mu)$, and $\varphi^8_{\rm abs}(x,\mu)$ can be produced convolution method. Smoothing functions $\varphi^9_{\rm abs}(x,\mu)$ and $\varphi^{10}_{\rm abs}(x,\mu)$ are constructed by using sigmoid functions $\tanh(x)$ and $\mathrm{erf}(x)$, respectively.

To illustrate how the above ten smoothing functions are obtained, we briefly review their concepts.

(A) Convex conjugate method. For more details, please refer to [4, 14, 21].

Definition 2.2. *A proper function d: $C \to \mathbb{R}\cup\{+\infty\}$, The function d is called σ-strongly convex with respect to $\|\cdot\|$ if there exists a constant $\sigma > 0$ (often called the modulus of strong convexity) such that $d((1-t)x+ty) \le (1-t)d(x)+td(y)$, for all $x, y \in C$ and $t \in (0,1)$.*

Definition 2.3. *For any extended real-valued function $f : \mathrm{dom}(f) \to \mathbb{R}\cup\{+\infty\}$, its convex conjugate $f^* : (\mathrm{dom}(f))^* \to \mathbb{R}\cup\{+\infty\}$ is defined by $f^*(x) = \sup\limits_{z\in \mathrm{dom}(f)} \{z^T x - f(z)\}$. for some 1-strongly convex and continuous function $d(\cdot)$ (called proximity function).*

The smooth approximation, denoted by f_μ, of f by convex conjugate method can be built up, i.e.,

$$f_\mu(x) := (g+\mu d)^*(x) = \sup_{z\in\mathrm{dom}(g+\mu d)} \{z^T x - g(z) - \mu d(z)\} = \sup_{z\in\mathrm{dom}(g)} \{z^T x - g(z) - \mu d(z)\},$$

where $g = f^*$. We demonstrate indicator function [12] to show up the construction procedure. First, the indicator function $\delta(\cdot \,|\, C)$ of C is

$$\delta(x|\ C) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{if } x \notin C. \end{cases}$$

By taking different proximity function d, it yields various smoothing functions.

(i) Take $g(z) = \delta(z \,|\, [-1,1])$, where g is the conjugate function of $|\cdot|$. Choose $d(z) = \frac{z^2}{2}$

as a proximity function. Then, we have

$$
\begin{aligned}
(g+\mu d)^*(x) &= \sup_{z\in \mathrm{dom}(g)} \{z^T x - g(z) - \mu d(z)\} \\
&= \sup_{|z|\le 1} \{z^T x - 0 - \mu \frac{z^2}{2}\} \\
&= \begin{cases} \dfrac{x^2}{2\mu} & \text{if } \ |x| \le \mu \\ |x| - \dfrac{\mu}{2} & \text{if } \ |x| > \mu \end{cases} \\
&= \varphi^1_{\text{abs}}(x,\mu)
\end{aligned}
$$

(ii) Take $g(z) = \delta(z \,|\, [-1,1])$ again, and choose $d(z) = 1 - \sqrt{1-z^2}$ as a proximity function. Then, we have

$$
\begin{aligned}
(g+\mu d)^*(x) &= \sup_{z\in \mathrm{dom}(g)} \{z^T x - g(z) - \mu d(z)\} \\
&= \sup_{|z|\le 1} \{z^T x - 0 - \mu(1 - \sqrt{1-z^2})\} \\
&= \sqrt{x^2+\mu^2} - \mu \\
&= \varphi^3_{\text{abs}}(x,\mu)
\end{aligned}
$$

(iii) Take $g(z) = \delta(z \,|\, C)$ with $C = \{z \ge 0 \,|\, z_1 + z_2 = 1\}$ Note that $|x| = g^*(Ax)$ with $A = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Now, choose $d(z) = z_1 \ln(z_1) + z_2 \ln(z_2) + \ln(2)$ as a proximity function. Then, we have

$$
\begin{aligned}
(g+\mu d)^*(Ax) &= \sup_{z\in \mathrm{dom}(g)} \{z^T(Ax) - g(z) - \mu d(z)\} \\
&= \sup_{z\ge 0, z_1+z_2=1} \{(z_1 x + z_2(-x)) - 0 - \mu(z_1 \ln(z_1) + z_2 \ln(z_2) + \ln(2))\} \\
&= \mu \ln\left(\cosh\left(\frac{x}{\mu}\right)\right) \\
&= \varphi^2_{\text{abs}}(x,\mu)
\end{aligned}
$$

(B) Convolution method. Please see [5, 14, 23] and [6, Section 11.8.2] for more details.

Definition 2.4. *A function $\rho : \mathbb{R} \to \mathbb{R}_+$ is a piecewise continuous density function (with finitely many discontinuous points) if $\int_{-\infty}^{\infty} \rho(t)dt = 1$ and $\int_{-\infty}^{\infty} |t|\rho(t)dt < +\infty$.*

The convolution method on constructing smoothing functions of $|x|$ is via

$$
f_\mu(x) := \frac{1}{\mu} \int_{-\infty}^{\infty} |x - \mu t| \rho(\frac{t}{\mu}) dt. \tag{30}
$$

the density function	smoothing function of $|x|$
ρ_1	φ^1_{abs}
ρ_2	φ^4_{abs}
ρ_3	φ^5_{abs}
ρ_4	φ^6_{abs}
ρ_5	φ^7_{abs}
ρ_6	φ^8_{abs}

Table 1: The density function ρ_i and its corresponding smoothing function

In fact, the formula (30) uses the smoothing function of plus function $(x)_+ = \max\{0, x\}$. To see this, we know that the absolute function $|x|$ is equal to the sum $(x)_+ + (x)_-$. Then, applying a smoothing function $\varphi_{\text{plus}}(x, \mu)$ of plus function, we can acquire a smoothing function of $|x|$ by $\varphi_{\text{plus}}(x, \mu) + \varphi_{\text{plus}}(-x, \mu)$. As below, we demonstrate six constructions by taking six different density function.

(i) For case of φ^1_{abs}, taking $\rho_1(t) = \begin{cases} 1 & \text{if} \quad 0 \le t \le 1 \\ 0 & \text{if} \quad t < 0 \text{ or } t > 1 \end{cases}$, then we obtain the smoothing function of absolute value function.

(ii) For case of φ^4_{abs}, taking $\rho_2(t) = \frac{e^{-t}}{(1+e^{-t})^2}$, then we obtain the smoothing function of the absolute value function.

(iii) For case of φ^5_{abs}, taking $\rho_3(t) = \begin{cases} 1 & \text{if} \quad -\frac{1}{2} \le t \le \frac{1}{2} \\ 0 & \text{if} \quad t < 0 \text{ or } t > \frac{1}{2} \end{cases}$, then we obtain the smoothing function of the absolute value function.

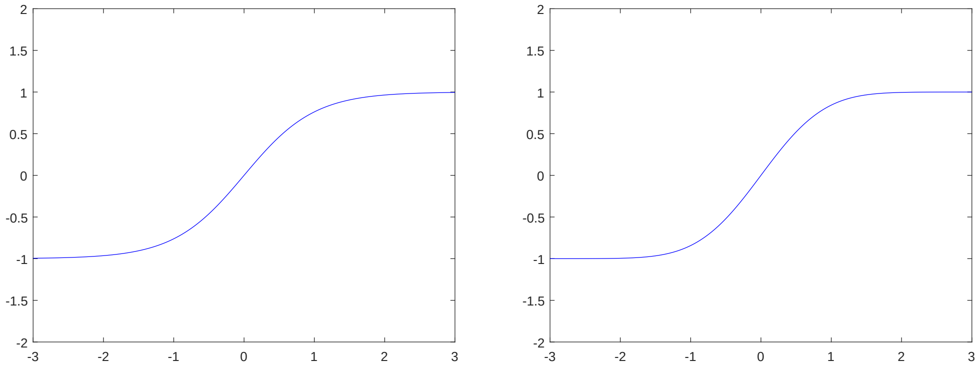
(iv) For case of φ^6_{abs}, taking $\rho_4(t) = \frac{2}{(t^2+4)^{\frac{3}{2}}}$, then we obtain the smoothing function of the absolute value function.

(v) For case of φ^7_{abs}, taking $\rho_5(t) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{if} \quad |t| < 1 \\ 0 & \text{if} \quad |t| \ge 1 \end{cases}$, and then we get the smoothing function of the absolute value function.

(vi) For case of φ^8_{abs}, taking $\rho_6(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$, then we obtain the smoothing function of the absolute value function.

To sum up, here is a table of the density function ρ_i and its corresponding smoothing function of $|x|$. See Table 1.

(C) Using sigmoid functions. A sigmoid function is bounded and differentiable real function that is defined for all real input values and has a non-negative derivative at each point, see [3, 8]. It is known that both $\tanh(x)$ and $\text{erf}(x)$ are sigmoid functions, whose graphs are shown in Figure 10.



(a) $\tanh(x)$ (b) $\text{erf}(x)$

Figure 10: Two important sigmoid functions

Inspired by an idea of smoothing technique from [3, pp. 13]:

$$\lim_{|x|\to\infty} \tanh(x) = \text{sgn}(x) \quad \text{and} \quad \lim_{|x|\to\infty} \text{erf}(x) = \text{sgn}(x).$$

Then, we can verify that

$$\lim_{\mu\to 0^+} \tanh\left(\frac{x}{\mu}\right) = \text{sgn}(x) \quad \text{and} \quad \lim_{\mu\to 0^+} \text{erf}\left(\frac{x}{\mu}\right) = \text{sgn}(x).$$

Now, using the fact that $|x| = x \cdot \text{sgn}(x)$, we have

$$\lim_{\mu\to 0^+} x \cdot \tanh\left(\frac{x}{\mu}\right) = x \lim_{\mu\to 0^+} \tanh\left(\frac{x}{\mu}\right) = x \cdot \text{sgn}(x) = |x|,$$

and

$$\lim_{\mu\to 0^+} x \cdot \text{erf}\left(\frac{x}{\mu}\right) = x \lim_{\mu\to 0^+} \text{erf}\left(\frac{x}{\mu}\right) = x \cdot \text{sgn}(x) = |x|.$$

By this way, $\varphi^{10}_{\text{abs}}(x,\mu) = x \cdot \text{erf}\left(\frac{x}{\mu}\right)$ can be constructed. We point out that this way also provides an alternative construction for $\varphi^{10}_{\text{abs}}(x,\mu)$ since $\varphi^{8}_{\text{abs}}(x,\mu) = x \cdot \text{erf}\left(\frac{x}{\sqrt{2}\mu}\right) + \sqrt{\frac{2}{\pi}}\mu e^{-\frac{x^2}{2\mu^2}}$. Please refer to [23, (2.16)] for more details.

Proposition 2.1. *Let $\varphi^{k}_{\text{abs}}(x,\mu)$ for $k = 1, 2, \cdots, 10$ be defined as in* (20)-(29)*, respectively. Then, $\varphi^{k}_{\text{abs}}(x,\mu)$ is continuously differentiable for all $(x,\mu) \in \mathbb{R} \times \mathbb{R}_{++}$.*

Proof. (i) For $i = 1, \ldots, 8$, the result can be obtained from [13, Proposition 2.1(a)] and [14, Proposition 2.1(a)].

(ii) Based on [1, Theorem 1], we know that

$$\nabla_x \varphi^{9}_{\text{abs}}(x,\mu) = \frac{x}{\mu}\text{sech}^2\left(\frac{x}{\mu}\right) + \tanh\left(\frac{x}{\mu}\right),$$

which is continuous. In addition, $\nabla_\mu \varphi^9_{\rm abs}(x,\mu) = -\left[\frac{x}{\mu}{\rm sech}(\frac{x}{\mu})\right]^2$ is also continuous. Thus, $\varphi^9_{\rm abs}(x,\mu)$ is continuously differentiable.

According to [3, Proposition 2.1], we know that

$$\nabla_x \varphi^{10}_{abs}(x,\mu) = \frac{2x}{\sqrt{\pi}\mu} e^{-\frac{x^2}{\mu^2}} + {\rm erf}\left(\frac{x}{\mu}\right)$$

which is continuous. Besides, $\nabla_\mu \varphi^{10}_{abs}(x,\mu) = -\frac{2}{\sqrt{\pi}}\left[\frac{x}{\mu}e^{-\frac{x}{\mu}}\right]^2$ is also continuous. Thus, $\varphi^{10}_{\rm abs}(x,\mu)$ is continuously differentiable. □

Lemma 2.1. *Let $\varphi^k_{\rm abs}(x,\mu)$ for $k = 1,2,\cdots,10$ be defined as in (20)-(29), respectively. Then, there exists $C_k > 0$ such that*

$$\left|\varphi^k_{\rm abs}(x,\mu) - |x|\right| \le C_k\,\mu$$

for all $x \in \mathbb{R}$ and for all positive parameter μ.

Proof. We verify the desired inequality case by case.

(1) From [23, Lemma 2.2], it is clear that

$$\left|\varphi^1_{\rm abs}(x,\mu) - |x|\right| \le \frac{1}{2}\mu.$$

(2) The desired result follows by

$$\begin{aligned}
\left|\varphi^2_{\rm abs}(x,\mu) - |x|\right| &= \left|\mu{\rm ln}\left(\cosh\left(\frac{x}{\mu}\right)\right) - \mu\frac{|x|}{\mu}\right| \\
&= \mu\left|{\rm ln}\left(\frac{e^{\frac{x}{\mu}} + e^{\frac{-x}{\mu}}}{2}\right) - {\rm ln}\left(e^{\frac{|x|}{\mu}}\right)\right| \\
&= \mu\left|{\rm ln}\left(e^{\frac{|x|}{\mu}}\right) - {\rm ln}\left(\frac{e^{\frac{x}{\mu}} + e^{\frac{-x}{\mu}}}{2}\right)\right| \\
&= \mu\left|{\rm ln}\left(\frac{2e^{\frac{|x|}{\mu}}}{2}\right) - {\rm ln}\left(\frac{e^{\frac{x}{\mu}} + e^{\frac{-x}{\mu}}}{2}\right)\right| \\
&= \mu\left({\rm ln}\left(\frac{2e^{\frac{|x|}{\mu}}}{2}\right) - {\rm ln}\left(\frac{e^{\frac{x}{\mu}} + e^{\frac{-x}{\mu}}}{2}\right)\right) \\
&= \mu{\rm ln}\left(\frac{2e^{\frac{|x|}{\mu}}}{e^{\frac{x}{\mu}} + e^{\frac{-x}{\mu}}}\right) \\
&\le \mu{\rm ln}\left(\frac{2e^{\frac{|x|}{\mu}}}{e^{\frac{|x|}{\mu}} + 0}\right) \\
&= ({\rm ln}2)\mu.
\end{aligned}$$

(3) Using [23, Lemma 2.1] yields $\left|\sqrt{x^2+\mu^2}-|x|\right| \le \mu$. Hence, we have

$$\begin{aligned}
\left|\varphi^3_{\mathrm{abs}}(x,\mu)-|x|\right| &= \left|(\sqrt{x^2+\mu^2}-\mu)-|x|\right| \\
&\le \left|\sqrt{x^2+\mu^2}-|x|\right| + |-\mu| \\
&\le (1+1)\mu \\
&= 2\mu.
\end{aligned}$$

(4) The desired result follows by

$$\begin{aligned}
\left|\varphi^4_{\mathrm{abs}}-|x|\right| &= \mu\left|\left[\ln\left(1+e^{\frac{-x}{\mu}}\right)+\ln\left(1+e^{\frac{x}{\mu}}\right)\right]-\frac{|x|}{\mu}\right| \\
&= \mu\left|\ln\left(\frac{2+e^{\frac{x}{\mu}}+e^{\frac{-x}{\mu}}}{e^{\frac{|x|}{\mu}}}\right)\right| \\
&\le \mu\left|\ln\left(\frac{2}{e^{\frac{|x|}{\mu}}}+\frac{e^{\frac{x}{\mu}}+e^{\frac{-x}{\mu}}}{e^{\frac{|x|}{\mu}}}\right)\right| \\
&\le \mu\cdot\ln\left(\frac{2}{e^0}+\frac{2e^{\frac{|x|}{\mu}}}{e^{\frac{|x|}{\mu}}}\right) \\
&= \mu\ln(2+2) = (\ln 4)\mu.
\end{aligned}$$

(5) For $|x| < \frac{\mu}{2}$, we have

$$\begin{aligned}
\left|\varphi^5_{\mathrm{abs}}(x,\mu)-|x|\right| &\le \left|\left(\frac{x^2}{\mu}+\frac{\mu}{4}\right)-|x|\right| \\
&\le \left|\frac{(\frac{\mu}{2})^2}{\mu}+\frac{\mu}{4}-0\right| \\
&= \left|\frac{\mu}{4}+\frac{\mu}{4}-0\right| \\
&= \frac{1}{2}\mu.
\end{aligned}$$

For $|x| \ge \frac{\mu}{2}$, there holds

$$\left|\varphi^5_{\mathrm{abs}}(x,\mu)-|x|\right| = ||x|-|x|| = 0.$$

(6) In light of [23, Lemma 2.1], we know

$$\left|\varphi^6_{\mathrm{abs}}(x,\frac{\mu}{2})-|x|\right| \le \mu.$$

Hence, we have

$$\left|\varphi^6_{\mathrm{abs}}(x,\mu)-|x|\right| = 2\mu.$$

(7) For $|x| > \mu$, we have

$$\left|\varphi_{\text{abs}}^7(x,\mu) - |x|\right| = ||x| - |x|| = 0 < \frac{11}{8}\mu.$$

For $|x| \le \mu$, we have

$$\begin{aligned}\left|\varphi_{\text{abs}}^7(x,\mu) - |x|\right| &= \left|\left(-\frac{x^4}{8\mu^3} + \frac{3x^2}{4\mu} + \frac{3\mu}{8}\right) - |x|\right| \\ &\le \left|-\frac{x^4}{8\mu^3} + \frac{3x^2}{4\mu}\right| + \left|\frac{3\mu}{8} - |x|\right|.\end{aligned}$$

Now, applying

$$\left|\frac{3\mu}{8} - |x|\right| \le \left|\frac{3\mu}{8} - \mu\right| = \frac{5}{8}\mu$$

and

$$\begin{aligned}\left|-\frac{x^4}{8\mu^3} + \frac{3x^2}{4\mu}\right| &= \frac{x^2}{\mu}\left|\frac{3}{4} - \frac{x^2}{8\mu^2}\right| \\ &\le \frac{\mu^2}{\mu}\left|\frac{3}{4} - \frac{0^2}{8\mu^2}\right| \\ &= \frac{3}{4}\mu,\end{aligned}$$

we obtain

$$\left|\varphi_{\text{abs}}^7(x,\mu) - |x|\right| \le \left(\frac{5}{8} + \frac{3}{4}\right)\mu = \frac{11}{8}\mu.$$

(8) From [3, Proposition 2.1], we have

$$\begin{aligned}\left|\varphi_{\text{abs}}^8(x,\mu) - |x|\right| &\le \left|x \operatorname{erf}\left(\frac{x}{\sqrt{2}\mu}\right) - |x|\right| + \left|\sqrt{\frac{2}{\pi}}\mu e^{\frac{-x^2}{2\mu^2}}\right| \\ &= \sqrt{2}\left|\frac{x}{\sqrt{2}} \operatorname{erf}\left(\frac{\left(\frac{x}{\sqrt{2}}\right)}{\mu}\right) - \left|\frac{x}{\sqrt{2}}\right|\right| + \left|\sqrt{\frac{2}{\pi}}\mu e^{\frac{-x^2}{2\mu^2}}\right| \\ &\le \sqrt{2} \cdot \frac{2}{(e\sqrt{\pi})}\mu + \sqrt{\frac{2}{\pi}}\mu \\ &= \left(\sqrt{2} \cdot \frac{2}{(e\sqrt{\pi})} + \sqrt{\frac{2}{\pi}}\right)\mu \\ &= \left(\sqrt{\frac{2}{\pi}}\left(\frac{2}{e} + 1\right)\right)\mu.\end{aligned}$$

(9) Using [2, Theorem 1], it is clear that

$$\left|\varphi_{\text{abs}}^{9}(x,\mu)-|x|\right|=\left|x\tanh\left(\frac{x}{\mu}\right)-|x|\right|\leq\mu.$$

(10) Using [3, Proposition 2.1], it is clear that

$$\left|\varphi_{abs}^{10}(x,\mu)-|x|\right|=\left|x\,\mathrm{erf}\left(\frac{x}{\mu}\right)-|x|\right|\leq\frac{2}{(e\sqrt{\pi})}\mu.$$

$\square$

Proposition 2.2. *Let* $\varphi_{\text{abs}}^{k}(x,\mu)$ *for* $k=1,2,\cdots,10$ *be defined as in* (20)-(29)*, respectively. Then, we have* $\lim\limits_{\mu\to 0^+,}\varphi_{abs}^{k}(x,\mu)=|x|$*, for all* $x\in\mathbb{R}$.

Proof. By Lemma 2.1, it implies $\lim\limits_{\mu\to 0^+}|\varphi_{abs}^{k}(x,\mu)-|x||\leq\lim\limits_{\mu\to 0^+}C_k\mu=0$. $\square$

From Proposition 2.1 and Proposition 2.2, we see that $\varphi_{\text{abs}}^{k}(x,\mu)$ for $k=1,2,\cdots,10$ defined as in (20)-(29), are smoothing functions of absolute value function $|x|$. For computational purpose, we write down those gradient functions of smoothing functions $\varphi_{\text{abs}}^{k}(x,\mu)$ for $k=1,2,\cdots,10$, which can be found in [1, 2, 3, 13, 14]. These will be used to prove Lemma 3.1 as well.

$$\nabla\varphi_{\text{abs}}^{1}(x,\mu)=\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix},\tag{31}$$

where

$$\alpha_1=\begin{cases}1 & \text{if} \quad x>\mu,\\ \frac{x}{\mu} & \text{if} \quad -\mu\leq x\leq\mu,\\ -1 & \text{if} \quad x<-\mu.\end{cases}\quad\text{and}\quad \alpha_2=\begin{cases}-\frac{1}{2} & \text{if} \quad x>\mu,\\ -\frac{1}{2}(\frac{x}{\mu})^2 & \text{if} \quad -\mu\leq x\leq\mu,\\ -\frac{1}{2} & \text{if} \quad x<-\mu.\end{cases}$$

$$\nabla\varphi_{\text{abs}}^{2}(x,\mu)=\begin{pmatrix}\tanh(\frac{x}{\mu})\\ \ln(\cosh(\frac{x}{\mu}))-\frac{x}{\mu^2}\tanh(\frac{x}{\mu})\end{pmatrix}.\tag{32}$$

$$\nabla\varphi_{\text{abs}}^{3}(x,\mu)=\begin{pmatrix}\frac{x}{\sqrt{x^2+\mu^2}}\\ \frac{\mu}{\sqrt{x^2+\mu^2}}-1\end{pmatrix}.\tag{33}$$

$$\nabla\varphi_{\text{abs}}^{4}(x,\mu)=\begin{pmatrix}\frac{1}{1+e^{-\frac{x}{\mu}}}-\frac{1}{1+e^{\frac{x}{\mu}}}\\ \left[\ln(1+e^{-\frac{x}{\mu}})+\ln(1+e^{\frac{x}{\mu}})\right]+\frac{x}{\mu}\left[\frac{-1}{1+e^{-\frac{x}{\mu}}}+\frac{1}{1+e^{\frac{x}{\mu}}}\right]\end{pmatrix}.\tag{34}$$

$$\nabla\varphi_{\text{abs}}^{5}(x,\mu)=\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix},\tag{35}$$

where

$$\beta_1 = \begin{cases} 1 & \text{if} \quad x \geq \frac{\mu}{2}, \\ \frac{2x}{\mu} & \text{if} \quad -\frac{\mu}{2} < x < \frac{\mu}{2}, \\ -1 & \text{if} \quad x \leq -\frac{\mu}{2}. \end{cases} \quad \text{and} \quad \beta_2 = \begin{cases} 0 & \text{if} \quad x \geq \frac{\mu}{2}, \\ -(\frac{x}{\mu})^2 + \frac{1}{4} & \text{if} \quad -\frac{\mu}{2} < x < \mu < \frac{\mu}{2}, \\ 0 & \text{if} \quad x \leq -\frac{\mu}{2}. \end{cases}$$

$$\nabla \varphi^6_{\text{abs}}(x, \mu) = \begin{pmatrix} \frac{x}{\sqrt{4\mu^2+x^2}} \\ \frac{4\mu}{\sqrt{4\mu^2+x^2}} \end{pmatrix}. \tag{36}$$

$$\nabla \varphi^7_{\text{abs}}(x, \mu) = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}, \tag{37}$$

where

$$\gamma_1 = \begin{cases} 1 & \text{if} \quad x > \mu \\ -\frac{x^3}{2\mu^3} + \frac{3x}{2\mu} & \text{if} \quad -\mu \leq x \leq \mu \;, \\ -1 & \text{if} \quad x < -\mu \end{cases}$$

$$\gamma_2 = \begin{cases} 0 & \text{if} \quad x > \mu \\ \frac{3x^4}{8\mu^4} - \frac{3x^2}{4\mu^2} + \frac{3}{8} & \text{if} \quad -\mu \leq x \leq \mu \;. \\ 0 & \text{if} \quad x < -\mu \end{cases}$$

$$\nabla \varphi^8_{\text{abs}}(x, \mu) = \begin{pmatrix} \text{erf}(\frac{x}{\sqrt{2}\mu}) \\ \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2\mu^2}} \end{pmatrix}. \tag{38}$$

$$\nabla \varphi^9_{\text{abs}}(x, \mu) = \begin{pmatrix} \frac{x}{\mu}\text{sech}^2(\frac{x}{\mu}) + \tanh(\frac{x}{\mu}) \\ -\left[\frac{x}{\mu}\text{sech}(\frac{x}{\mu})\right]^2 \end{pmatrix}. \tag{39}$$

$$\nabla \varphi^{10}_{\text{abs}}(x, \mu) = \begin{pmatrix} \frac{2x}{\sqrt{\pi}\mu} e^{-\frac{x^2}{\mu^2}} + \text{erf}(\frac{x}{\mu}) \\ -\frac{2}{\sqrt{\pi}} \left[\frac{x}{\mu} e^{-\frac{x}{\mu}}\right]^2 \end{pmatrix}. \tag{40}$$

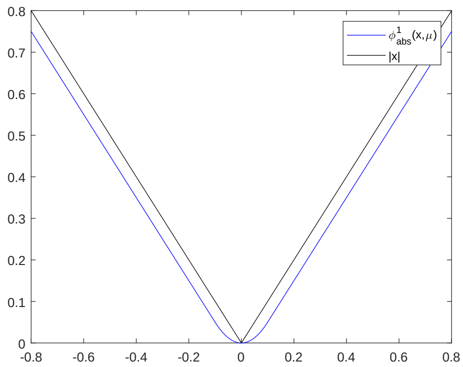
Next, we talk about some features of $\varphi^k_{\text{abs}}(x, \mu)$. Geometric views are depicted in Figure 11, Figure 12, and Figure 13. Below are for algebraic checking.

(i) We know that $\nabla_x \varphi^k_{\text{abs}}(x, \mu)$ for $k = 1, 2$ for fixed $\mu > 0$ are non-decreasing. From [20, Corollary 2.1], we know $\varphi^k_{\text{abs}}(x, \mu)$ for $k = 1, 2$ for fixed $\mu > 0$ are convex.

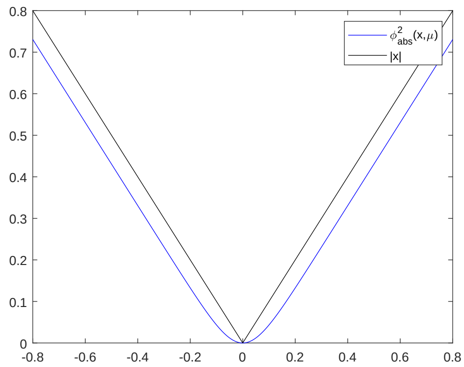
(ii) We observe that $\nabla_{xx} \varphi^k_{\text{abs}}(x, \mu)$ for $k = 3, 4$ for fixed $\mu > 0$ are non-negative. Thus, from [20, Corollary 2.1], we know $\varphi^k_{\text{abs}}(x, \mu)$ for $k = 3, 4$ for fixed $\mu > 0$ are convex.

(iii) It is known that $\nabla_x \varphi^k_{\text{abs}}(x, \mu)$ for $k = 5, 8$ for fixed $\mu > 0$ are non-decreasing. Based on [20, Corollary 2.1], we have $\varphi^k_{\text{abs}}(x, \mu)$ for $k = 5, 8$ for fixed $\mu > 0$ are convex.

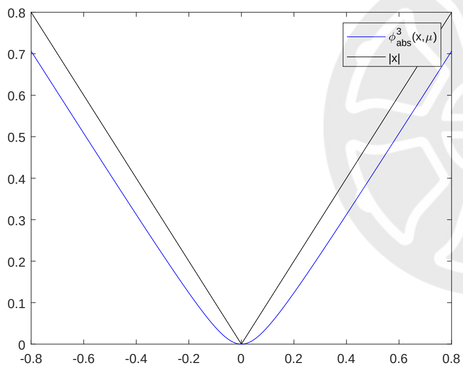
(iv) We know that $\nabla_{xx} \varphi^k_{\text{abs}}(x, \mu)$ for $k = 6, 7$ for fixed $\mu > 0$ are non-negative. In light of [20, Corollary 2.1] again, we have $\varphi^k_{\text{abs}}(x, \mu)$ for $k = 6, 7$ for fixed $\mu > 0$ are convex.



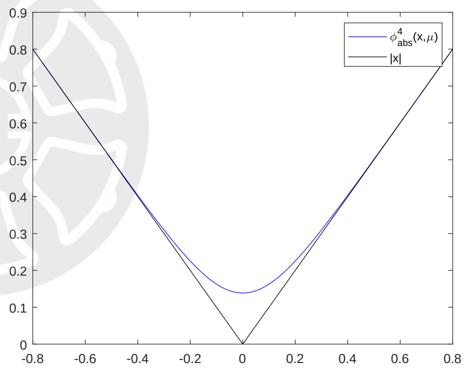
(a) $|x|$ and $\varphi^1_{abs}(x,\mu)$:convex function



(b) $|x|$ and $\varphi^2_{abs}(x,\mu)$



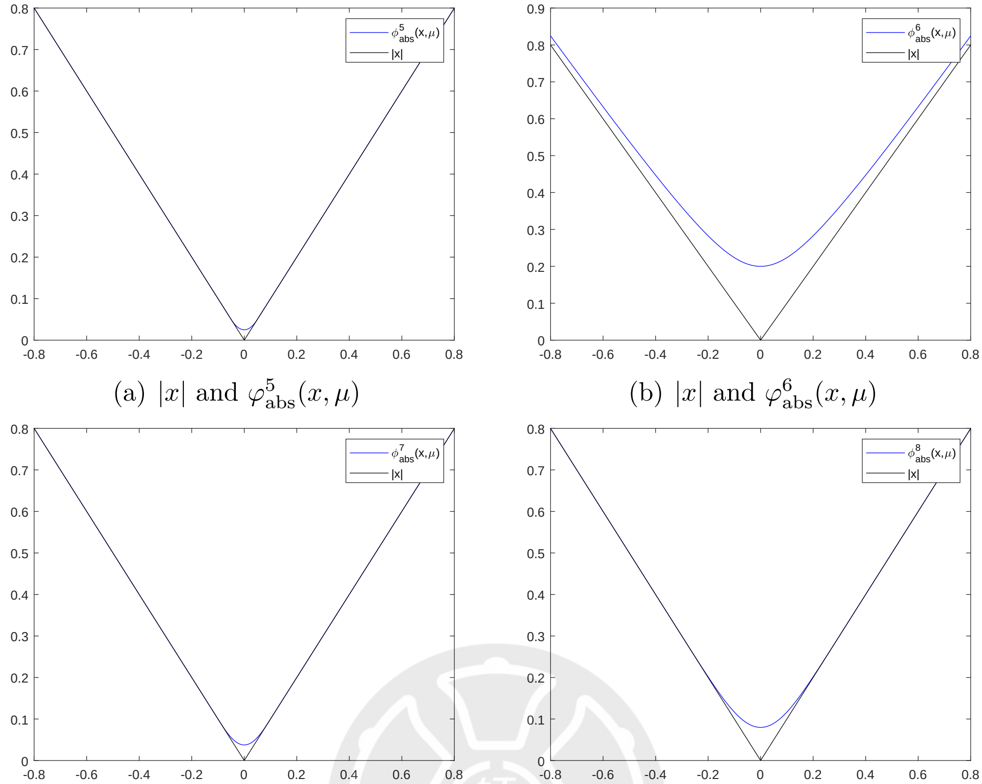
(c) $|x|$ and $\varphi^3_{abs}(x,\mu)$



(d) $|x|$ and $\varphi^4_{\text{abs}}(x,\mu)$

Figure 11: $|x|$ and $\varphi^k_{\text{abs}}(x,\mu)$ for $k = 1, 2, 3, 4$ with $\mu = 0.1$

(v) Because both $\nabla_x \varphi^9_{\mathrm{abs}}(x,\mu)$ and $\nabla_x \varphi^{10}_{\mathrm{abs}}(x,\mu)$ with $\mu = 0.1$ are not non-decreasing, neither $\varphi^9_{abs}(x,\mu)$ nor $\varphi^{10}_{abs}(x,\mu)$ is not convex. See Figure 14.



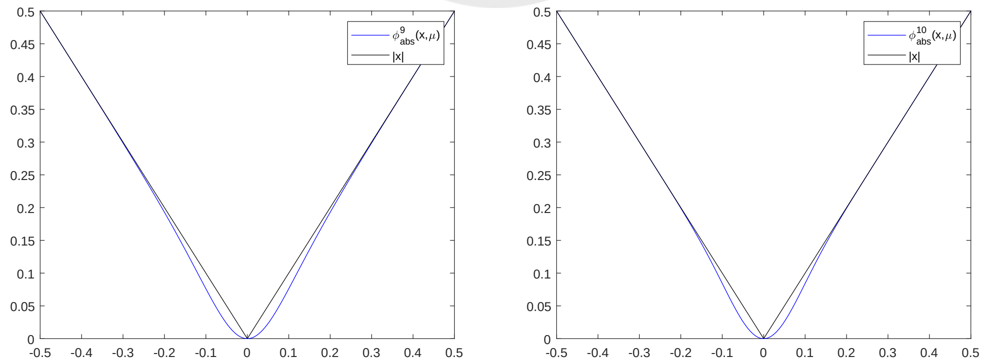
(a) $|x|$ and $\varphi^5_{\mathrm{abs}}(x,\mu)$

(b) $|x|$ and $\varphi^6_{\mathrm{abs}}(x,\mu)$

(c) $|x|$ and $\varphi^7_{\mathrm{abs}}(x,\mu)$

(d) $|x|$ and $\varphi^8_{\mathrm{abs}}(x,\mu)$

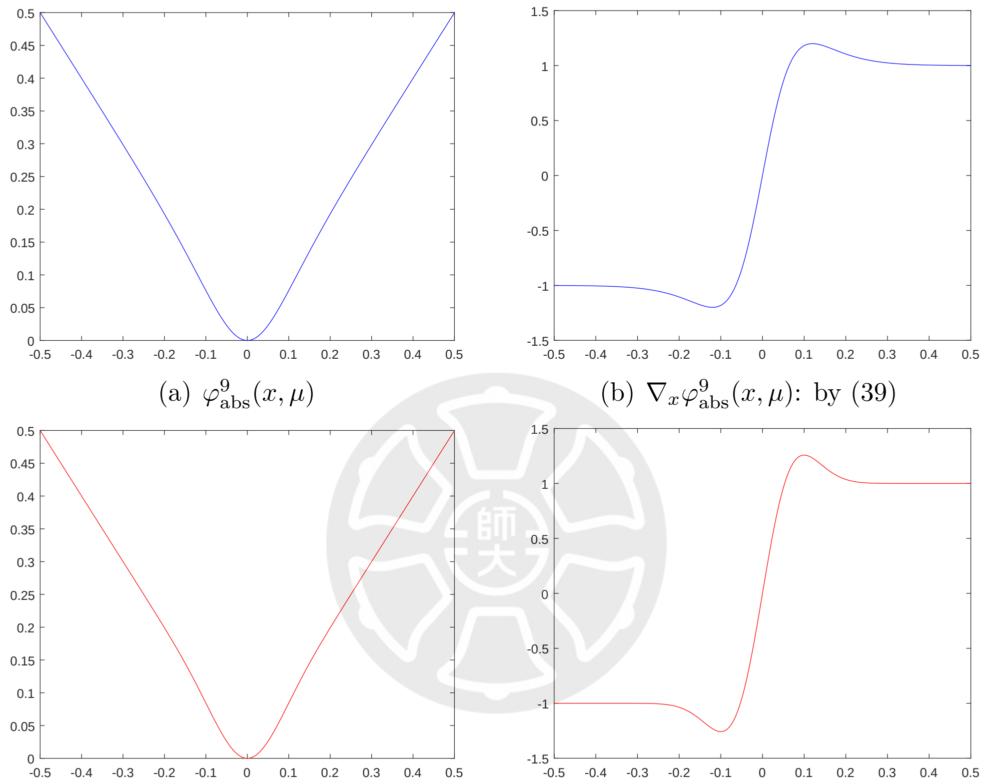
Figure 12: $|x|$ and $\varphi^k_{\mathrm{abs}}(x,\mu)$ for $k = 5,6,7,8$ with $\mu = 0.1$



(a) $\varphi^9_{\mathrm{abs}}(x,\mu)$

(b) $\varphi^{10}_{\mathrm{abs}}(x,\mu)$

Figure 13: $\varphi^k_{\mathrm{abs}}(x,\mu)$ is non-convex for $k = 9,10$ with $\mu = 0.1$



(a) $\varphi^{9}_{\text{abs}}(x,\mu)$

(b) $\nabla_x \varphi^{9}_{\text{abs}}(x,\mu)$: by (39)

(c) $\varphi^{10}_{\text{abs}}(x,\mu)$

(d) $\nabla_x \varphi^{10}_{\text{abs}}(x,\mu)$: by (40)

Figure 14: $\varphi^{k}_{\text{abs}}(x,\mu)$ and $\nabla_x \varphi^{k}_{\text{abs}}(x,\mu)$ for $k = 9, 10$ with $\mu = 0.1$

3 Main Results

This section is devoted to deriving smoothing functions for the truncated pinball loss function $P_{\tau,s}(x)$. The following proposition says that the truncated pinball loss function can be expressed as the sum of two absolute functions and an affine function.

Proposition 3.1. *Let $P_{\tau,s}(x)$ be defined as in* (14). *There is an alternative expression for $P_{\tau,s}(x)$, i.e., $P_{\tau,s}(x) = -\frac{\tau}{2}|x+s| + \frac{1+\tau}{2}|x| + \frac{\tau s+x}{2}$.*

Proof. From the definition of the truncated pinball loss function $P_{\tau,s}$, we express $P_{\tau,s}(x)$ as the sum of two continuous functions. In particular, we can express $P_{\tau,s}(x)$ as $f_1(x)+f_2(x)$, where

$$f_1(x) = \begin{cases} x + \dfrac{\tau x}{2} & \text{if } \ x \geq 0, \\ -\dfrac{\tau x}{2} & \text{if } \ x < 0. \end{cases} \quad \text{and} \quad f_2(x) = \begin{cases} -\dfrac{\tau x}{2} & \text{if } \ x > -s, \\ \tau s + \dfrac{\tau x}{2} & \text{if } \ x \leq -s. \end{cases}$$

The first continuous function $f_1(x)$ is a piecewise function, which has two pieces: $x \geq 0$ and $x < 0$. The second continuous function $f_2(x)$ is a piecewise function, which has two pieces: $x > -s$ and $x \leq -s$. The graphs of f_1, f_2 and $P_{\tau,s}$ are shown in Figure 15 and Figure 16. These pictures motivate us an idea that $P_{\tau,s}(x)$ is the sum of a convex function and a concave function. Therefore, we have

$$f_1(x) = \begin{cases} x + \dfrac{\tau x}{2} & \text{if } \ x \geq 0 \\ -\dfrac{\tau x}{2} & \text{if } \ x < 0 \end{cases} = \begin{cases} \dfrac{x}{2} + \dfrac{(1+\tau)x}{2} & \text{if } \ x \geq 0 \\ \dfrac{x}{2} - \dfrac{(1+\tau)x}{2} & \text{if } \ x < 0 \end{cases} = \frac{x}{2} + \frac{1+\tau}{2}|x|,$$

and

$$f_2(x) = \begin{cases} -\dfrac{\tau x}{2} & \text{if } \ x > -s \\ \tau s + \dfrac{\tau x}{2} & \text{if } \ x \leq -s \end{cases} = \begin{cases} \dfrac{\tau s}{2} - \dfrac{(\tau(x+s)}{2} & \text{if } \ x > -s \\ \dfrac{\tau s}{2} + \dfrac{(\tau(x+s)}{2} & \text{if } \ x \leq -s \end{cases} = \frac{\tau s}{2} - \frac{\tau}{2}|x+s|.$$

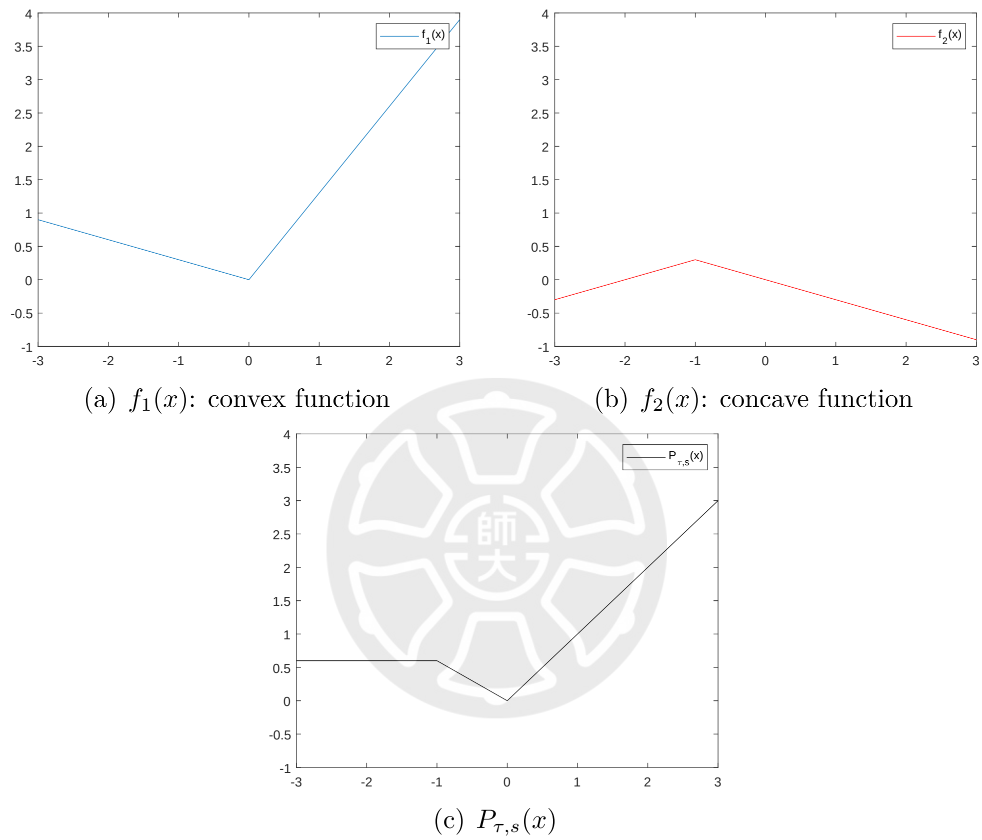
Thus, the proof is complete. $\square$

Observe that Proposition 3.1 expresses the truncated pinball loss function $P_{\tau,s}(x)$ in terms of the absolute value function and an affine function. Moreover, it is equivalent to say

$$P_{\tau,s}(x) = -\frac{\tau}{2}|x+s| + \frac{1+\tau}{2}|x| + \frac{\tau s + x}{2} = \left(\frac{1+\tau}{2}|x|\right) + \left(-\frac{\tau}{2}|x+s| + \frac{\tau s + x}{2}\right).$$

From which, we observe two facts:

(i) The first term $\frac{1+\tau}{2}|x|$ is convex.



(a) $f_1(x)$: convex function

(b) $f_2(x)$: concave function

(c) $P_{\tau,s}(x)$

Figure 15: $f_1(x) + f_2(x) = P_{\tau,s}(x)$ with $\tau = 0.6, s = 1$

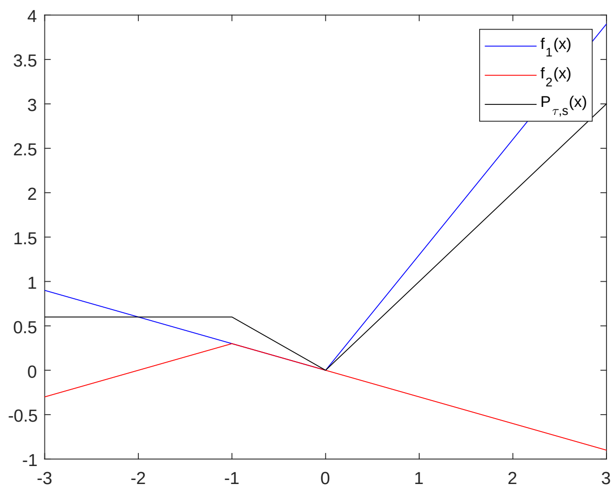


Figure 16: Graphs of $P_{\tau,s}(x)$, $f_1(x)$, and $f_2(x)$ with $\tau = 0.6, s = 1$

(ii) The second term $-\frac{\tau}{2}|x+s| + \frac{\tau s + x}{2}$ is concave.

See Figure 17 and Figure 18.

Furthermore, in view of Proposition 3.1, we can rewrite our target minimization problem (16) as below:

$$\min_{w\in\mathbb{R}^m, b\in\mathbb{R}} \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{l} \frac{1+\tau}{2}|(1-y_i(w^T\Phi(x_i)+b))|\right)$$
$$+ C\left(\sum_{i=1}^{l} -\frac{\tau}{2}|(1-y_i(w^T\Phi(x_i)+b))+s|\right) + C\left(\sum_{i=1}^{l} \frac{\tau s + (1-y_i(w^T\Phi(x_i)+b))}{2}\right). \tag{41}$$

Now, we apply smoothing functions of the absolute value function presented in Section 2 to derive smoothing function for $P_{\tau,s}(x)$. To this end, we define

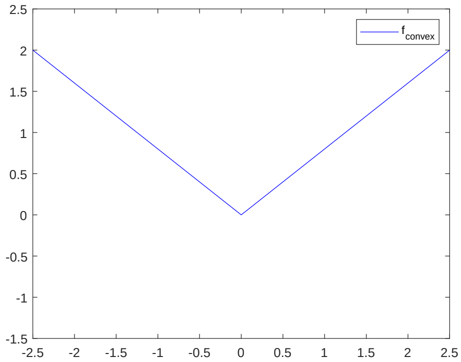
$$\varphi_{\tau,s}(x,\mu) = \left(\frac{1+\tau}{2}\varphi_{\text{abs}}(x,\mu)\right) + \left(-\frac{\tau}{2}\varphi_{abs}(x+s,\mu) + \frac{\tau s + x}{2}\right), \tag{42}$$

where $\varphi_{\text{abs}}(x,\mu)$ is any smoothing function of absolute value function $|x|$.

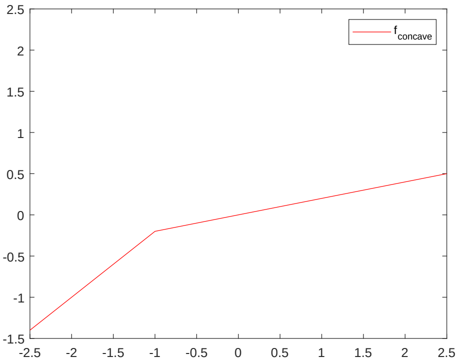
Proposition 3.2. *Let $\varphi_{\tau,s}(x,\mu)$ be defined as in (42). Then, $\varphi_{\tau,s}(x,\mu)$ is continuously differentiable for all $(x,\mu) \in \mathbb{R} \times \mathbb{R}_{++}$.*

Proof. Note that $\varphi_{\text{abs}}(x+s,\mu)$ and $\varphi_{\text{abs}}(x,\mu)$ are continuously differentiable. Besides, the function $\frac{\tau s + x}{2}$ is obviously continuously differentiable. Hence, $\varphi_{\tau,s}(x,\mu)$ is continuously differentiable. □

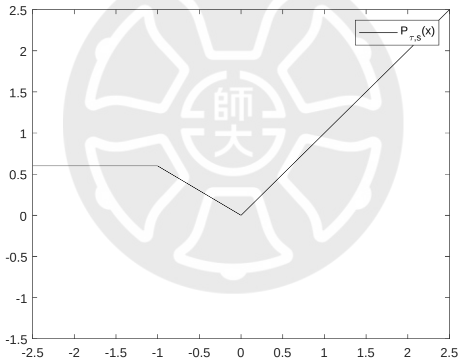
Proposition 3.3. *Let $\varphi_{\tau,s}(x,\mu)$ be defined as in (42). Then, $\lim_{\mu\to 0^+} \varphi_{\tau,s}(x,\mu) = P_{\tau,s}(x)$ for all $(x,\mu) \in \mathbb{R} \times \mathbb{R}_{++}$.*



(a) $f_{\text{convex}}(x)$: convex function



(b) $f_{\text{concave}}(x)$: concave function



(c) $P_{\tau,s}(x)$

Figure 17: $f_{\text{convex}}(x) + f_{\text{concave}}(x) = P_{\tau,s}(x)$ with $\tau = 0.6, s = 1$

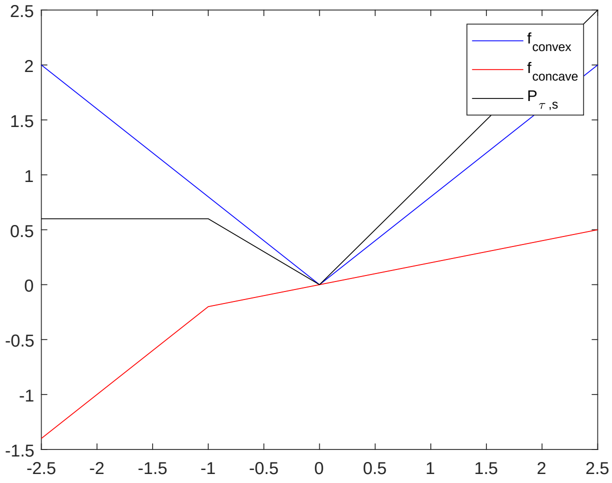


Figure 18: Graphs of $\mathrm{P}_{\tau,s}(x)$, $\mathrm{f}_{\mathrm{convex}}(x)$, and $\mathrm{f}_{\mathrm{concave}}(x)$ with $\tau = 0.6, s = 1$

Proof. By Definition 2.1, we know that

$$\lim_{\mu\to 0^+} \varphi_{\mathrm{abs}}(x,\mu) = |x|,$$

$$\lim_{\mu\to 0^+} \varphi_{\mathrm{abs}}(x+s,\mu) = |x+s|,$$

and

$$\lim_{\mu\to 0^+} \frac{\tau s + x}{2} = \frac{\tau s + x}{2}.$$

Hence, we have

$$\begin{aligned}
\lim_{\mu\to 0^+} \varphi_{\tau,s}(x,\mu) &= \lim_{\mu\to 0^+} \left[\left(\frac{1+\tau}{2}\varphi_{\mathrm{abs}}(x,\mu) \right) + \left(-\frac{\tau}{2}\varphi_{\mathrm{abs}}(x+s,\mu) + \frac{\tau s + x}{2} \right) \right] \\
&= \frac{1+\tau}{2}|x| - \frac{\tau}{2}|x+s| + \frac{\tau s + x}{2} \\
&= P_{\tau,s}(x).
\end{aligned}$$

Then, the proof is complete. $\square$

Based on Proposition 3.2 and Proposition 3.3, we can say that the function $\varphi_{\tau,s}(x,\mu)$ is a smoothing function of the truncated pinball loss function $P_{\tau,s}(x)$. Accordingly, we obtain

$$\varphi^k_{\tau,s}(x,\mu) = \left(\frac{1+\tau}{2}\varphi^k_{\mathrm{abs}}(x,\mu) \right) + \left(-\frac{\tau}{2}\varphi^k_{\mathrm{abs}}(x+s,\mu) + \frac{\tau s + x}{2} \right), \tag{43}$$

where $\varphi^k_{\mathrm{abs}}(x,\mu)$ defined as (20)-(29) for $k = 1, \cdots, 10$. For $\overline{pin}$-SVM, we replace $P_{\tau,s}(x)$ with $\varphi^k_{\tau,s}(x,\mu)$ to achieve the following problem:

$$\min_{w\in\mathbb{R}^m, b\in\mathbb{R}} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \varphi^k_{\tau,s}\left(1 - y_i(w^T\Phi(x_i) + b), \mu\right). \tag{44}$$

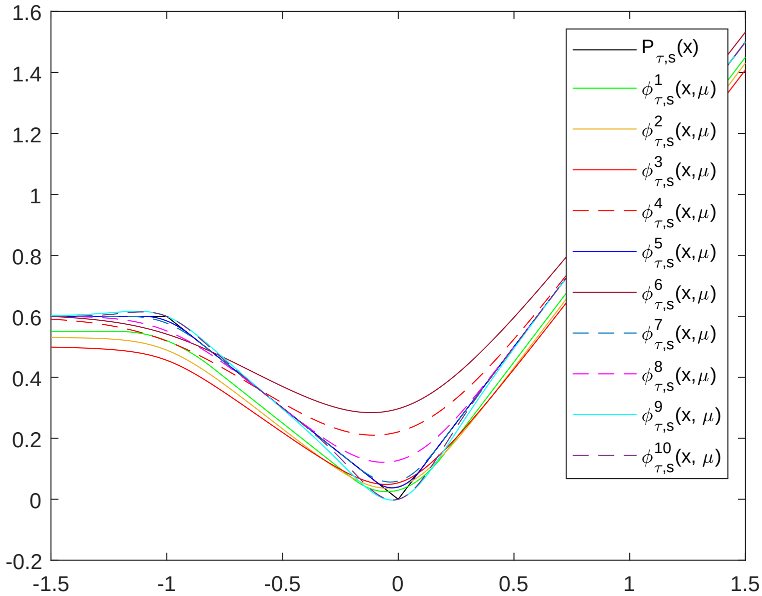


Figure 19: Graphs of $P_{\tau,s}(x)$ and $\varphi^k_{\tau,s}(x,\mu)$ with $\mu = 0.2, \tau = 0.6, s = 1$

Next proposition indicates that the difference of $\varphi^k_{\tau,s}(x,\mu)$ and $P_{\tau,s}(x)$ have a upper bound $C_k^{\tau,s}\,\mu$.

Proposition 3.4. *For $k = 1, 2, \cdots, 10$, there exists $C_k^{\tau,s} > 0$ such that*

$$\left|\varphi^k_{\tau,s}(x,\mu) - P_{\tau,s}(x)\right| \leq C_k^{\tau,s}\,\mu$$

for all $x \in \mathbb{R}$. In fact, $C_k^{\tau,s} = \frac{1+2\tau}{2}C_k$.

Proof. Applying Lemma 2.1 yields

$$\begin{aligned}
& \left|\varphi^k_{\tau,s}(x,\mu) - P_{\tau,s}(x)\right| \\
= & \left|(\frac{-\tau}{2}\varphi^k_{abs}(x+s,\mu) + \frac{1+\tau}{2}\varphi^k_{abs}(x,\mu) + \frac{\tau s + x}{2}) - (\frac{-\tau}{2}|x+s| + \frac{1+\tau}{2}|x| + \frac{\tau s + x}{2})\right| \\
= & \left|\frac{-\tau}{2}(\varphi^k_{\text{abs}}(x+s,\mu) - |x+s|) + \frac{1+\tau}{2}(\varphi^k_{\text{abs}}(x,\mu) - |x|) + (\frac{\tau s + x}{2} - \frac{\tau s + x}{2})\right| \\
\leq & \left|\frac{-\tau}{2}(\varphi^k_{\text{abs}}(x+s,\mu) - |x+s|\right| + \left|\frac{1+\tau}{2}(\varphi^k_{abs}(x,\mu) - |x|\right| + |0| \\
= & \frac{\tau}{2}\left|\varphi^k_{\text{abs}}(x+s,\mu) - |x+s|\right| + \frac{1+\tau}{2}\left|\varphi^k_{abs}(x,\mu) - |x|\right| \\
\leq & \frac{\tau}{2}C_k\,\mu + \frac{1+\tau}{2}C_k\,\mu \\
= & \frac{1+2\tau}{2}C_k\,\mu
\end{aligned}$$

Then, the proof is complete. □

We point out Proposition 3.4 is an alternative way to prove Proposition 3.3. Now, we study the derivative of $\varphi^k_{\tau,s}(x,\mu)$ with respect to x as μ approaches 0^+. We will use the following lemma to prove Proposition 3.5.

Lemma 3.1. *For $k = 1, 2, \cdots, 10$, there hold*

$$\lim_{\mu\to 0^+} \nabla_x \varphi^k_{\text{abs}}(x,\mu) = \begin{cases} 1 & \text{if} \quad x > 0, \\ 0 & \text{if} \quad x = 0, \\ -1 & \text{if} \quad x < 0. \end{cases}$$

Proof. For $k \neq 9$, the proof is easy and we omit them.

For $k = 9$, we discuss two cases. If $x = 0$, then

$$\lim_{\mu\to 0^+} \nabla_x \varphi^9_{\text{abs}}(x,\mu) = \lim_{\mu\to 0^+} \left(\frac{0}{\mu} \text{sech}^2 \left(\frac{0}{\mu} \right) + \tanh \left(\frac{0}{\mu} \right) \right) = 0 + 0 = 0.$$

If $x \neq 0$, then

$$\lim_{\mu\to 0^+} \nabla_x \varphi^9_{\text{abs}}(x,\mu) = \lim_{\mu\to 0^+} \left(\frac{x}{\mu} \text{sech}^2 \left(\frac{x}{\mu} \right) + \tanh \left(\frac{x}{\mu} \right) \right) = 0 + 1 = 1.$$

In the above, we use the fact

$$\lim_{\mu\to 0^+} \left(\frac{x}{\mu} \text{sech}^2 \left(\frac{x}{\mu} \right) \right) = 0,$$

which is due to

$$\begin{aligned} \lim_{\mu\to 0^+} \left(\frac{x}{\mu} \text{sech}^2 \left(\frac{x}{\mu} \right) \right) &= \lim_{\mu\to 0^+} \left(\frac{\text{sech}^2 \left(\frac{x}{\mu} \right)}{\frac{\mu}{x}} \right) \\ &= \lim_{\mu\to 0^+} \left(\frac{2\text{sech} \left(\frac{x}{\mu} \right) \tanh \left(\frac{x}{\mu} \right) \text{sech} \left(\frac{x}{\mu} \right) \frac{-x^2}{\mu^2}}{\frac{1}{x}} \right) \\ &= \lim_{\mu\to 0^+} \left(\frac{\text{sech}^2 \left(\frac{x}{\mu} \right)}{\mu^2} \left(-2x^2 \tanh \left(\frac{x}{\mu} \right) \right) \right) \\ &= \lim_{\mu\to 0^+} \left(\frac{\frac{2}{\mu}}{e^{\frac{x}{\mu}} + e^{\frac{-x}{\mu}}} \frac{\frac{2}{\mu}}{e^{\frac{x}{\mu}} + e^{\frac{-x}{\mu}}} \left(-2x^2 \tanh \left(\frac{x}{\mu} \right) \right) \right) \\ &= 0 \cdot 0 \cdot (-2x^2) \cdot \text{sgn}(x) = 0. \end{aligned}$$

Then, the proof is complete. □

Proposition 3.5. *For $k = 1, 2, \cdots, 10$, there holds*

$$\lim_{\mu\to 0^+} \nabla_x \varphi^k_{\tau,s}(x,\mu) = \begin{cases} 1 & \text{if } \ x > 0, \\ \dfrac{1-\tau}{2} & \text{if } \ x = 0, \\ -\tau & \text{if } \ -s < x < 0, \\ -\dfrac{\tau}{2} & \text{if } \ x = -s, \\ 0 & \text{if } \ x < -s. \end{cases} \tag{45}$$

Proof. First, we observe that

$$\begin{aligned} &\lim_{\mu\to 0^+} \nabla_x \varphi^k_{\tau,s}(x,\mu) \\ =& \lim_{\mu\to 0^+} \nabla_x \left(\frac{-\tau}{2} \varphi^k_{\text{abs}}(x+s,\mu) + \frac{1+\tau}{2} \varphi^k_{\text{abs}}(x,\mu) + \frac{\tau s + x}{2} \right) \\ =& \frac{-\tau}{2} \left(\lim_{\mu\to 0^+} \nabla_x \varphi^k_{\text{abs}}(x+s,\mu) \right) + \frac{1+\tau}{2} \left(\lim_{\mu\to 0^+} \nabla_x \varphi^k_{\text{abs}}(x,\mu) \right) + \left(\lim_{\mu\to 0^+} \nabla_x \frac{\tau s + x}{2} \right) \end{aligned}$$

Then, by Lemma 3.1, we have

$$\begin{aligned} \lim_{\mu\to 0^+} \nabla_x \varphi^k_{\text{abs}}(x+s,\mu) &= \lim_{\mu\to 0^+} \nabla_{(x+s)} \varphi^k_{\text{abs}}(x+s,\mu) \\ &= \begin{cases} 1 & \text{if } \ x+s > 0 \\ 0 & \text{if } \ x+s = 0 \\ -1 & \text{if } \ x+s < 0 \end{cases} \\ \lim_{\mu\to 0^+} \nabla_x \varphi^k_{\text{abs}}(x,\mu) &= \begin{cases} 1 & \text{if } \ x > 0 \\ 0 & \text{if } \ x = 0 \\ -1 & \text{if } \ x < 0 \end{cases} \\ \lim_{\mu\to 0^+} \nabla_x (\frac{\tau s + x}{2}) &= \frac{1}{2} \end{aligned}$$

Hence, if $x > 0$, $\lim\limits_{\mu\to 0^+} \nabla_x \varphi^k_{\tau,s}(x,\mu) = \frac{-\tau}{2} \cdot 1 + \frac{1+\tau}{2} \cdot 1 + \frac{1}{2} = \frac{1}{2} + \frac{1}{2} = 1$.
if $x = 0$, $\lim\limits_{\mu\to 0^+} \nabla_x \varphi^k_{\tau,s}(x,\mu) = -\frac{\tau}{2} \cdot 1 + \frac{1+\tau}{2} \cdot 0 + \frac{1}{2} = \frac{1-\tau}{2}$.
if $x < 0, x+s > 0$, $\lim\limits_{\mu\to 0^+} \nabla_x \varphi^k_{\tau,s}(x,\mu) = \frac{-\tau}{2} \cdot 1 + \frac{1+\tau}{2} \cdot (-1) + \frac{1}{2} = -\tau$.
if $x = -s$, $\lim\limits_{\mu\to 0^+} \nabla_x \varphi^k_{\tau,s}(x,\mu) = -\frac{\tau}{2} \cdot 0 + \frac{1+\tau}{2} \cdot (-1) + \frac{1}{2} = -\frac{\tau}{2}$.
if $x + s < 0$, $\lim\limits_{\mu\to 0^+} \nabla_x \varphi^k_{\tau,s}(x,\mu) = \frac{-\tau}{2} \cdot (-1) + \frac{1+\tau}{2} \cdot (-1) + \frac{1}{2} = -\frac{1}{2} + \frac{1}{2} = 0$.
The proof is complete. $\square$

Note that the equation (45) is similar to (17). In fact, from [14, inequality(20)], we learn that

$$\begin{aligned} \varphi^6_{\text{abs}}(x,\mu) \geq \varphi^4_{\text{abs}}(x,\mu) \geq \varphi^8_{\text{abs}}(x,\mu) &\geq \varphi^7_{\text{abs}}(x,\mu) \geq \varphi^5_{\text{abs}}(x,\mu) \geq |x| \\ &\geq \varphi^1_{\text{abs}}(x,\mu) \geq \varphi^2_{\text{abs}}(x,\mu) \geq \varphi^3_{\text{abs}}(x,\mu) \end{aligned}$$

for all $x \in \mathbb{R}$ and for fixed $\mu > 0$. Where are the positions of $\varphi^9_{\rm abs}(x,\mu)$ and $\varphi^{10}_{\rm abs}(x,\mu)$? We answer it in next proposition, which talks about the order of $\varphi^9_{\rm abs}(x,\mu)$ and $\varphi^{10}_{\rm abs}(x,\mu)$ between $|x|$ and $\varphi^1_{\rm abs}(x,\mu)$. It may help us to study the order of $P_{\tau,s}$ and $\varphi^k_{\tau,s}(x,\mu)$, see Figure 20.

Proposition 3.6. *For all $x \in \mathbb{R}$ and for fixed $\mu > 0$, there holds*

$$|x| \geq \varphi^{10}_{\rm abs}(x,\mu) \geq \varphi^9_{\rm abs}(x,\mu) \geq \varphi^1_{\rm abs}(x,\mu).$$

Proof. From [3, inequality (3.1)], we know $|x| \geq \varphi^{10}_{\rm abs}(x,\mu) \geq \varphi^9_{\rm abs}(x,\mu)$ for all $x \in \mathbb{R}$ and for fixed $\mu > 0$. Thus, it suffices to show that $\varphi^9_{\rm abs}(x,\mu) \geq \varphi^1_{\rm abs}(x,\mu)$.

(i) : Consider $x \in (-\infty, -\mu] \cup [\mu, \infty)$.
If $x \geq \mu$, then we will claim that

$$g(x) = \frac{\mu}{2}\ln\left(\cosh\left(\frac{2x}{\mu}\right)\right)$$

is between $\varphi^9_{\rm abs}(x,\mu)$ and $\varphi^1_{\rm abs}(x,\mu)$. Based on [18, Theorem 3], we obtain

$$\begin{aligned}\varphi^9_{\rm abs}(x,\mu) &= x\tanh\left(\frac{x}{\mu}\right)\\ &= x\tanh\left(\frac{\left(\frac{2x}{\mu}\right)}{2}\right)\\ &= \frac{\mu}{2}\left(\frac{2x}{\mu}\right)\tanh\left(\frac{\left(\frac{2x}{\mu}\right)}{2}\right)\\ &\geq \frac{\mu}{2}\ln\left(\cosh\left(\frac{2x}{\mu}\right)\right)\\ &= g(x)\end{aligned} \tag{46}$$

Now, we prove that $g(x) \geq \varphi^1_{\rm abs}(x,\mu)$. We claim that

$$\ln\left(\cosh\left(\frac{2x}{\mu}\right)\right) + 1 > \frac{2x}{\mu}. \tag{47}$$

Due to

$$\begin{aligned}\ln\left(\cosh\left(\frac{2x}{\mu}\right)\right) + 1 &= \ln\left(e\frac{e^{\frac{2x}{\mu}} + e^{-\frac{2x}{\mu}}}{2}\right)\\ &= \ln\left(\left(\frac{e}{2}\right)\left(e^{\frac{2x}{\mu}} + e^{-\frac{2x}{\mu}}\right)\right)\\ &> \ln\left(e^{\frac{2x}{\mu}} + e^{-\frac{2x}{\mu}}\right)\\ &> \ln\left(e^{\frac{2x}{\mu}} + 0\right)\\ &= \frac{2x}{\mu},\end{aligned}$$

(47) holds. Besides, based on (47), we can obtain the following inequality.

$$\begin{aligned}
&\ln\left(\cosh\left(\frac{2x}{\mu}\right)\right)+1>\frac{2x}{\mu}\\
\Rightarrow&\frac{\mu}{2}\ln\left(\cosh\left(\frac{2x}{\mu}\right)\right)+\frac{\mu}{2}>x\\
\Rightarrow&\frac{\mu}{2}\ln\left(\cosh\left(\frac{2x}{\mu}\right)\right)>x-\frac{\mu}{2}.
\end{aligned}$$

Therefore , if $x \geq \mu$,

$$g(x) \geq \varphi^1_{\text{abs}}(x,\mu). \tag{48}$$

Based on (46) and (48), $\varphi^9_{\text{abs}}(x,\mu) \geq \varphi^1_{\text{abs}}(x,\mu)$ holds for $x \geq \mu$. By symmetry of $\varphi^9_{\text{abs}}(x,\mu)$ and $\varphi^1_{\text{abs}}(x,\mu)$, we know that $\varphi^9_{\text{abs}}(x,\mu) \geq \varphi^1_{\text{abs}}(x,\mu)$ holds for $x \in (-\infty,-\mu] \cup [\mu,\infty)$.

(ii) : For fixed μ, we let $F(y) = \varphi^9_{\text{abs}}(y,\mu) - \varphi^1_{\text{abs}}(y,\mu)$. We will claim that $F(y) \geq 0$ for all $y \in (-\mu,\mu)$. If $0 \leq y < \mu$, then $F(y) = y\tanh\left(\frac{y}{\mu}\right) - \frac{y^2}{2\mu}$. It is enough to show that F is increasing function, so we compute the derivative F.

$$\begin{aligned}
F'(y) &= \tanh\left(\frac{y}{\mu}\right) + y\operatorname{sech}^2\left(\frac{y}{\mu}\right)\frac{1}{\mu} - \frac{y}{\mu}\\
&= \tanh\left(\frac{y}{\mu}\right) + \frac{y}{\mu}\left(\operatorname{sech}^2\left(\frac{y}{\mu}\right) - 1\right)\\
&= \tanh\left(\frac{y}{\mu}\right) + \frac{y}{\mu}\left(-\tanh^2\left(\frac{y}{\mu}\right)\right)\\
&= \tanh\left(\frac{y}{\mu}\right)\left[1 - \frac{y}{\mu}\tanh\left(\frac{y}{\mu}\right)\right].
\end{aligned}$$

Based on $0 < \frac{y}{\mu} < 1$ and $0 < \tanh\left(\frac{y}{\mu}\right) < 1$, we know $1 - \frac{y}{\mu}\tanh(\frac{y}{\mu}) > 0$. Then, we have

$$F'(y) = \tanh\left(\frac{y}{\mu}\right)\left[1 - \frac{y}{\mu}\tanh\left(\frac{y}{\mu}\right)\right] > 0.$$

Because F is a continuous function and

$$F(0) = 0 \cdot \tanh\left(\frac{0}{\mu}\right) - \frac{0^2}{2\mu} = 0 - 0 = 0,$$

we prove that $F(y) \geq 0, \forall y \in [0,\mu)$. Hence, $y\tanh\left(\frac{y}{\mu}\right) > \frac{y^2}{2\mu}$, $\forall y \in [0,\mu)$. By symmetry of $F(y)$, it is true that $F(y) \geq 0$ for all $y \in (-\mu,\mu)$.

Thus, the proof is complete. □

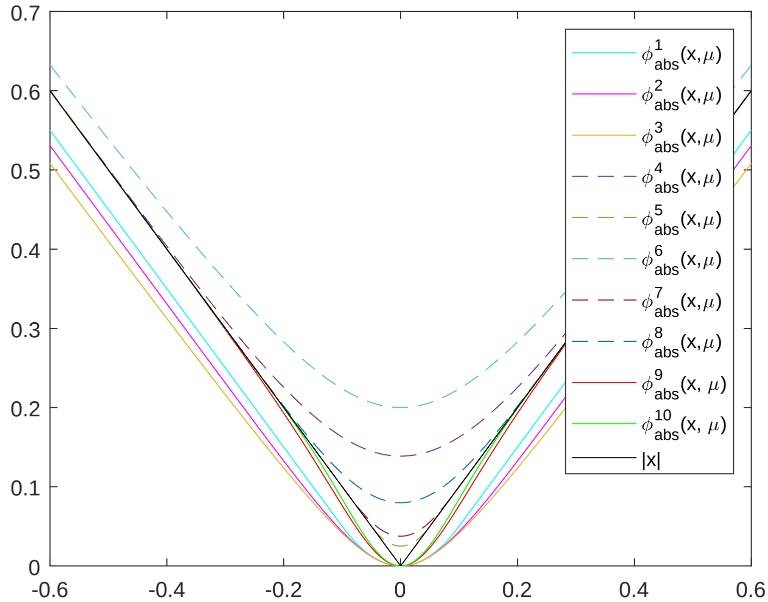


Figure 20: Graphs of $|x|$ and all ten $\varphi^k_{\rm abs}(x, \mu)$ with $\mu = 0.1$

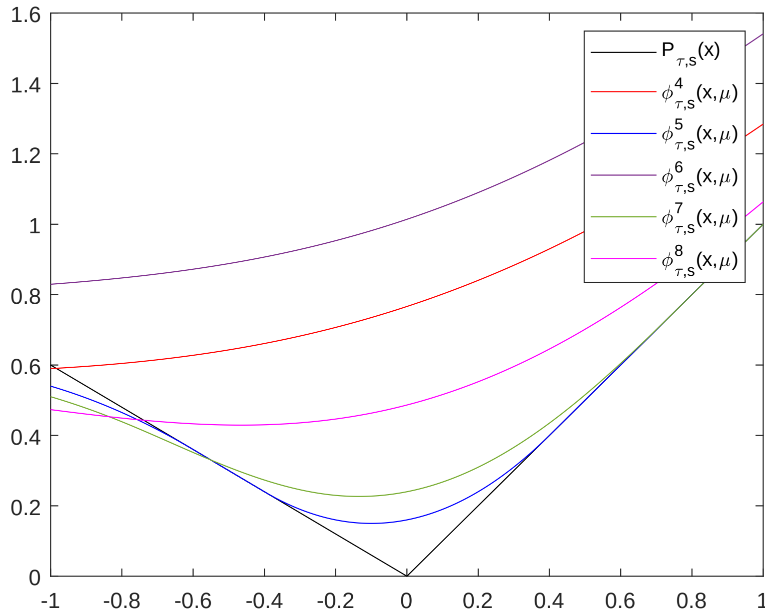
From Figure 21 regarding the smoothing functions $\varphi^k_{\tau,s}(x, \mu)$ for $k = 4, 5, 6, 7, 8$ and the truncated pinball loss function $P_{\tau,s}(x)$, we have the observation: there exists an interval $[a_1, b_1]$ with $a_1 < 0 < b_1$,

$$\varphi^6_{\tau,s}(x, \mu) \geq \varphi^4_{\tau,s}(x, \mu) \geq \varphi^8_{\tau,s}(x, \mu) \geq \varphi^7_{\tau,s}(x, \mu) \geq \varphi^5_{\tau,s}(x, \mu) \geq P_{\tau,s}(x).$$

We are not able to provide the algebraic verifications yet, which is our future direction. From Figure 22 regarding the smoothing functions $\varphi^k_{\tau,s}(x, \mu)$ for $k = 1, 2, 3, 9$ and the truncated pinball loss function $P_{\tau,s}(x)$, we observe: there exists an interval $[a_2, b_2]$ with $a_2 < 0 < b_2$,

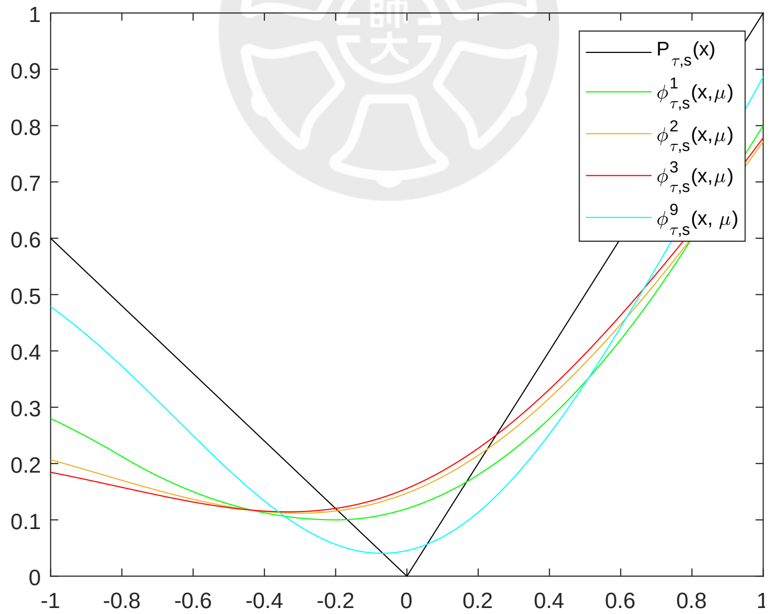
$$\varphi^3_{\tau,s}(x, \mu) \geq \varphi^2_{\tau,s}(x, \mu) \geq \varphi^1_{\tau,s}(x, \mu) \geq \varphi^9_{\tau,s}(x, \mu) \geq P_{\tau,s}(x).$$

We are not able to provide the algebraic verifications yet, which is our future direction.



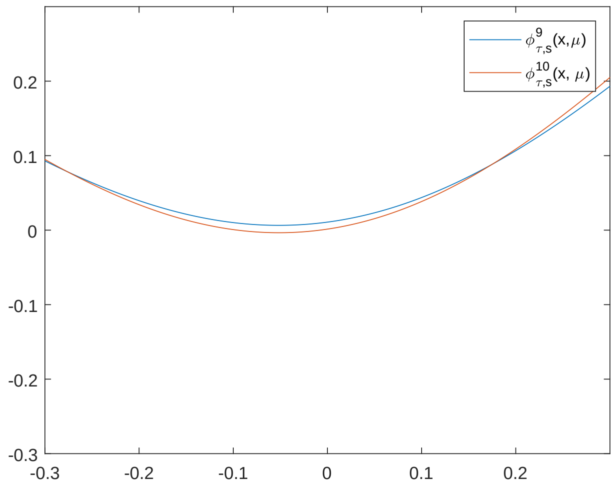
(a) $\varphi^k_{\tau,s}(x,\mu)$ for $k = 4, 5, 6, 7, 8$ for $\mu = 0.8$

Figure 21: Conjecture one

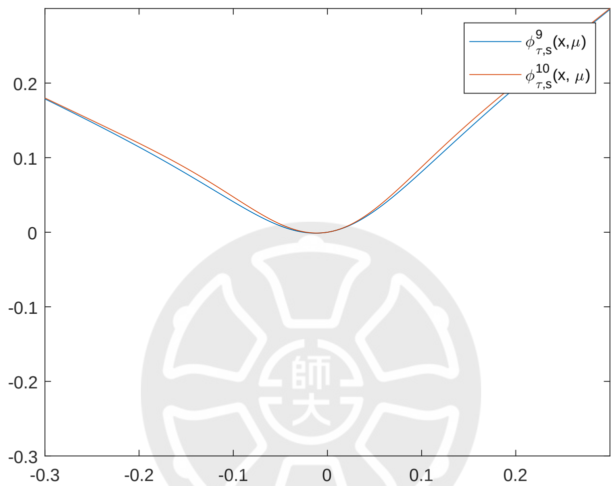


(a) $\varphi^k_{\tau,s}(x,\mu)$ for $k = 1, 2, 3, 9$ for $\mu = 0.8$

Figure 22: Conjecture two



(a) $\varphi^k_{\tau,s}(x,\mu)$ for $k = 9, 10$ for $\mu = 0.5$



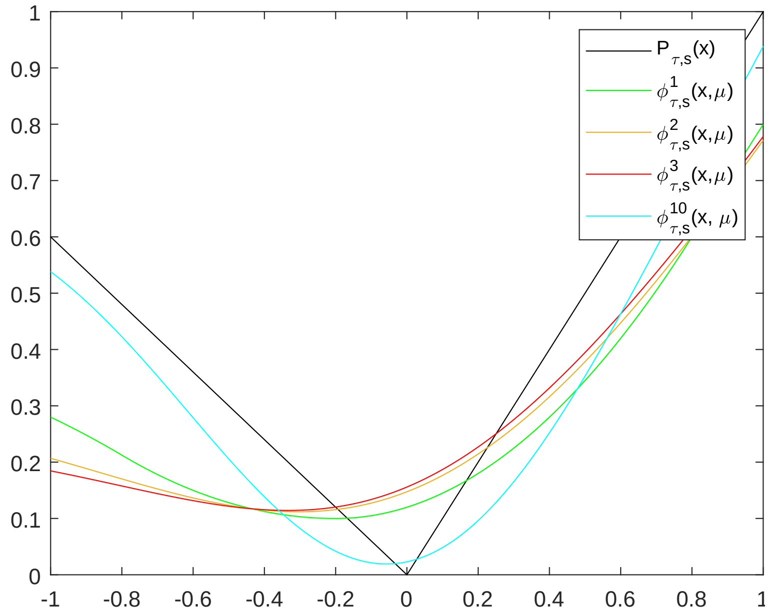
(b) $\varphi^k_{\tau,s}(x,\mu)$ for $k = 9, 10$ for $\mu = 0.1$

Figure 23: Ordering $\varphi^k_{\tau,s}(x,\mu)$ for $k = 9, 10$ with different μ

As seen in Figure 23, the graphs of $\varphi^9_{\tau,s}(x,\mu)$ and $\varphi^{10}_{\tau,s}(x,\mu)$ are very close, ordering $\varphi^9_{\tau,s}(x,\mu)$ and $\varphi^{10}_{\tau,s}(x,\mu)$ near $x = 0$ is difficult. In Figure 24, we will order $\varphi^k_{\tau,s}(x,\mu)$ for $k = 1, 2, 3, 10$ and $P_{\tau,s}(x)$. From this graph of smoothing functions $\varphi^k_{\tau,s}(x,\mu)$ for $k = 1, 2, 3, 10$ and the truncated pinball loss function $P_{\tau,s}(x)$ in Figure 24, we observe: there exists an interval $[a_3, b_3]$ with $a_3 < 0 < b_3$,

$$\varphi^3_{\tau,s}(x,\mu) \geq \varphi^2_{\tau,s}(x,\mu) \geq \varphi^1_{\tau,s}(x,\mu) \geq \varphi^{10}_{\tau,s}(x,\mu) \geq P_{\tau,s}(x).$$

Again, it still needs algebraic verifications.



(a) $\varphi^k_{\tau,s}(x,\mu)$ for $k = 1, 2, 3, 10$ for $\mu = 0.8$

Figure 24: Conjecture three

4 Conclusion

In this thesis, we present 10 smoothing functions $\varphi^k_{\tau,s}(x,\mu)$ of the truncated pinball loss function $P_{\tau,s}(x)$, which are constructed by the smoothing functions of absolute value function $|x|$. The first eight smoothing functions $\varphi^k_{\tau,s}(x,\mu)$ can be expressed as the sum of a convex function and a concave function. However, $\varphi^9_{\tau,s}(x,\mu)$ and $\varphi^{10}_{\tau,s}(x,\mu)$ may not easily or cannot be expressed as the sum such as the above. Geometric views are provided to verify their features. Possible direction is applying these smoothing functions to work along some algorithms and compare their numerical performance.

References

[1] Y.J. Bagul, *A smooth transcendental approximation to $|x|$*, International Journal of Mathematical Sciences and Engineering Applications(IJMSEA), vol. 11, pp. 213–217, 2017.

[2] Y.J. Bagul and B.K. Khairnar, *A note on smooth transcendental approximation to $|x|$*, `https://www.preprints.org/manuscript/201902.0190/download/final_file`, 2019.

[3] Y.J. Bagul and C. Chesneau, *Sigmoid functions for the smooth approximation to the absolute value function*, Moroccan Journal of Pure and Applied Analysis, vol. 7, pp. 12–19, 2021.

[4] A. Beck and M. Teboulle, *Smoothing and first order methods: A unified framework*, SIAM Journal on Optimization, vol. 22, pp. 557-580, 2012.

[5] C.-H. Chen and O.L. Mangasarian, *A class of smoothing functions for nonlinear and mixed complementarity problems*, Computational Optimization and Applications, vol. 5, pp. 97–138, 1996.

[6] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag New York, vol. 2, 2003.

[7] W.-Z. Gu, W.-P. Chen, C.-H. Ko, Y.-J. Lee, and J.-S. Chen, *Two smooth support vector machines for ε-insensitive regression*, Computational Optimization and Applications, vol. 70, pp. 171–199, 2018.

[8] J. Han and C. Moraga, *The influence of the sigmoid function parameters on the speed of backpropagation learning*, International Workshop on Artificial Neural Networks, pp. 195–201, 1995.

[9] X.-L. Huang, L. Shi, and J. A.K. Suykens, *Support vector machine classifier with pinball loss*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, pp. 984–997, 2013.

[10] Y. LIN, *A note on margin-based loss functions in classification*, Statistics & Probability Letters, vol. 68, pp. 73-82, 2004.

[11] C.T. NGUYEN, B. SAHEYA, Y.-L. CHANG, AND J.-S. CHEN, *Unified smoothing functions for absolute value equation associated with second-order cone*, Applied Numerical Mathematics, vol. 135, pp. 206–227, 2019.

[12] R.T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, 1970.

[13] B. SAHEYA, C.-H. YU, AND J.-S. CHEN, *Numerical comparisons based on four smoothing functions for absolute value equation*, Journal of Applied Mathematics and Computing, vol. 56, pp. 131–149, 2018.

[14] B. SAHEYA, C.T. NGUYEN, J.-S. CHEN, *Neural network based on systematically generated smoothing functions for absolute value equation*, Journal of Applied Mathematics and Computing, vol. 61, pp. 533–558, 2019.

[15] X. SHEN, L.-F. NIU, Z.-Q. QI, AND Y.-J. TIAN, *Support vector machine classifier with truncated pinball loss*, Pattern Recognition, vol. 68, pp.199-210, 2017.

[16] I. STEINWART, *Sparseness of Support Vector Machines—some asymptotically sharp bounds*, Advances in Neural Information Processing Systems, vol. 16, pp. 1069–1076, 2004.

[17] I. STEINWART, AND A. CHRISTMANN, *Support Vector Machines*, Springer, 1st edition, 2008.

[18] K.B. STOLARSKY, *Hölder Means, Lehmer Means, and x^{-1} log cosh x*, Journal of mathematical analysis and applications, vol. 202, pp. 810-818, 1996.

[19] A. THARWAT, *Behavioral analysis of support vector machine classifier with Gaussian kernel and imbalanced data*, arXiv preprint arXiv:2007.05042, 2020.

[20] H. TUY, *Convex Analysis and Global Optimization*, Springer International Publishing, 2nd, 2016.

[21] L. VANDENBERGHE, *Smoothing*, `seas.ucla.edu/~vandenbe/236C/lectures/smoothing.pdf`, Spring 2013-14.

[22] V. VAPNIK, *The Nature of Statical Learning Theory*, Springer-Verlag New York, 2nd, 2000.

[23] S. VORONIN, G. OZKAYA, AND D. YOSHIDA, *Convolution based smooth approximations to the absolute value function with application to non-smooth regularization*, arXiv preprint arXiv:1408.6795 [math.NA], 2014.

[24] X.-G. Zhang, *Using class-center vectors to build support vector machines*, Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No. 98TH8468), pp. 3–11, 1999.