A new regularization for sparse optimization

Josef Dick[1] Guoyin Li[2] Dinh Duy Tran[3]

(Received 20 December 2020; revised 6 September 2021)

Abstract

Several numerical studies have shown that non-convex sparsity-induced regularization can outperform the convex ℓ_1-penalty. In this article, we introduce a new non-convex and non-smooth regularization. This new regularization is a continuous and separable function which provides a tighter approximation to the cardinality function than any ℓ_q-penalty ($0 < q < 1$). We then apply the Proximal Gradient Method to solve a regularized optimization problem with the new regularization. The convergence analysis shows that the algorithm converges to a critical point and we also provide a pseudo-code for fast implementation. In addition, we conduct a simple numerical experiment with a regularized least square problem to illustrate the performance of the new regularization.

DOI:10.21914/anziamj.v62.16076, Published 2022-02-07, as part of the Proceedings of the 19th Biennial Computational Techniques and Applications Conference. ISSN 1445-8810. (Print two pages per sheet of paper.)

Contents

1 Introduction

In recent years, sparse models have become popular in many applications such as signal processing, statistics, and machine learning. To obtain sparsity, a suitable regularization is applied to the optimization problem. The most familiar sparsity-inducing regularization is the cardinality function and its convex-relaxation ℓ_1-penalty [8]. Optimization with the ℓ_0-penalty can be unstable due to its intrinsic discontinuity. Meanwhile, the ℓ_1-penalty usually produces biased models because it tends to shrink large-valued parameters excessively [7]. Furthermore, it only achieves reliable sparse recovery under a strong condition, namely a low coherent sensing matrix [5].

Lately, there are lines of research focusing on non-convex, continuous and symmetric regularizations that are concave on $[0, +\infty]$. Some notable examples are smoothly clipped absolute deviation [7], the minimax concave penalty [13], the ℓ_q-penalty with $0 < q < 1$ [4], and the transformed ℓ_1-penalty [9]. Recent developments in the theory of non-convex and non-smooth optimization

have encouraged practical implementation of non-convex regularization. Recently, Wen et al. [12] provided a list of applications to which non-convex regularization had been applied, and they showed non-convex regularizations produce unbiased models which usually achieve better sparse recovery under relaxed conditions. As illustrated in Figure 1, a potential reason is that the curves of these non-convex regularizations are bending towards the origin and the curves become flatter away from the origin. In other words, they are continuous surrogates which provide a better approximation to the cardinality function than the ℓ_1-penalty. Furthermore, through the use of a relative convergence ratio introduced in Section 2, one can also find that around the origin the ℓ_0-penalty has the slowest decay to zero, followed by the ℓ_q-penalty and the ℓ_1-penalty. By using these observations as motivation, we hypothesize that a continuous, non-convex regularization which decreases to zero slower when approaching the origin can better model the jump of the cardinality function. To the best of our knowledge, in the current literature, the ℓ_q-penalty ($0 < q < 1$) has the slowest decaying speed among the existing non-convex regularizations with good numerical performances. In Section 2, we therefore introduce a new non-convex and non-smooth regularization which converges to zero slower than any ℓ_q-penalty when approaching the origin.

In Section 3, we use ideas from Marjanovic and Solo [10] to set up an algorithm which solves regularized optimization problems with the new regularization. We also provide a convergence analysis and pseudo-code for the algorithm. Section 4 covers simulations for regularized least square problems with synthetic data. Finally, Section 5 provides a conclusion and suggestions for future work.

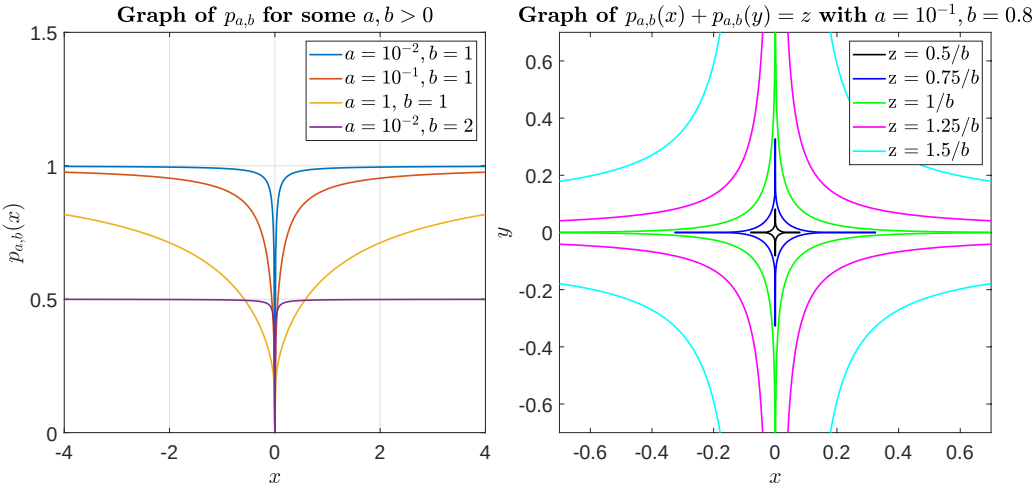

Figure 1: The graph of penalty $p_{a,b}$ and its level sets.

2 The new regularization

Let $a, b \in \mathbb{R}_{++}$, where $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$. We define the new regularization $p_{a,b} : \mathbb{R} \to \mathbb{R}$ by

$$p_{a,b}(x) = \begin{cases} \frac{1}{\ln(a|x|^{-1}+1)+b}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

As shown in Figure 1, the regularization $p_{a,b}$ is non-smooth and non-convex on $\mathbb{R}$. By examining the limit of $p_{a,b}(x)$ when $|x|$ approaches zero, one can verify that the new regularization $p_{a,b}$ is continuous. Additionally, if the scaling parameter a is small, then the gradient of $p_{a,b}(x)$ will quickly approach zero as $|x|$ increases. According to Fan and Li [7], the latter property indicates that this new regularization will result in a nearly unbiased model.

The new regularization has two notable characteristics. Firstly, $p_{a,b}$ is symmetric around zero and strictly increasing on $\mathbb{R}_{++}$. By examining the limit when x approaches infinity, we find that $p_{a,b}$ is bounded between zero and $\frac{1}{b}$.

Secondly, $p_{a,b}$ decreases to zero slower than the ℓ_q-penalty for any $0 < q < 1$. To verify this property, we define the following ratio which compares the convergence rates as $|x|$ approaches zero:

$$\text{Relative Convergence Ratio} = \lim_{|x|\to 0} \frac{f(x)}{g(x)},$$

where f and g are continuous, non-convex regularization functions with $f(0) = g(0) = 0$. If the ratio is positive infinity, then f converges to zero slower than g around the origin. Next, using this ratio, we deduce that the new regularization converges to zero slower than the ℓ_q-penalty because

$$\begin{aligned}
\lim_{|x|\to 0} \frac{p_{a,b}(x)}{|x|^q} &= \lim_{|x|\to 0} \frac{|x|^{-q}}{\ln\left(a|x|^{-1} + 1\right) + b} \\
&= \lim_{t\to 0^+} \frac{t^{-q}}{\ln\left(at^{-1} + 1\right) + b} \\
&= \lim_{t\to 0^+} \frac{-qt^{-q-1}}{-\left(t + \frac{t^2}{a}\right)^{-1}} \\
&= \lim_{t\to 0^+} q\left(t^{-q} + \frac{t^{1-q}}{a}\right) \\
&= +\infty,
\end{aligned}$$

where the third equality follows from the L'Hôpital's rule. As a result, we expect that this new regularization function gives a better approximation to the cardinality function than the ℓ_q penalty in the literature. We set the parameter $b = 1$ in the regularization $p_{a,b}$ to simplify the parameter tuning process for the later experiments, and so that the maximum value of $p_{a,b}$ matches with that of the cardinality function. To simplify the notation, we write p_a as the new regularization when $b = 1$.

3 The problem and algorithm

In this article, we consider

$$\min_{x\in\mathbb{R}^n} \; f(x) + \lambda \sum_{i=1}^{n} p_a(x_i)\,, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and has a Lipschitz continuous gradient. Here x_i are the ith coordinate of $x \in \mathbb{R}^n$ for $i = 1, \ldots, n$, and λ is a positive real number. The formulation (1) is a regularized optimization problem and the property of the cost function f covers a wide range of applications [12]. To solve the non-smooth non-convex problem (1), we apply the Proximal Gradient Method (PGM) [2, Chapter 10], which is a type of iterative majorization-minimization algorithm.

The majorization step involves finding the upper bound of the objective function (1). To do this step, we require the following lemma.

Lemma 1. *[2, Lemma 5.7] Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ have a Lipschitz-continuous gradient with Lipschitz constant $L \geqslant 0$ over a given convex set $D \subseteq \mathbb{R}^n$. Then*

$$f(x) \leqslant f(y) + \nabla f(y)^T(x-y) + \frac{L}{2}\|x-y\|^2\,, \quad \textit{for all} \quad x, y \in D\,.$$

Let $x^{(k)} \in \mathbb{R}^n$ be the iterate. Then, we implement Lemma 1 to the cost function f in problem (1) to obtain the majorized problem:

$$\min_{x\in\mathbb{R}^n} \; f(x^{(k)}) + \nabla f(x^{(k)})^T(x-x^{(k)}) + \frac{L}{2}\|x-x^{(k)}\|^2 + \lambda \sum_{i=1}^{n} p_a(x_i)\,.$$

In general, it is impractical or computationally expensive to determine the Lipschitz constant L precisely. Hence, one often uses a constant L_k with $L_k > L$ at each iteration step and this can be achieved by using line search

techniques. In combination with this information, the majorized problem is equivalent to

$$\min_{x\in\mathbb{R}^n} \frac{L_k}{2}\|x - z^{(k)}\|_2^2 + \lambda \sum_{i=1}^{n} p_a(x_i)\,, \tag{2}$$

where $z^{(k)}$ is defined by

$$z^{(k)} = x^{(k)} - \frac{1}{L_k}\nabla f\left(x^{(k)}\right).$$

The minimization step requires solving problem (2), whose objective function is separable. Thus, it is sufficient to solve n many one-dimensional problems

$$\min_{x\in\mathbb{R}} \frac{L_k}{2}(x - z_i^{(k)})^2 + \lambda p_a(x)\,, \tag{3}$$

where $z_i^{(k)}$ is the ith coordinate of $z^{(k)}$. It can be directly verified that the objective function in problem (3) is coercive so it has at least one global minimum. Furthermore, the candidates for the global minimum are the non-smooth point zero and the stationary points found from

$$z_i^{(k)} = \underbrace{x + \frac{\lambda}{L_k}\left(\frac{a}{a|x| + x^2}\right) p_a^2(x)\ \mathrm{sign}(x)}_{g(x)}\,, \quad x \neq 0\,, \tag{4}$$

which is derived from the first order necessary condition, and we refer to $z_i^{(k)}$ as z to simplify the notation. Then, we check the behaviour of $g(x) - z$, which is the gradient of the objective function (3) to deduce which candidate is the global minimum. Further analysis shows that the function $g : \mathbb{R}\setminus\{0\} \mapsto \mathbb{R}$ from (4) is an odd function and $g(x)$ has the same sign as its variable x. Moreover, g is smooth, strictly convex and has a strictly positive global minimum over $\mathbb{R}_{++}$. Using these properties, one has the following lemma regarding the solutions of problem (3).

Lemma 2 (Solution of (3)). *Define function* g *as in* (4) *and denote*

$$T^* = \min_{x \in \mathbb{R}_{++}} g(x) .$$

Then, there are two cases for the solution of problem (3).

- *If* $|z| \leqslant T^*$, *then zero is the global minimizer of problem* (3).
- *If* $|z| > T^*$, *then equation* (4) *has two distinct real roots, which are not zero and have the same signs. The root with larger absolute value is a local minimizer of problem* (3) *while the other root is a local maximizer. In addition, the non-smooth point zero is the second local minimizer. Thus, the global minimizer of problem* (3) *is one of the two local minimizers that has the lowest objective function value.*

There are some important points to consider before establishing an algorithm for problem (3). Firstly, the function g from (4) is an odd function. Hence, instead of solving (4) directly, we solve $|z| = g(x)$ and then set the sign of roots to be the sign of z because $g(x)$ and x have the same sign. Secondly, the second point in Lemma 2 indicates it is possible that problem (3) has two global minimizers with identical function values, and we choose zero in this case to promote sparsity. Finally, if $|z| > T^*$, then equation $|z| = g(x)$ has two unique positive roots and we need to estimate the root with larger absolute value. Since g is smooth over $\mathbb{R}_{++}$, we use the bisection method to ensure convergence. Let x^* be the minimizer of $g(x)$ when $x \in \mathbb{R}_{++}$. Using the intermediate value theorem, it follows that x^* is inbetween the two roots. Thus, the first initial point for the bisection method is x^* and $g(x^*) - |z|$ is negative. Hence, the second point should be greater than x^* and $g - |z|$ has to be positive at that point.

Algorithm 1 is the Proximal Gradient Algorithm, and Algorithm 2 is the solution method of problem (3) In Algorithm 2, the first two inputs of Bisection() are two points which bracket the root and the last input is the function to which we apply bisection method. Additionally, in our computation, we choose L_k by using line search.

Algorithm 1: Proximal Gradient Algorithm.

Set λ, a and $obj(\boldsymbol{x}) = f(\boldsymbol{x}) + \lambda \sum_{i=1}^{n} p_a(x_i)$; $k = 0$;
while *Stopping condition is not satisfied* **do**
  Set $L_k > L$;
  $\boldsymbol{z}^{(k)} = \boldsymbol{x}^{(k)} - \frac{1}{L_k} \nabla f(\boldsymbol{x}^{(k)})$;
  $\boldsymbol{x}^{(k+1)} = \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmin}} \ \frac{L_k}{2} ||\boldsymbol{x} - \boldsymbol{z}^{(k)}||_2^2 + \lambda \sum_{i=1}^{n} p_a(x_i)$;
  $k = k + 1$;
end

Algorithm 2: Solution of problem (3).

$f(x) = \frac{1}{2}(x - z)^2 + \lambda p_a(x)$;
$g(x) = f'(x) + z$;
$T^* = \min_{x>0} g(x)$; $x^* = \underset{x>0}{\operatorname{argmin}} \ g(x)$;
if $|z| > T^*$ **then**
  Choose $x^{(1)} > x^*$ such that $g(x^{(1)}) - |z| > 0$;
  $y = \text{sign}(z) \times \text{Bisection}(x^*, x^{(1)}, g(x) - |z|)$;
  if $f(0) \leqslant f(y)$ **then**
    $y = 0$;
  end
else
  $y = 0$;
end

3.1 Convergence analysis

In this subsection, we outline how the convergence analysis of the proposed algorithm is deduced from the existing literature. Firstly, Attouch, Bolte, and Svaiter [1] studied the convergence of several gradient-descent methods, including the PGM in the non-convex and non-smooth setting. In particular, they introduced the Kurdyka–Łojasiewicz (KL) inequality for the non-smooth functions and demonstrated that there are many classes of functions that satisfy such property. Secondly, Attouch, Bolte, and Svaiter [1, Theorem 2.9] proved that the sequence generated by a numerical method will converge to a stationary point of the problem given that: the objective function satisfies the KL property and the sequence satisfies sufficient decrease, relative error and continuity conditions [1, Page 8 (H1, H2, H3)]. To apply this theorem, we note that Attouch, Bolte, and Svaiter [1, Section 5] already proved that PGM produces a sequence that satisfies the sufficient decreasing and relative error conditions, and the continuity condition is guaranteed because the objective function (1) is continuous. Thus, we only need to verify that the objective function (1) satisfies the KL property. Bolte et al. [3, Section 4] deduced that the KL property is satisfied for non-smooth functions which are definable in o-minimal structure. In addition, Dries and Speissegger [6] provided several examples of o-minimal structure and one can directly verify that the objective function (1) is definable in such structure.

Therefore, applying PGM to problem (1) will generate a sequence $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ of iterates that converge $\boldsymbol{x}^{(k)} \to \overline{\boldsymbol{x}}$ as $k \to \infty$ where $\overline{\boldsymbol{x}}$ is a stationary point of the problem in the sense of Bolte et al. [3, Definition 2], and

$$\sum_{k=0}^{\infty} \|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\| < \infty\,.$$

4 Experiment

4.1 Problem statement and settings

As an illustration for the proposed regularization, we consider the penalized least square problem

$$\min_{x\in\mathbb{R}^{100}} \frac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^{n} p_a(x_i)\,, \tag{5}$$

where A is a 75-by-100 standard Gaussian matrix and p_a is the new penalty. To generate a sample vector $\mathbf{y}$, we create a true sparse model $\mathbf{x}^*$ and set $\mathbf{y} = A\mathbf{x}^* + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon}$ is a vector of normally distributed error with mean zero and variance 0.1^2. Problem (5) is a non-convex problem and PGM globally converges to a stationary point. Thus, the solution may depend on the location of the initial point of the iteration. By following suggestion from Mazumder, Friedman, and Hastie [11], we firstly solve

$$\min_{x\in\mathbb{R}^{100}} \tfrac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda_{\ell_1}\|\mathbf{x}\|_1\,,$$

which is a ℓ_1-regularized least square. Then, we use this ℓ_1 solution as the initial iterate for problem (5). Another important setting is the tuning parameter λ which controls the effect of the regularization. In this experiment, we set $\lambda = \lambda_{\ell_1}$, which is chosen by cross-validation with the one-standard-error rule [8, Section 2.3] to encourage sparsity. There are three criteria to assess the performance of our regularization: support recovery, ℓ_2-norm error and computational time. Support recovery is the percentage of correct signs that the solution has. Finally, we do 100 simulations and plot the median of those criteria.

4.2 Numerical results

This subsection consists of two parts and each part consists of two cases where we vary s, which is the sparsity of the true model $\mathbf{x}^*$.

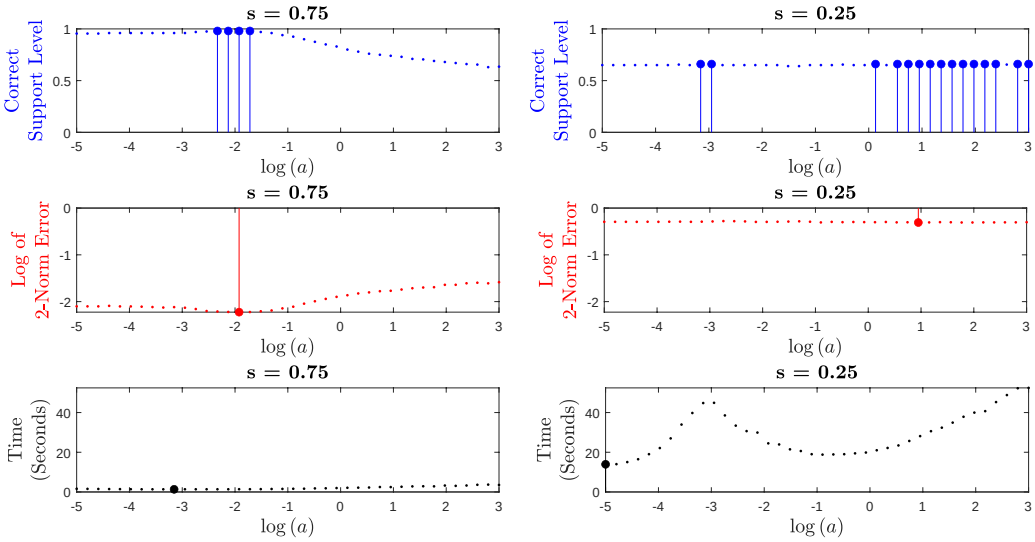

Figure 2: *Performance of regularization* $\mathfrak{p}_\mathfrak{a}$ *with different values of* $\mathfrak{a}$. *Each color represents one of the testing criteria: support recovery,* ℓ_2*-norm error or computational time, and big dots with stem highlight the best results. The value of* $\mathfrak{s}$ *is the sparsity of the true model.*

4.2.1 Performance of regularization $\mathfrak{p}_\mathfrak{a}$

The left three graphs in Figure 2 present the case $\mathfrak{s} = 0.75$ and they show that $\mathfrak{p}_\mathfrak{a}$ performs better when the parameter $\mathfrak{a}$ is roughly smaller than 10^{-2} and that is when the shape of $\mathfrak{p}_\mathfrak{a}$ is more similar to the cardinality function. When $\mathfrak{s} = 0.75$ the best results for support recovery are nearly one, which means that the solution almost has the same signs as the true model.

The overall results across all three criteria become worse when the sparsity of the true model is reduced to $\mathfrak{s} = 0.25$. Furthermore, there is no clear distinction in support recovery and ℓ_2-norm error across different values of the parameter $\mathfrak{a}$. However, the computational time has an interesting pattern when $\mathfrak{s} = 0.25$ and we suspect that it is due to instability of the algorithm.

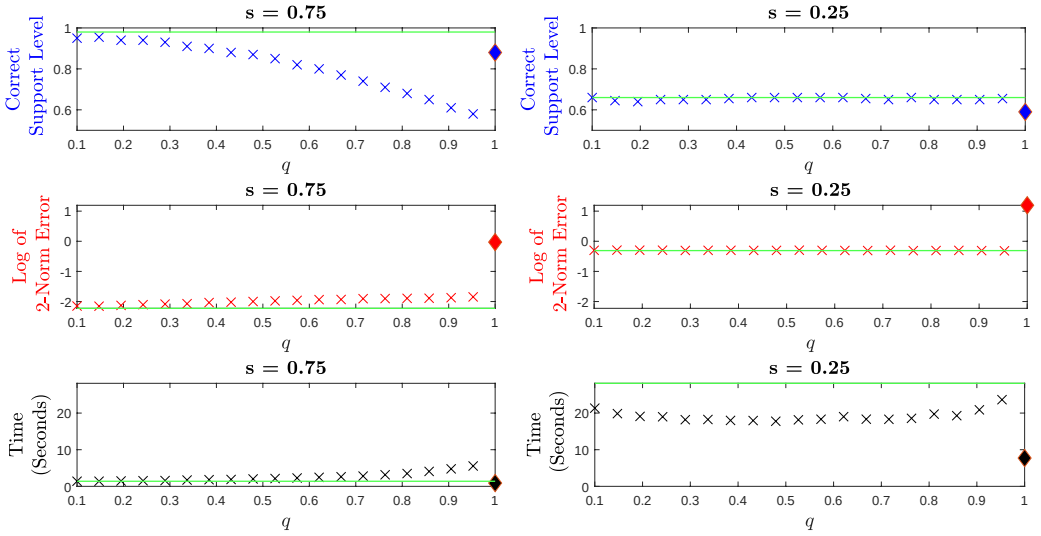

Figure 3: *Performance comparison between regularization* p_a *and* ℓ_q *penalty. The crosses and diamond denote the results for the* ℓ_q*-penalty when* $0 < q < 1$ *and* $q = 1$*, respectively. Each colour red, blue and black represents a testing criteria. The green line denotes performance of* p_a *for the* a *that give smallest 2-norm error in Figure 2. The value of* s *is the sparsity of the true model.*

In the next part, we compare the performance of the regularization p_a with the ℓ_q-penalty for $0 < q \leqslant 1$. In addition, for comparison we choose the parameter value a that has the best performance. In this experiment, we prefer accuracy over computational efficiency so we choose the value a which gives the smallest ℓ_2-norm error and highest support recovery level.

4.2.2 Comparison between p_a and ℓ_q-penalty for $0 < q \leqslant 1$

In comparison to the outcome for the ℓ_1-penalty, the optimal results of p_a are better in terms of support recovery level and ℓ_2-norm error criteria. When $s = 0.25$, p_a requires longer computational times because the one dimensional sub-problem (3) is non-convex and was solved numerically. However, when

$s = 0.75$, the computational times of the new penalty p_a and ℓ_1-penalty are not significantly different. We believe it is because the value of s is high and the ℓ_1-result is a good initial iteration for the algorithm.

In comparison to the optimal results for the ℓ_q-penalty with $0 < q < 1$, p_a performs similarly in terms of the support recovery level and ℓ_2-norm error criteria. Such similarities also appear in the computational time criteria when $s = 0.75$. When we compare the computational time when $s = 0.25$, the new regularization has a much longer computational time for some values of a.

5 Conclusion and future works

In summary, we proposed a new regularization and established some of its properties. We also provided pseudo-code to solve the regularized optimization problem (1) and provided a brief convergence analysis for the algorithm. Finally, we performed numerical experiment for regularized least square problems with synthetic data, and compared the performance of p_a-regularization, ℓ_1-penalty and ℓ_q-penalty for $0 < q < 1$. As demonstrated by the numerical experiment, in terms of accuracy and support recovery, the optimization model with new regularization performs better than the one with ℓ_1-penalty. On the other hand, there is no significant difference in comparison to the case of ℓ_q-penalty.

There are several future directions for this research. Firstly, more high-dimensional numerical experiments should be conducted. In addition, the provided pseudo-code is suitable for many cost functions so we can test regularization p_a in some other applications. Secondly, Algorithm 2 still needs some improvements which may speed up the computational time. A stability analysis is also desirable. Finally, we can study constrained optimization models using the new regularization $p_{a,b}$. In regularized optimization problems, the boundary parameter b is not considered here because it is influenced by the tuning parameter λ. However, we speculate that the boundary parameter b may have a more prominent role in constrained optimization problems.

References

[1] H. Attouch, J. Bolte, and B. F. Svaiter. "Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods". In: *Math. Program.* 137 (2013), pp. 91–129. DOI: 10.1007/s10107-011-0484-9 (cit. on p. C185).

[2] A. Beck. "First-order methods in optimization". In: *MOS-SIAM Series on optimization*. Society for Industrial and Applied Mathematics, 2017. DOI: 10.1137/1.9781611974997 (cit. on p. C181).

[3] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota. "Clarke subgradients of stratifiable functions". In: *SIAM J. Optim.* 18 (2007), pp. 556–572. DOI: 10.1137/060670080 (cit. on p. C185).

[4] R. Chartrand. "Exact reconstruction of sparse signals via nonconvex minimization". In: *IEEE Sig. Process. Lett.* 14.10 (2007), pp. 707–710. DOI: 10.1109/LSP.2007.898300 (cit. on p. C177).

[5] D. L. Donoho, M. Elad, and V. N. Temlyakov. "Stable recovery of sparse overcomplete representations in the presence of noise". In: *IEEE Trans. Inform. Theory* 52.1 (2006), pp. 6–18. DOI: 10.1109/TIT.2005.860430 (cit. on p. C177).

[6] L. van den Dries and P. Speissegger. "The field of reals with multisummable series and the exponential function". In: *Proc. London Math. Soc.* 81 (2000), pp. 513–565. DOI: 10.1112/S0024611500012648 (cit. on p. C185).

[7] J. Fan and R. Li. "Variable selection via nonconcave penalized likelihood and Its oracle properties". In: *J. Am. Stat. Assoc.* 96 (2001), pp. 1348–1360. DOI: 10.1198/016214501753382273 (cit. on pp. C177, C179).

[8] T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical learning with sparsity: The lasso and generalizations*. Chapman and Hall, 2015. DOI: 10.1201/b18401 (cit. on pp. C177, C186).

[9] J. Lv and Y. Fan. "A unified approach to model selection and sparse recovery using regularized least squares". In: *Annal. Stat.* 37.6A (2009), pp. 3498–3528. DOI: 10.1214/09-AOS683 (cit. on p. C177).

[10] G. Marjanovic and V. Solo. "On ℓ_q optimization and matrix completion". In: *IEEE Trans. Signal Process.* 60.11 (2012), pp. 5714–5724. DOI: 10.1109/TSP.2012.2212015 (cit. on p. C178).

[11] R. Mazumder, J. H. Friedman, and T. Hastie. "SparseNet: Coordinate descent with nonconvex penalties". In: *J. Am. Stat. Assoc.* 106 (2011), pp. 1125–1138. DOI: 10.1198/jasa.2011.tm09738 (cit. on p. C186).

[12] F. Wen, L. Chu, P. Liu, and R. C. Qiu. "A Survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning". In: *IEEE Access* 6 (2018), pp. 69883–69906. DOI: 10.1109/ACCESS.2018.2880454 (cit. on pp. C178, C181).

[13] C.-H. Zhang. "Nearly unbiased variable selection under minimax concave penalty". In: *Annal. Stat.* 38.2 (2010), pp. 894–942. DOI: 10.1214/09-aos729 (cit. on p. C177).

Author addresses

1. **Josef Dick**, University of New South Wales, AUSTRALIA.
mailto:josef.dick@unsw.edu.au
orcid:https://orcid.org/0000-0003-0142-6022

2. **Guoyin Li**, University of New South Wales, AUSTRALIA.
mailto:g.li@unsw.edu.au
orcid:https://orcid.org/0000-0002-2099-7974

3. **Dinh Duy Tran**, University of New South Wales, AUSTRALIA.
mailto:dinh.tran@unsw.edu.au