

Towards An Efficient Approach for the Nonconvex ℓ_p Ball Projection: Algorithm and Analysis

Xiangyu Yang

*School of Information Science and Technology
ShanghaiTech University
Shanghai, 201210, China
Center for Applied Mathematics of Henan Province
Henan University
Zhengzhou 450046, China*

YANGXY3@SHANGHAITECH.EDU.CN

Jiashan Wang

*Department of Mathematics
University of Washington
Seattle, WA 98195, USA*

JSW1119@GMAIL.COM

Hao Wang

*School of Information Science and Technology
ShanghaiTech University
Shanghai, 201210, China*

WANGHAO1@SHANGHAITECH.EDU.CN

Editor: Lorenzo Rosasco

Abstract

This paper primarily focuses on computing the Euclidean projection of a vector onto the ℓ_p ball in which $p \in (0, 1)$. Such a problem emerges as the core building block in statistical machine learning and signal processing tasks because of its ability to promote the sparsity of the desired solution. However, efficient numerical algorithms for finding the projections are still not available, particularly in large-scale optimization. To meet this challenge, we first derive the first-order necessary optimality conditions of this problem. Based on this characterization, we develop a novel numerical approach for computing the stationary point by solving a sequence of projections onto the reweighted ℓ_1 -balls. This method is practically simple to implement and computationally efficient. Moreover, the proposed algorithm is shown to converge uniquely under mild conditions and has a worst-case $O(1/\sqrt{k})$ convergence rate. Numerical experiments demonstrate the efficiency of our proposed algorithm.

Keywords: Nonsmooth optimization, Sparse regularization, Nonconvex ℓ_p ball projection, Weighted ℓ_1 ball projection, Iterative reweighting methods

1. Introduction

In this paper, we are principally interested in computing the Euclidean projections of a point onto the ℓ_p ball with $p \in (0, 1)$, which can be formulated as the following optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_p^p \leq \gamma, \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^n$ is given, $\|\mathbf{x}\|_p^p := \sum_{i=1}^n |x_i|^p$ and $\gamma > 0$ is a prescribed scalar and is referred to as the radius of ℓ_p ball. Throughout this paper, \mathbf{y} is assumed to be outside the ℓ_p ball, that is, $\|\mathbf{y}\|_p^p > \gamma$; otherwise the computation of such a projection is trivial. Notice that (1) may have multiple minimizers.

Problems of the form (1) arise in inverse problems with a broad spectrum of applications, including image denoising in signal processing, computer vision, and statistical machine learning. In general, the imposed ℓ_p ball constraint can significantly promote sparsity, reduce system complexity, and improve the generalization performance Jain et al. (2017); Liang et al. (2013). We recall several motivating applications with (1) in the following.

1. *The ℓ_p -constrained sparse coding:* In the context of sparse coding Thom et al. (2015), the task is to reconstruct the unknown sparse code word $\mathbf{x}^* \in \mathbb{R}^n$ from the linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \boldsymbol{\xi}$, where $\mathbf{y} \in \mathbb{R}^m$ represents the data with m features, $\boldsymbol{\xi} \in \mathbb{R}^m$ denotes the noise vector, and \mathbf{A} corresponds to the fixed dictionary that consists of n atoms with respect to its columns. This problem can be formulated as

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_p^p \leq \gamma, \quad (2)$$

where the ℓ_p ball constraint is to induce sparsity in the code word. To solve (2), the projected gradient descent (PGD) method is often employed Bahmani and Raj (2013). The efficiency of such an algorithm heavily relies on the computation of projections onto the ℓ_p ball.

2. *Projection-based transform domain denoising:* In standard wavelet denoising Cetin and Tofghi (2015); Grandits and Pock (2017), it could be used as a prior information that the unknown signal \mathbf{x} is sparse Chang et al. (2000). Suppose \mathbf{y} represents the wavelet signals and the radius γ is predetermined, problem (1) is usually a pivotal procedure to be solved.
3. *Adversarial attacks for deep learning:* Adversarial attacks have become a critical task for designing robust neural networks, which considers the security issues of deep neural networks Yuan et al. (2019). Among many advances in generating adversarial examples, a line of works considers the bounded norm-constraints of the perturbation Su et al. (2019), which endeavors to operate a few entries merely. In particular, a central step is to compute the projection of the perturbation vector onto a ℓ_p ball in the black-box attacks setting Balda et al. (2020).

Despite its potential wide applicability, the nonsmooth and nonconvex nature of (1) makes it challenging to solve in general. In principle, it is non-trivial to derive optimality conditions to characterize the optimal solutions of (1). This introduces difficulties in numerical algorithm design and the analysis of the convergence properties of the proposed algorithms.

Due to these aforementioned difficulties, to the extent of our knowledge, not much has been achieved for solving (1). A pioneer study Das Gupta and Kumar (2013) proposed an exhaustive search algorithm, which combines branching and root-finding techniques. By focusing on the local optimality with respect to the nonzero components of the iterates for each branch, they obtained the optimal solution via finding the roots of a univariate

nonlinear equation. It is claimed that the proposed algorithm could find a globally optimal solution for most p with high probability. A recent study Chen et al. (2019) derived a unified characterization of the optimal solution and applied the Lagrangian method to improve Das Gupta and Kumar (2013), where a bisection method is used for finding the optimal multiplier; for each multiplier, this method needs to solve a system of nonlinear equations. By postulating that (1) is addressed by a practical algorithm, the authors in Bahmani and Raj (2013) theoretically investigated the performance of the PGD method in solving the ℓ_p -constrained least squares problems. At this point, the authors of Bahmani and Raj (2013) commented, “*We believe developing an efficient algorithm for projection onto ℓ_p balls with $p \in (0, 1)$ is an interesting problem that can provide a building block for other methods of sparse signal estimation involving the ℓ_p -norm. Furthermore, studying this problem may help to find an insight on how the complexity of these algorithms vary in terms of p .*”

The typical algorithmic frameworks for solving the nonconvex and nonsmooth optimization problems include the majorization-minimization (MM) framework Lange (2016) and the difference-of-convex (DC) framework Le Thi and Dinh (2018). For the MM framework, a great variety of works considered the nonconvex ℓ_p -penalized minimization problems and addressed it by solving a sequence of reweighted ℓ_1 -penalized subproblems Figueiredo et al. (2007); Chartrand and Yin (2008). Our proposed algorithm also falls into the MM algorithmic framework since it generates the sequence by iteratively minimizing a sequence of weighted ℓ_1 -ball projection problems in which the constructed weighted ℓ_1 ball is contained in the ℓ_p ball. It should be highlighted that the ℓ_p ball projection problem generalizes the ℓ_p -penalized problem, as evidenced by Das Gupta and Kumar (2013). On the other hand, the convergence results of the MM framework for constrained optimization investigated in Bolte and Pauwels (2016) are not directly applicable to our algorithm due to the semi-algebraic constraint set assumption. Another line of works investigated the DC programming framework for the sparse constrained optimization problems Boob et al. (2020), but their surrogate functions for the ℓ_p ball constraint are different from ours. We would also like to mention that the updating strategy for the perturbation in our algorithm is novel and is crucial to the convergence analysis.

It is worthy to point out that fast convergence and low complexity properties are usually required when performing ℓ_p ball projections, since (1) usually is a subroutine that is embedded in various algorithmic frameworks. The existing Lagrangian based methods mentioned above generally can not guarantee the feasibility of the obtained solution, and its effectiveness is only demonstrated empirically. On the other hand, the performance of root-finding within Lagrangian based methods degrades in a large-scale setting. To overcome these challenges, we develop the first-order necessary optimality conditions to characterize the solutions of (1), design numerical approaches for solving (1) and analyze the global convergence and convergence rate. The core idea of our work is based on the observation that the ℓ_p ball is locally equivalent to a weighted ℓ_1 ball. Therefore, our approach solves (1) via a sequence of projections onto weighted ℓ_1 -balls, which are constructed by first adding perturbation to smooth the ℓ_p norm and then linearizing the ℓ_p norm at current iterate. Notably, the weighted ℓ_1 ball projection subproblem can be solved accurately by finite terminating algorithms Perez et al. (2020) with $O(n)$ complexity, and these make our methods suitable for large-scale problems. Comparison with the existing state-of-the-art

method highlights the advantage of the proposed algorithm which improves the CPU time by at least 2 orders of magnitude. Moreover, we develop a novel updating strategy for the perturbation parameters such that the algorithm is guaranteed to converge to a first-order stationary solution. The optimality errors are shown to converge with the rate of $O(1/\sqrt{k})$.

1.1 Organization

In the remainder of this section, we outline our notation that will be employed throughout the paper. First-order optimality conditions are derived in section 2. The proposed reweighted ℓ_1 -ball projection algorithm is detailed in section 3. The global convergence and complexity are analyzed in section 4. The results of numerical experiments are presented in section 5. Concluding remarks are provided in section 6.

1.2 Notation and Preliminaries

Throughout this paper, we limit our discussion to \mathbb{R}^n with the standard inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i$, and all boldface lower case letters represent vectors. In particular, $\mathbf{0}$ denotes the zero vector of appropriate size. Let \mathbb{R}_+^n be the nonnegative orthant of \mathbb{R}^n , and \mathbb{R}_{++}^n denote the interior of \mathbb{R}_+^n . Let $\mathbb{N} = \{1, 2, \dots\}$ be the set of natural numbers. For any $n \in \mathbb{N}$, we use $[n]$ to represent the set of integers $\{1, 2, \dots, n\}$. We use $\|\cdot\|_1$ and $\|\cdot\|_2$ to represent the ℓ_1 norm and the ℓ_2 norm of any vector $\mathbf{x} \in \mathbb{R}^n$, respectively. Define $|\mathbf{x}| = (|x_1|, \dots, |x_n|)^T$, $\text{sign}(\mathbf{x}) = (\text{sign}(x_1), \dots, \text{sign}(x_n))^T$ and $[\mathbf{x}]^p = (x_1^p, \dots, x_n^p)^T$ with $p > 0$. Given an index set $\mathcal{S} \subset \mathbb{N}$, we use $|\mathcal{S}|$ to denote the cardinality of \mathcal{S} . Using this notation, we use $\mathbf{I}_{\mathcal{S}\mathcal{S}}$ to represent the submatrix of the identity matrix \mathbf{I} , whose rows and columns are indexed by \mathcal{S} . In addition, $\text{Diag}(\mathbf{x}) \in \mathbb{R}^{n \times n}$ denotes a diagonal matrix with \mathbf{x} forming the diagonal elements. The Hadamard product $\mathbf{a} \circ \mathbf{b} \in \mathbb{R}^n$ for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ is defined element-wisely as $(\mathbf{a} \circ \mathbf{b})_i = a_i b_i$, $i \in [n]$. The active and inactive index set of \mathbf{x} are respectively defined as

$$\mathcal{A}(\mathbf{x}) := \{i \mid x_i = 0, i \in [n]\} \quad \text{and} \quad \mathcal{I}(\mathbf{x}) := \{i \mid x_i \neq 0, i \in [n]\}.$$

Moreover, $x \sim N(\mu, \sigma^2)$ denotes a Gaussian random variable with mean μ and variance σ^2 ; $z \sim U[a, b]$ corresponds to a random variable uniformly distributed on a closed interval $[a, b]$.

We next recall the concept of normal cone presented in Rockafellar and Wets (2009), which plays a crucial role in establishing optimality conditions in nonlinear programming.

Definition 1 *Let $C \subset \mathbb{R}^n$ be closed and $\bar{\mathbf{x}} \in C$. A vector $\mathbf{v} \in \mathbb{R}^n$ is normal to C at $\bar{\mathbf{x}}$ in the regular sense, written $\mathbf{v} \in \widehat{\mathcal{N}}_C(\bar{\mathbf{x}})$, if*

$$\limsup_{\substack{\mathbf{x} \rightarrow \bar{\mathbf{x}} \\ \mathbf{x} \in C \\ \mathbf{x} \neq \bar{\mathbf{x}}}} \frac{\langle \mathbf{v}, \mathbf{x} - \bar{\mathbf{x}} \rangle}{\|\mathbf{x} - \bar{\mathbf{x}}\|_2} \leq 0, \quad \text{for } \mathbf{x} \in C.$$

It is normal to C at $\bar{\mathbf{x}}$ in the general sense, written $\mathbf{v} \in \mathcal{N}_C(\bar{\mathbf{x}})$, if there are sequences $\mathbf{x}^\nu \xrightarrow{C} \bar{\mathbf{x}}$ and $\mathbf{v}^\nu \rightarrow \mathbf{v}$ with $\mathbf{v}^\nu \in \widehat{\mathcal{N}}_C(\mathbf{x}^\nu)$.

2. Proposed Algorithm

In this section, we characterize the properties of the optimal solutions of (1) and derive the optimality conditions. We first show that the projection of \mathbf{y} lies in the same orthant due to the symmetry of the ℓ_p ball. This property is proven to be true for ℓ_1 -ball projections Condat (2016); Duchi et al. (2008). We show this property for a more general weighted ℓ_p -ball projection problem, i.e.,

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \sum_{i=1}^n w_i |x_i|^p \leq \gamma \quad (3)$$

with $p \in (0, 1]$ and $w_i \geq 0$, for all $i \in [n]$. If $w_i = 1$ for all $i \in [n]$ and $p \in (0, 1)$, this problem reverts to (1). If $p = 1$, we have a weighted ℓ_1 -ball projection problem.

Lemma 2 *Let \mathbf{y} be a point satisfying $\sum_{i=1}^n w_i |y_i|^p > \gamma$ and \mathbf{x}^* be a global optimal solution of (3). Then, the following properties hold:*

- (i) $x_i^* y_i \geq 0$ and $|x_i^*| \leq |y_i|$ for all $i \in [n]$.
- (ii) $|\mathbf{x}^*|$ is a global optimal solution of

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - |\mathbf{y}|\|_2^2 \quad \text{s.t.} \quad \sum_{i=1}^n w_i x_i^p \leq \gamma, \quad x_i \in \mathbb{R}_+. \quad (4)$$

Proof We prove (i) by contradiction. Assume there exists $j \in [n]$ for which $x_j^* y_j < 0$. Now consider $\tilde{\mathbf{x}}$ such that

$$\tilde{x}_j = 0 \quad \text{and} \quad \tilde{x}_i = x_i^*, i \neq j.$$

We can see that $\sum_{i=1}^n w_i |\tilde{x}_i|^p \leq \sum_{i=1}^n w_i |x_i^*|^p \leq \gamma$; hence $\tilde{\mathbf{x}}$ is feasible for (3). Note that $(x_j^* - y_j)^2 = (x_j^*)^2 - 2x_j^* y_j + y_j^2 > y_j^2 = (\tilde{x}_j - y_j)^2$ since $x_j^* y_j < 0$, implying $\|\mathbf{x}^* - \mathbf{y}\|_2^2 > \|\tilde{\mathbf{x}} - \mathbf{y}\|_2^2$. This contradicts the global optimality of \mathbf{x}^* . Hence $x_i^* y_i \geq 0$ for all $i \in [n]$.

We also assume by contradiction that there exists $j \in [n]$ for which $|x_j^*| > |y_j|$. Now consider $\hat{\mathbf{x}}$ such that

$$\hat{x}_j = y_j \quad \text{and} \quad \hat{x}_i = x_i^*, \forall i \in [n], i \neq j.$$

We can see that $\sum_{i=1}^n w_i |\hat{x}_i|^p < \sum_{i=1}^n w_i |x_i^*|^p = \gamma$; hence $\hat{\mathbf{x}}$ is feasible. Note that $(x_j^* - y_j)^2 > (\hat{x}_j - y_j)^2 = 0$, which implies that $\|\mathbf{x}^* - \mathbf{y}\|_2^2 > \|\hat{\mathbf{x}} - \mathbf{y}\|_2^2$, contradicting the global optimality of \mathbf{x}^* . Therefore, $|x_i^*| \leq |y_i|$ for all $i \in [n]$.

As for (ii), notice that $\|\mathbf{x}^* - \mathbf{y}\|_2^2 \leq \|\mathbf{x} - \mathbf{y}\|_2^2$ for any feasible \mathbf{x} for (4), which along with statement (i), implies that $\|\mathbf{x}^* - \mathbf{y}\|_2^2 = \|\mathbf{x}^*\|_2^2 - 2\langle \mathbf{x}^*, \mathbf{y} \rangle + \|\mathbf{y}\|_2^2 = \|\mathbf{x}^*| - |\mathbf{y}|\|_2^2 \leq \|\mathbf{x} - \mathbf{y}\|_2^2$ for any \mathbf{x} feasible for (4). This proves that $|\mathbf{x}^*|$ is global optimal for (4). \blacksquare

Lemma 2 (i) indicates that we can restrict the discussions of ℓ_p ball projection problem on \mathbb{R}_+^n . In other words, instead of solving the original projection problem (3), we can instead compute the projection of $|\mathbf{y}|$ onto the ℓ_p ball in the positive orthant. Letting $\bar{\mathbf{y}} := |\mathbf{y}|$, we

can then focus on the following optimization problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) := \frac{1}{2} \|\mathbf{x} - \bar{\mathbf{y}}\|_2^2 \\ \text{s.t.} \quad & \sum_{i=1}^n x_i^p \leq \gamma, \quad \mathbf{x} \in \mathbb{R}_+^n. \end{aligned} \tag{\mathcal{P}}$$

After obtaining the optimal solution \mathbf{x}^* to (\mathcal{P}) , the solution of (1) can thus be recovered by setting $\text{sign}(\mathbf{y}) \circ \mathbf{x}^*$. In addition, we can see that the optimal solution satisfies $x_i^* = 0$ if $y_i = 0$ by Lemma 2 (i).

2.1 First-order Necessary Conditions for Optimality

In this subsection, we derive the first-order necessary optimality conditions to characterize the solutions of (\mathcal{P}) by referring to Wang et al. (2021). For briefly, from now on, we use Ω to denote the feasible region of (\mathcal{P}) , i.e.,

$$\Omega = \{\mathbf{x} \mid \sum_{i=1}^n x_i^p \leq \gamma, \mathbf{x} \in \mathbb{R}_+^n\}. \tag{5}$$

By Rockafellar and Wets (2009, Theorem 6.12), any locally optimal solution of (\mathcal{P}) satisfies the following first-order necessary optimality condition.

Theorem 3 *Let \mathbf{x} be a local solution of (\mathcal{P}) . Then the following holds*

$$\mathbf{x} - \bar{\mathbf{y}} + \mathcal{N}_\Omega(\mathbf{x}) \ni \mathbf{0}. \tag{6}$$

To verify the satisfaction of the necessary condition (6), we need to further investigate the properties of $\mathcal{N}_\Omega(\mathbf{x})$. Note from our assumption that the point to be projected is not in the ball, and the local optimal solution \mathbf{x} to (\mathcal{P}) lies on the boundary of the ℓ_p ball, i.e., $\sum_{i \in \mathcal{I}(\mathbf{x})} x_i^p = \gamma$. The following proposition characterizes the elements of $\mathcal{N}_\Omega(\mathbf{x})$, and its proof mainly follows from Wang et al. (2021, Theorem 2.2 & Theorem 2.6 (c)) and Rockafellar and Wets (2009, Theorem 6.42).

Proposition 4 *Let $\mathbf{x} \in \Omega$. It holds that*

$$\mathcal{N}_\Omega(\mathbf{x}) = \begin{cases} \{\mathbf{v} \in \mathbb{R}_+^n \mid v_i = \lambda p x_i^{p-1}, i \in \mathcal{I}(\mathbf{x}); \lambda \geq 0\}, & \text{if } \sum_{i=1}^n x_i^p = \gamma, \\ \{\mathbf{0}\}, & \text{if } \sum_{i=1}^n x_i^p < \gamma. \end{cases}$$

We depict the geometrical representation of the normal cone at $\mathbf{x} \in \Omega$ for a two-dimensional example in Figure 1.

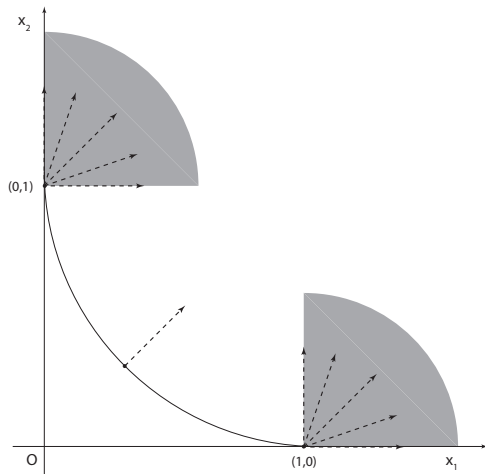


Figure 1: A two-dimensional illustration of the normal cone for the unit $\ell_{0.5}$ -ball constraint in the nonnegative orthant. The normal cone has infinitely many normal directions at $(0, 1)$ and $(1, 0)$ as represented in the shaded region, and it has a single normal direction besides those two points.

Using Proposition 4 and by Theorem 3, the first-order necessary conditions for (\mathcal{P}) are stated as

$$(\bar{y}_i - x_i)x_i = \lambda p x_i^p, \quad i \in [n], \quad (7a)$$

$$\sum_{i \in [n]} x_i^p = \gamma, \quad (7b)$$

$$\mathbf{x} \geq 0, \quad \lambda \geq 0. \quad (7c)$$

where λ is the dual variable/multiplier.

3. An Iteratively Reweighted ℓ_1 Ball Projection Algorithm

In this section, we describe our proposed algorithm for solving (\mathcal{P}) , which can be viewed as an instance of the MM framework. We use a weighted ℓ_1 ball as a majorizer of the perturbed ℓ_p ball in the majorization step, and the majorizer is constructed by relaxing and linearizing the ℓ_p ball constraint. In the minimization step, the next iterate is computed as the projection of $\bar{\mathbf{y}}$ onto the weighted ℓ_1 ball. In this way, our algorithm needs to solve a sequence of weighted ℓ_1 ball projection subproblems, which can be solved efficiently within a finite number of iterations.

3.1 Weighted ℓ_1 -ball Approximation

Now we describe our weighted ℓ_1 ball projection subproblem. First, we add a perturbation $\epsilon_i > 0$ to have a continuous differentiable approximate model with respect to $|x_i|$,

$$|x_i|^p \approx (|x_i| + \epsilon_i)^p.$$

Correspondingly, we denote the perturbed ℓ_p ball constraint in (\mathcal{P}) by $\Omega_\epsilon := \{\mathbf{x} \mid \sum_{i=1}^n (x_i + \epsilon_i)^p \leq \gamma, \mathbf{x} \in \mathbb{R}_+^n\}$. Then at \mathbf{x}^k with perturbation $\epsilon^k > \mathbf{0}$, by concavity of the ℓ_p norm, we approximate $(|x_i| + \epsilon_i^k)^p$ by linearizing it with respect to $|x_i^k|$ to have

$$(|x_i| + \epsilon_i^k)^p \leq (|x_i^k| + \epsilon_i^k)^p + p(|x_i^k| + \epsilon_i^k)^{p-1}(|x_i| - |x_i^k|).$$

Therefore, at the k th iterate \mathbf{x}^k , the constraint in (\mathcal{P}) can be replaced by

$$\sum_{i=1}^n (|x_i^k| + \epsilon_i^k)^p + \sum_{i=1}^n w_i^k (|x_i| - |x_i^k|) \leq \gamma \quad (8)$$

with $w_i^k := w(x_i^k, \epsilon_i^k) = p(|x_i^k| + \epsilon_i^k)^{p-1}$.

Defining $\gamma^k := \gamma - \sum_{i=1}^n (x_i^k + \epsilon_i^k)^p + \sum_{i=1}^n w_i^k |x_i^k|$ and rearranging, we have the k th convex relaxation of (\mathcal{P}) , which reads

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2} \|\mathbf{x} - \bar{\mathbf{y}}\|_2^2 \\ \text{s.t.} \quad & \mathbf{x} \in \Omega_{\mathbf{w}^k} := \{\mathbf{x} \mid \sum_{i=1}^n w_i^k x_i \leq \gamma^k, \mathbf{x} \in \mathbb{R}_+^n\}. \end{aligned} \quad (\mathcal{P}_{\text{sub}})$$

The feasible region of $(\mathcal{P}_{\text{sub}})$ is in fact a weighted ℓ_1 ball intersected with the nonnegative orthant. Since $\bar{\mathbf{y}} \in \mathbb{R}_+^n$, $(\mathcal{P}_{\text{sub}})$ is clearly equivalent to the projection of $\bar{\mathbf{y}}$ onto the weighted ℓ_1 ball $\Omega_{\mathbf{w}^k}$.

The next iterate \mathbf{x}^{k+1} is then computed as the optimal solution of $(\mathcal{P}_{\text{sub}})$. Therefore, \mathbf{x}^{k+1} satisfies the following Karush-Kuhn-Tucker (KKT) conditions for $(\mathcal{P}_{\text{sub}})$

$$\bar{y}_i - x_i^{k+1} - \lambda^{k+1} w_i^k + \mu_i^{k+1} = 0, \quad i \in [n] \quad (9a)$$

$$\langle \boldsymbol{\mu}^{k+1}, \mathbf{x}^{k+1} \rangle = 0, \quad (9b)$$

$$\mu_i^{k+1} \geq 0, \quad i \in [n] \quad (9c)$$

$$\sum_{i \in \mathcal{I}(\mathbf{x}^{k+1})} w_i^k x_i^{k+1} \leq \gamma^k, \quad (9d)$$

$$\lambda^{k+1} \geq 0, \quad (9e)$$

$$\lambda^{k+1} \left(\sum_{i \in \mathcal{I}(\mathbf{x}^{k+1})} w_i^k x_i^{k+1} - \gamma^k \right) = 0, \quad (9f)$$

where λ^{k+1} and $\boldsymbol{\mu}^{k+1}$ are multipliers. Conditions (9a)–(9c) can be equivalently written as

$$\bar{y}_i - \lambda^{k+1} w_i^k \leq 0, \quad i \in \mathcal{A}(\mathbf{x}^{k+1}), \quad (10a)$$

$$\bar{y}_i - x_i^{k+1} - \lambda^{k+1} w_i^k = 0, \quad i \in \mathcal{I}(\mathbf{x}^{k+1}). \quad (10b)$$

Clearly, λ^{k+1} satisfies

$$\lambda^{k+1} = \frac{\sum_{i \in \mathcal{I}(\mathbf{x}^{k+1})} (\bar{y}_i - x_i^{k+1})}{\sum_{i \in \mathcal{I}(\mathbf{x}^{k+1})} w_i^k}. \quad (11)$$

3.2 A Dynamic Updating Strategy for ϵ

Our algorithm solves a sequence of weighted ℓ_1 ball projection subproblems with the relaxation parameters $\epsilon_i, i \in [n]$ driven to 0. However, if we decrease $\epsilon_i, i \in [n]$ too slowly, then the algorithm may suffer from unsatisfactorily slow convergence. On the other hand, if $\epsilon_i, i \in [n]$ were driven to 0 too fast, then the algorithm may quickly get trapped in local solutions or even fail to converge to stationary points. Unlike the usual parameter tuning strategy in the penalized model which tries a sequence of penalty parameters to approximately find one satisfying the needs, we design a dynamic updating strategy Burke et al. (2015) that can automatically determine when $\epsilon_i, i \in [n]$ should be reduced. We update $\epsilon_i, i \in [n]$ according to the following rule

$$\begin{cases} \epsilon^{k+1} \in (\mathbf{0}, \theta \epsilon^k], & \text{if } \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 \|\text{sign}(\mathbf{x}^{k+1} - \mathbf{x}^k) \circ \mathbf{w}^k\|_2^\tau \leq M, \\ \epsilon^{k+1} = \epsilon^k, & \text{otherwise.} \end{cases} \quad (12)$$

with $0 < \theta < 1$, $M > 0$ and $\tau > 1$. As discussed in the numerical study, we can determine these parameters with a rough tuning, in particular τ could be selected close to 1.

Our proposed Iteratively Reweighted ℓ_1 -Ball Projection (IRBP) algorithm for solving (\mathcal{P}) is stated in Algorithm 1.

Algorithm 1 Iteratively Reweighted ℓ_1 Ball Projection Algorithm (IRBP)

- 1: Input: \mathbf{y} , p , and γ . Let $\bar{\mathbf{y}} = |\mathbf{y}|$. Choose $\tau > 1$, $0 < \theta < 1$ and $M > 0$.
- 2: (Initialization) Choose $(\mathbf{x}^0, \epsilon^0) \in \mathbb{R}_+^n \times \mathbb{R}_{++}^n$ such that $\|\mathbf{x}^0\|_p^p \leq \gamma$, $\|\epsilon^0\|_p^p \leq \gamma$ and $\|\mathbf{x}^0 + \epsilon^0\|_p^p \leq \gamma$. Set $k = 0$.
- 3: (Compute the weights) Set $w_i^k = p(|x_i^k| + \epsilon_i^k)^{p-1}$ for $i \in [n]$.
- 4: (Solve the subproblem for \mathbf{x}^{k+1}) Compute \mathbf{x}^{k+1} as the optimal solution of $(\mathcal{P}_{\text{sub}})$.
- 5: (Set the new relaxation vector ϵ^{k+1}) Obtain ϵ^{k+1} according to

$$\begin{cases} \epsilon^{k+1} \in (\mathbf{0}, \theta \epsilon^k], & \text{if } \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 \|\text{sign}(\mathbf{x}^{k+1} - \mathbf{x}^k) \circ \mathbf{w}^k\|_2^\tau \leq M, \\ \epsilon^{k+1} = \epsilon^k, & \text{otherwise.} \end{cases}$$

- 6: Set $k \leftarrow k + 1$ and go to step 3.
-

The initialization requirement of $(\mathbf{x}^0, \epsilon^0)$ in Algorithm 1 is easy to fulfill. For example, we can simply choose $\epsilon_i^0 = (\frac{\gamma}{n})^{1/p}$ such that $\|\epsilon^0\|_p^p < \gamma$ and then set $\mathbf{x}^0 = \mathbf{0}$. On the other hand, the weighted ℓ_1 -ball projection subproblem of the form $(\mathcal{P}_{\text{sub}})$ can be addressed by calling the exact projection algorithms proposed in Michelot (1986); Condat (2016); Zhang et al. (2020) with a straightforward modification.

4. Convergence Analysis of IRBP

This section is devoted to the convergence analysis of the proposed Algorithm 1, including the well-posedness, global convergence, conditions guaranteeing unique convergence, and the worst-case complexity of the optimality residuals.

4.1 Well-posedness and Basic Properties

The following theorem guarantees that the subproblems of our proposed algorithm are always feasible so that the algorithm is well defined.

Theorem 5 (Well-posedness) *It holds true that $\|\mathbf{x}^k + \boldsymbol{\epsilon}^k\|_p^p \leq \gamma$, $\forall k \in \mathbb{N}$. Hence $\gamma^k > 0$, and all subproblems in Algorithm 1 are feasible.*

Proof First, $\|\mathbf{x}^k + \boldsymbol{\epsilon}^k\|_p^p \leq \gamma$ implies that the k th subproblem is feasible, since in this case \mathbf{x}^k satisfies the constraint of the k th subproblem

$$\|\mathbf{x}^k + \boldsymbol{\epsilon}^k\|_p^p + \sum_{i=1}^n p(x_i^k + \epsilon_i^k)^{p-1}(x_i - x_i^k) \leq \gamma.$$

Therefore, we only have to prove $\|\mathbf{x}^k + \boldsymbol{\epsilon}^k\|_p^p \leq \gamma$ holds true for all k .

Since the initial point is chosen such that $\|\mathbf{x}^0 + \boldsymbol{\epsilon}^0\|_p^p \leq \gamma$, by induction, we only have to show that $\|\mathbf{x}^{k+1} + \boldsymbol{\epsilon}^{k+1}\|_p^p \leq \gamma$ provided that $\|\mathbf{x}^k + \boldsymbol{\epsilon}^k\|_p^p \leq \gamma$. \mathbf{x}^{k+1} satisfies the constraint of the k th subproblem, since \mathbf{x}^{k+1} is the optimal solution to the k th subproblem. By concavity of $(\cdot)^p$, we have

$$\|\mathbf{x}^{k+1} + \boldsymbol{\epsilon}^k\|_p^p \leq \|\mathbf{x}^k + \boldsymbol{\epsilon}^k\|_p^p + \sum_{i=1}^n p(x_i^k + \epsilon_i^k)^{p-1}(x_i^{k+1} - x_i^k) \leq \gamma,$$

which, together with $\epsilon_i^{k+1} \leq \epsilon_i^k$ for all $i \in [n]$ and $\mathbf{x}^{k+1} \in \mathbb{R}_+^n$, implies that $\|\mathbf{x}^{k+1} + \boldsymbol{\epsilon}^{k+1}\|_p^p \leq \|\mathbf{x}^{k+1} + \boldsymbol{\epsilon}^k\|_p^p \leq \gamma$. This completes the proof. \blacksquare

In the following lemma, we prove that the objective of (\mathcal{P}) has a quadratic improvement after each iteration, and the displacement of the iterates vanishes in the limit.

Lemma 6 (Descent property) *Let the sequence $\{\mathbf{x}^k\}$ be generated by Algorithm 1. It holds that*

(i) *The sequence $\{\|\mathbf{x}^k - \bar{\mathbf{y}}\|_2^2\}$ monotonically decreases. Indeed,*

$$\|\mathbf{x}^k - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^{k+1} - \bar{\mathbf{y}}\|_2^2 \geq \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_2^2, \forall k \in \mathbb{N}. \quad (13)$$

(ii) *The sequence $\{\mathbf{x}^k\}$ satisfies $\sum_{k=0}^{+\infty} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2 < +\infty$, and consequently, $\lim_{k \rightarrow +\infty} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2 = 0$.*

Proof Since \mathbf{x}^k is feasible to $(\mathcal{P}_{\text{sub}})$, the first-order optimality condition for projection onto a closed convex set yields the desired result of statement (i). For example, see Burke et al. (2015, Theorem 2.1 (2)) and Zarantonello (1971) for the proof details.

As for statement (ii), summing up both sides of (13) from $k = 0$ to t and rearranging, we have

$$\begin{aligned} \sum_{k=0}^t \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2 &\leq \sum_{k=0}^t (\|\mathbf{x}^k - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^{k+1} - \bar{\mathbf{y}}\|_2^2) = \|\mathbf{x}^0 - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^{t+1} - \bar{\mathbf{y}}\|_2^2 \\ &\leq \|\mathbf{x}^0 - \bar{\mathbf{y}}\|_2^2 < +\infty. \end{aligned}$$

Letting $t \rightarrow \infty$, we conclude that $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2 \rightarrow 0$ as $k \rightarrow +\infty$, completing the proof. \blacksquare

The next lemma enumerates relevant properties of the primal-dual solution to the subproblem (\mathcal{P}_{sub}).

Lemma 7 *The following statements hold true:*

- (i) $\mathcal{I}(\mathbf{x}^k) \neq \emptyset$ and $0 \leq x_i^k \leq \bar{y}_i$ for any $i \in [n]$ and $k \in \mathbb{N}$.
- (ii) $\{\mathbf{x}^k\}_{k \geq 1}$ is uniformly bounded away from $\mathbf{0}$.
- (iii) $\sum_{i=1}^n w_i^k x_i^{k+1} = \gamma^k$.
- (iv) $\{\lambda^k\}$ is bounded above.

Proof

- (i) For any k , $w_i^k < +\infty$ since $\epsilon_i^k > 0$. Therefore, the interior of the feasible set of the projection subproblem (\mathcal{P}_{sub}) is not empty. Lemma 2 shows that $\mathbf{0}$ is not the projection of $\bar{\mathbf{y}}$ onto the weighted ℓ_1 ball, meaning $\mathcal{I}(\mathbf{x}^{k+1}) \neq \emptyset$. The rest of statement (i) is simply a special case of Lemma 2 with $p = 1$.
- (ii) We first show that $\{\mathbf{x}^k\}_{k \geq \bar{k}}$ is uniformly bounded away from $\mathbf{0}$, where \bar{k} is such that $\mathbf{x}^{\bar{k}} \neq \mathbf{0}$. From Lemma 6 (i), we have $\|\mathbf{x}^k - \bar{\mathbf{y}}\|_2 \leq \|\mathbf{x}^{\bar{k}} - \bar{\mathbf{y}}\|_2$ for any $k \geq \bar{k}$. Therefore, for any $k \geq \bar{k}$, $\|\mathbf{x}^k\|_2 \geq \|\hat{\mathbf{x}}\|_2$ with

$$\hat{\mathbf{x}} := \arg \min_{\mathbf{x}} \{\|\mathbf{x}\|_2^2 \mid \|\mathbf{x} - \bar{\mathbf{y}}\|_2^2 \leq \|\mathbf{x}^{\bar{k}} - \bar{\mathbf{y}}\|_2^2\}. \quad (14)$$

It can be easily shown that $0 < \|\hat{\mathbf{x}}\|_2 = \|\bar{\mathbf{y}}\|_2 - \|\mathbf{x}^{\bar{k}} - \bar{\mathbf{y}}\|_2$ (please refer to Appendix A.1 for its proof), where the inequality is by $\mathbf{x}^{\bar{k}} \neq \mathbf{0}$ and statement (i). Therefore, $\{\|\mathbf{x}^k\|_2\}_{k \geq 1}$ is uniformly bounded away from 0.

Now it suffices to show that $\mathbf{x}^1 \neq \mathbf{0}$. If $\mathbf{x}^0 \neq \mathbf{0}$, then it is done. If $\mathbf{x}^0 = \mathbf{0}$, then $\|\epsilon^0\|_p^p < \gamma$ and the initial subproblem is

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \bar{\mathbf{y}}\|_2^2, \quad \text{s.t.} \quad \sum_{i=1}^n p(\epsilon_i^0)^{p-1} x_i \leq \gamma - \|\epsilon^0\|_p^p, \quad \mathbf{x} \in \mathbb{R}_+^n. \quad (15)$$

Consider $\hat{\mathbf{x}} := \frac{\gamma - \|\epsilon^0\|_p^p}{\sum_{i=1}^n p(\epsilon_i^0)^{p-1} \bar{y}_i} \bar{\mathbf{y}}$. It is clear that $\hat{\mathbf{x}}$ is feasible for (15) and $\frac{1}{2} \|\hat{\mathbf{x}} - \bar{\mathbf{y}}\|_2^2 < \frac{1}{2} \|\mathbf{0} - \bar{\mathbf{y}}\|_2^2$, implying $\mathbf{x} = \mathbf{0}$ is not optimal for (15). Therefore, $\mathbf{x}^1 \neq \mathbf{0}$, completing the proof of statement (ii).

- (iii) Suppose by contradiction that the projection of $\bar{\mathbf{y}}$ onto the weighted ℓ_1 ball of (\mathcal{P}_{sub}) satisfies $\sum_{i=1}^n w_i^k x_i^{k+1} < \gamma^k$, it happens that $\bar{\mathbf{y}}$ is also in the weighted ℓ_1 ball, i.e., $\sum_{i=1}^n w_i^k \bar{y}_i \leq \gamma^k$. Therefore,

$$\|\bar{\mathbf{y}}\|_p^p \leq \|\bar{\mathbf{y}} + \epsilon^k\|_p^p \leq \|\mathbf{x}^k + \epsilon^k\|_p^p + \sum_{i=1}^n p(x_i^k + \epsilon_i^k)^{p-1} (\bar{y}_i - x_i^k) \leq \gamma, \quad (16)$$

contradicting our assumption that $\|\bar{\mathbf{y}}\|_p^2 > \gamma$. This completes the proof of statement (iii).

- (iv) For any $i \in \mathcal{I}(\mathbf{x}^k)$, $\{w_i^k\}$ is bounded away from 0 and $\sum_{i \in \mathcal{I}(\mathbf{x}^k)} (\bar{y}_i - x_i^k)$ is bounded above according to statement (i). It follows from (11) that $\{\lambda^k\}$ is bounded above. ■

4.2 Convergence Results

In this section, we prove that any cluster point of the sequence generated by the IRBP is first-order stationary for (1). To enforce the convergence of the iterates to the stationary points of the original problem, our goal is to drive $\boldsymbol{\epsilon}^k$ to $\mathbf{0}$ as the algorithm proceeds, so that the relaxed ball $\|\mathbf{x} + \boldsymbol{\epsilon}\|_p \leq \gamma$ can approximate the ℓ_p ball in the limit. We first show that the update (12) of $\boldsymbol{\epsilon}$ is triggered infinitely many times. Define the set of iterations where $\boldsymbol{\epsilon}$ is decreased as

$$\mathcal{U} := \{k \in \mathbb{N} \mid \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 \|\text{sign}(\mathbf{x}^{k+1} - \mathbf{x}^k) \circ \mathbf{w}^k\|_2^\tau \leq M\}.$$

Then we have the following result, and its proof is provided in Appendix A.2.

Lemma 8 *The updating condition (12) is triggered infinitely many times, i.e., $|\mathcal{U}| = +\infty$, indicating $\lim_{k \rightarrow +\infty} \boldsymbol{\epsilon}^k = \mathbf{0}$.*

For ease of presentation, we define four cluster point sets as follows.

$$\begin{aligned} \Gamma &:= \{\text{cluster points of } \{(\mathbf{x}^k, \lambda^k)\}\}, & \Gamma_{\mathcal{U}} &:= \{\text{cluster points of } \{(\mathbf{x}^k, \lambda^k)\}_{\mathcal{U}}\}, \\ \Lambda &:= \{\text{cluster points of } \{\lambda^k\}\}, & \Lambda_{\mathcal{U}} &:= \{\text{cluster points of } \{\lambda^k\}_{\mathcal{U}}\}. \end{aligned}$$

The following property shows that for sufficiently large k , the zero and nonzero components of $\{\mathbf{x}^k\}_{\mathcal{U}}$ remain unchanged. For simplicity, we use shorthands \mathcal{I}^k and \mathcal{A}^k to represent $\mathcal{I}(\mathbf{x}^k)$ and $\mathcal{A}(\mathbf{x}^k)$, respectively. Please refer to Appendix A.3 for its proof.

Proposition 9 *Assume $0 \notin \Lambda_{\mathcal{U}}$, then the following statements hold:*

- (i) *There exists \bar{k} such that $\mathcal{U} \supseteq \{k \mid k \geq \bar{k}\}$, which implies the update of $\boldsymbol{\epsilon}$ in (12) is always triggered for sufficiently large k .*
- (ii) *If $\{w_i^k\}$ is unbounded, then $x_i^k \equiv 0$ for all sufficiently large k . If $\{w_i^k\}$ is bounded above, then there exists $\delta_i > 0$ such that $x_i^k > \delta_i > 0$ for all sufficiently large k .*
- (iii) *There exist index sets $\mathcal{I}^* \cup \mathcal{A}^* = [n]$ such that $\mathcal{I}(\mathbf{x}^*) = \mathcal{I}^*$ and $\mathcal{A}(\mathbf{x}^*) = \mathcal{A}^*$ for any $(\mathbf{x}^*, \lambda^*) \in \Gamma$. Furthermore, $\mathcal{I}(\mathbf{x}^k) = \mathcal{I}^*$ and $\mathcal{A}(\mathbf{x}^k) = \mathcal{A}^*$ for all sufficiently large k .*

The next corollary follows directly from Proposition 9.

Corollary 10 *It holds that*

(i) $\Lambda_{\mathcal{U}} \subseteq \Lambda$.

(ii) If $0 \notin \Lambda$, then $\Lambda = \Lambda_{\mathcal{U}}$.

(iii) If $0 \in \Lambda$, then $0 \in \Lambda_{\mathcal{U}}$.

Considering a primal-dual pair (\mathbf{x}, λ) with $\mathbf{x} \geq 0$ and $\lambda > 0$, we define the following metrics to measure the optimality residuals at (\mathbf{x}, λ)

$$\alpha(\mathbf{x}, \lambda) := \sum_{i=1}^n |(\bar{y}_i - x_i)x_i - \lambda p x_i^p|, \quad \beta(\mathbf{x}) := \left| \sum_{i=1}^n x_i^p - \gamma \right|. \quad (17)$$

It is obvious that (\mathbf{x}, λ) is first-order stationary if and only if $\alpha(\mathbf{x}, \lambda) = 0$ and $\beta(\mathbf{x}) = 0$.

The following property of the optimality residuals is needed in our analysis, and its proof is provided in Appendix A.4.

Lemma 11 *There exists $\tilde{\mathbf{x}}$ with \tilde{x}_i^k between x_i^k and $x_i^{k-1} + \epsilon_i^{k-1}$, $i \in \mathcal{I}^k$ such that*

$$\alpha(\mathbf{x}^k, \lambda^k) \leq \lambda^k p(1-p) \|\tilde{\mathbf{x}}_{\mathcal{I}^k}^k\|^{p-2} \|\bar{\mathbf{y}}\|_{\infty} \left[\|\mathbf{x}_{\mathcal{I}^k}^{k-1} - \mathbf{x}_{\mathcal{I}^k}^k\|_2 + \|\boldsymbol{\epsilon}_{\mathcal{I}^k}^k\|_2 \right]. \quad (18)$$

There exists $\hat{\mathbf{x}}$ with \hat{x}_i^k between x_i^k and $x_i^{k-1} + \epsilon_i^{k-1}$, $i \in \mathcal{I}^k$ such that

$$\begin{aligned} \beta(\mathbf{x}^k) &\leq (\|\mathbf{w}_{\mathcal{I}^k}^{k-1}\|_2 + p \|\hat{\mathbf{x}}_{\mathcal{I}^k}^k\|^{p-1}) \|\mathbf{x}_{\mathcal{I}^k}^{k-1} - \mathbf{x}_{\mathcal{I}^k}^k\|_2 \\ &\quad + p \|\hat{\mathbf{x}}_{\mathcal{I}^k}^k\|^{p-1} \|\boldsymbol{\epsilon}_{\mathcal{I}^k}^{k-1}\|_1 + \|\boldsymbol{\epsilon}_{\mathcal{A}^k}^{k-1}\|^p + \|\boldsymbol{\epsilon}_{\mathcal{A}^k}^{k-1}\|^{p-1} \circ \mathbf{x}_{\mathcal{A}^k}^{k-1} \|_1. \end{aligned} \quad (19)$$

The main result on the global convergence is stated in the following theorem.

Theorem 12 *Let $\{(\mathbf{x}^k, \lambda^k)\}$ be the sequence generated by Algorithm 1. Then for all $(\mathbf{x}^*, \lambda^*) \in \Gamma_{\mathcal{U}}$, $(\mathbf{x}^*, \lambda^*)$ is first-order stationary for (\mathcal{P}) . Consequently, we have*

$$\lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{U}}} \alpha(\mathbf{x}^k, \lambda^k) = 0 \quad \text{and} \quad \lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{U}}} \beta(\mathbf{x}^k) = 0. \quad (20)$$

Proof Given $(\mathbf{x}^*, \lambda^*) \in \Gamma_{\mathcal{U}}$, there exists $\mathcal{S} \subseteq \mathcal{U}$ such that $\lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{S}}} (\mathbf{x}^k, \lambda^k) = (\mathbf{x}^*, \lambda^*)$. It is easily seen that there exists $\bar{k} > 0$ such that $\mathcal{I}(\mathbf{x}^*) \subseteq \mathcal{I}(\mathbf{x}^k)$ for all $k \geq \bar{k}$ and $k \in \mathcal{S}$. Lemma 7 (i) implies that $\mathcal{I}(\mathbf{x}^*)$ is non-empty.

$$\begin{aligned} \alpha(\mathbf{x}^*, \lambda^*) &= \lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{S}}} \alpha(\mathbf{x}^k, \lambda^k) \\ &\leq \lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{S}}} \lambda^k p(1-p) \|\tilde{\mathbf{x}}_{\mathcal{I}^k}^k\|^{p-2} \|\bar{\mathbf{y}}\|_{\infty} \left[\|\mathbf{x}_{\mathcal{I}^k}^{k-1} - \mathbf{x}_{\mathcal{I}^k}^k\|_2 + \|\boldsymbol{\epsilon}_{\mathcal{I}^k}^k\|_2 \right] \\ &= 0, \end{aligned} \quad (21)$$

where the inequality is by Lemma 11 and the second equality is by Lemma 6 (ii) and Lemma 8.

Applying the same argument for $\beta(\mathbf{x}^k)$, we have

$$\beta(\mathbf{x}^*) = \lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{S}}} \beta(\mathbf{x}^k) = 0.$$

Hence $(\mathbf{x}^*, \lambda^*)$ is stationary for (\mathcal{P}) .

Overall, we have shown for any convergent subsequence $\mathcal{S} \subseteq \mathcal{U}$,

$$\lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{S}}} \alpha(\mathbf{x}^k, \lambda^k) = 0 \quad \text{and} \quad \lim_{\substack{k \rightarrow +\infty \\ k \in \mathcal{S}}} \beta(\mathbf{x}^k) = 0,$$

hence, both $\alpha(\mathbf{x}^k, \lambda^k)$ and $\beta(\mathbf{x}^k)$ converge to 0 on \mathcal{U} . ■

4.3 Uniqueness of the Cluster Points

Theorem 12 shows the convergence of optimality residuals and all cluster points of $\{\mathbf{x}^k\}_{\mathcal{U}}$ are first-order stationary for (\mathcal{P}) . In this subsection, we discuss conditions to guarantee the convergence of the entire sequence $\{\mathbf{x}^k\}$.

The following lemma states the properties of the cluster set Γ .

Lemma 13 *Γ is non-empty, compact and connected. For every cluster point $(\mathbf{x}^*, \lambda^*) \in \Gamma$, they have the same objective value.*

Proof $\{(\mathbf{x}^k, \lambda^k)\}$ is bounded by Lemma 7 (i) and (iv), hence it has cluster points. Γ is compact and connected by Lemma 6 (ii) and Bauschke et al. (2015, Lemma 2.6) Ostrowski (1973, Theorem 26.1). Lemma 6 (i) shows all points in Γ have the same objective value. This completes the proof. ■

The following theorem asserts the convergence property of the entire sequence generated by IRBP, and its proof can be found in Appendix A.5.

Theorem 14 *Assume $0 \notin \Lambda$ and there exists a cluster point \mathbf{x}^* of $\{\mathbf{x}^k\}$ such that $x_i^* \neq [\lambda^* p(1-p)]^{\frac{1}{2-p}}, \forall i \in \mathcal{I}^*$. Then $\lim_{k \rightarrow +\infty} \mathbf{x}^k = \mathbf{x}^*$.*

We have shown that when $0 \notin \Lambda$, $\{\mathbf{x}^k\}$ converges to a unique point. Next we demonstrate that when $0 \in \Lambda$, to check if a point $(\mathbf{x}^*, 0)$ is stationary for (\mathcal{P}) or not is NP-hard.

Proposition 15 (NP-hardness) *It is NP-hard to determine whether (\mathcal{P}) has a first-order stationary point $(\mathbf{x}^*, \lambda^*)$ with $\lambda^* = 0$.*

Proof We reduce the NP-hard subset-sum problem Toth and Martello (1990); Wojtczak (2018) to our problem. Consider a set of positive real numbers $\mathbf{s} = \{s_1, \dots, s_n\}$ and $\gamma > 0$. The subset sum problem is to determine whether there exists a subset $s_i, i \in \mathcal{I}$ satisfying $\sum_{i \in \mathcal{I}} s_i = \gamma$. Obviously, this problem can be reduced to the problem of determining whether (7) with $\bar{y}_i := s_i^{1/p}, i \in [n]$ has a solution $(\mathbf{x}^*, \lambda^*)$ with $\lambda^* = 0$. Hence, to determine whether (\mathcal{P}) has a first-order stationary point $(\mathbf{x}^*, \lambda^*)$ with $\lambda^* = 0$ is NP-hard. ■

4.4 Local Complexity Analysis

This subsection establishes a $O(1/\sqrt{k})$ convergence rate for the optimality residuals in an ergodic sense.

The proof of the following local complexity result can be found in Appendix A.6.

Lemma 16 *Assume $0 \notin \Lambda$. Then there exist $\bar{k} \in \mathcal{U}$ and positive constants C_α and C_β such that for all $k \geq \bar{k}$*

$$\min_{\bar{k} \leq t \leq \bar{k}+k} \alpha(\mathbf{x}^t, \lambda^t) \leq \sqrt{C_\alpha/k} \quad \text{and} \quad \min_{\bar{k} \leq t \leq \bar{k}+k} \beta(\mathbf{x}^t) \leq \sqrt{C_\beta/k}.$$

4.5 Worst-case Complexity Analysis

The purpose of this subsection is to analyze the worst-case complexity of IRBP. A primal-dual pair $(\tilde{\mathbf{x}}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}_+$ is an ε -optimal solution to (\mathcal{P}) if

$$\alpha(\tilde{\mathbf{x}}, \tilde{\lambda}) \leq \varepsilon \quad \text{and} \quad \beta(\tilde{\mathbf{x}}) \leq \varepsilon \quad (22)$$

with given tolerance $\varepsilon > 0$.

The iteration complexity of Algorithm 1 is established in the following theorem.

Theorem 17 *Given tolerance $\varepsilon > 0$. Let $\boldsymbol{\epsilon} \in \mathbb{R}_{++}^n$ satisfy the following*

$$\|\boldsymbol{\epsilon}\|_p^p < \gamma, \quad p\|\bar{\mathbf{y}}\|_1(\|\bar{\mathbf{y}}\|_\infty + \|\boldsymbol{\epsilon}\|_\infty)^{1-p}\|\boldsymbol{\epsilon}\|_p^p \leq \varepsilon/2 \quad \text{and} \quad 2^p\|\boldsymbol{\epsilon}\|_p^p \leq \varepsilon/2. \quad (23)$$

Assume that $\{(\mathbf{x}^k, \boldsymbol{\epsilon}^k)\}$ is generated by Algorithm 1 with initial relaxation vector $\boldsymbol{\epsilon}^0 = \boldsymbol{\epsilon}$, and the relaxation vector is kept fixed so that $\boldsymbol{\epsilon}^k = \boldsymbol{\epsilon}$ for all $k \in \mathbb{N}$. Then in at most $O(1/\varepsilon^2)$ iterations, $(\mathbf{x}^k, \boldsymbol{\epsilon}^k)$ is an ε -optimal solution to (\mathcal{P}) .

The proof of such a result depends upon a few preliminary results. To proceed, we first define the relaxed optimal residuals with relaxation vector $\boldsymbol{\epsilon}$ as

$$\begin{aligned} \alpha_\boldsymbol{\epsilon}(\mathbf{x}, \lambda) &= \sum_{i=1}^n |(\bar{y}_i - x_i)x_i - \lambda p(x_i + \epsilon_i)^{p-1}x_i|, \\ \beta_\boldsymbol{\epsilon}(\mathbf{x}) &= \left| \sum_{i=1}^n (x_i + \epsilon_i)^p - \gamma \right|. \end{aligned} \quad (24)$$

In the following analysis, we first show that $\alpha_\boldsymbol{\epsilon}(\mathbf{x}, \lambda)$ and $\beta_\boldsymbol{\epsilon}(\mathbf{x})$ can be used to approximate $\alpha(\mathbf{x}, \lambda)$ and $\beta(\mathbf{x})$, respectively. Moreover, the approximation errors can be controlled by $\boldsymbol{\epsilon}$ as shown in the following lemma and its proof is provided in Appendix A.7.

Lemma 18 *Suppose $(\mathbf{x}, \boldsymbol{\epsilon})$ with $\mathbf{0} \leq \mathbf{x} \leq \bar{\mathbf{y}}$, $\boldsymbol{\epsilon} \in \mathbb{R}_{++}^n$ and $\|\boldsymbol{\epsilon}\|_p^p < \gamma$. It then satisfies*

$$|\alpha(\mathbf{x}, \lambda) - \alpha_\boldsymbol{\epsilon}(\mathbf{x}, \lambda)| \leq \lambda p \|\boldsymbol{\epsilon}\|_p^p \quad \text{and} \quad |\beta(\mathbf{x}) - \beta_\boldsymbol{\epsilon}(\mathbf{x})| < 2^p \|\boldsymbol{\epsilon}\|_p^p.$$

For ease of exposition throughout this subsection, we assume that the hypotheses of Theorem 17 hold. Next, we prove that $\{\alpha_\boldsymbol{\epsilon}(\mathbf{x}^k, \lambda^k)\}$ and $\{\beta_\boldsymbol{\epsilon}(\mathbf{x}^k)\}$ vanish in the following theorem, and the proof can be found in Appendix A.8.

Lemma 19 *Let the hypotheses of Theorem 17 hold and $\{(\mathbf{x}^k, \lambda^k)\}$ be generated by Algorithm 1. The following statements hold:*

$$(i) \quad 0 \leq \lambda^k \leq \kappa_3/p,$$

$$(ii) \quad \alpha_\epsilon(\mathbf{x}^k, \lambda^k) \leq \kappa_3 \|\bar{\mathbf{y}}\|_2 (1-p) \|\epsilon^{p-2}\|_2 \|\mathbf{x}^{k-1} - \mathbf{x}^k\|_2,$$

$$(iii) \quad \beta_\epsilon(\mathbf{x}^k) \leq p \|\epsilon^{p-1}\|_2 \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2,$$

where $\kappa_3 := \frac{\|\bar{\mathbf{y}}\|_1}{(\|\bar{\mathbf{y}}\|_\infty + \|\epsilon\|_\infty)^{p-1}}$.

The following corollary follows directly from Lemma 19.

Corollary 20 *Given that ϵ is a vector of ϵ , i.e., $\epsilon = (\epsilon, \dots, \epsilon)^T$, we have $\alpha_\epsilon(\mathbf{x}^k, \lambda^k) \leq \frac{\sqrt{n}(1-p)\epsilon^{p-2} \|\bar{\mathbf{y}}\|_1 \|\bar{\mathbf{y}}\|_2}{(\|\bar{\mathbf{y}}\|_\infty + \epsilon)^{p-1}} \|\mathbf{x}^{k-1} - \mathbf{x}^k\|_2$.*

In what follows, we prove the iteration complexity for an approximate critical point satisfying (22), and the proof can be found in Appendix A.9.

Lemma 21 *Let the hypotheses of Theorem 17 hold and $\{(\mathbf{x}^k, \lambda^k)\}$ be generated by Algorithm 1, and let \mathbf{x}^* be a cluster point of the sequence $\{\mathbf{x}^k\}_U$. Given the tolerance $\varepsilon > 0$, then Algorithm 1 terminates in finite many iterations to reach ε -optimality, i.e. $\max\{\alpha_\epsilon(\mathbf{x}^k, \lambda^k), \beta_\epsilon(\mathbf{x}^k)\} < \varepsilon/2$. Furthermore, the maximum iteration number to reach ε -optimality is given by $\max(\kappa_\alpha/\varepsilon^2, \kappa_\beta/\varepsilon^2)$ where κ_α and κ_β are constants as defined below*

$$\begin{aligned} \kappa_\alpha &= 4(1-p)^2 \kappa_3^2 \|\bar{\mathbf{y}}\|_2^2 \|\epsilon^{p-2}\|_2^2 (\|\mathbf{x}^0 - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^* - \bar{\mathbf{y}}\|_2^2), \\ \kappa_\beta &= 4p^2 \|\epsilon^{p-1}\|_2^2 (\|\mathbf{x}^0 - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^* - \bar{\mathbf{y}}\|_2^2), \end{aligned}$$

and κ_3 is as defined in Lemma 19.

Now we prove Theorem 17 in the following.

Proof By Lemma 21, IRBP performs at most $O(1/\varepsilon^2)$ iterations to reach

$$\alpha_\epsilon(\mathbf{x}^k, \lambda^k) \leq \varepsilon/2 \quad \text{and} \quad \beta_\epsilon(\mathbf{x}^k) \leq \varepsilon/2. \quad (25)$$

On the other hand, by Lemma 18, Lemma 19 (i) and (23), we have

$$|\alpha(\mathbf{x}^k, \lambda^k) - \alpha_\epsilon(\mathbf{x}^k, \lambda^k)| \leq \lambda^k p \|\epsilon\|_p^p \leq \kappa_3 \|\epsilon\|_p^p \leq \varepsilon/2, \quad (26)$$

$$|\beta(\mathbf{x}^k) - \beta_\epsilon(\mathbf{x}^k)| < 2^p \|\epsilon\|_p^p \leq \varepsilon/2. \quad (27)$$

Combining (25), (26) and (27), it gives

$$\alpha(\mathbf{x}^k, \lambda^k) \leq |\alpha(\mathbf{x}^k, \lambda^k) - \alpha_\epsilon(\mathbf{x}^k, \lambda^k)| + \alpha_\epsilon(\mathbf{x}^k, \lambda^k) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon, \quad (28)$$

and similarly,

$$\beta(\mathbf{x}^k) \leq |\beta(\mathbf{x}^k) - \beta_\epsilon(\mathbf{x}^k)| + \beta_\epsilon(\mathbf{x}^k) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon. \quad (29)$$

This completes the proof. ■

5. Numerical Experiments

In this section, we conduct a set of numerical experiments on synthetic data to demonstrate the efficiency of the proposed IRBP for solving the ℓ_p ball projection problems. The code is implemented in Python.¹ and all experiments are performed on an Intel Core CPU i7-7500U at 2.7GHz×4 with 7.5GB of main memory, running under a Ubuntu 64-bit laptop.

5.1 Data Description and Implementation Details

In all tests, we consider projecting $\mathbf{y} \in \mathbb{R}^n$ onto the unit ℓ_p ball (i.e., $\gamma = 1$). Motivated by the tests on ℓ_1 -ball projections in Condat (2016), the projection vector is generated as $\mathbf{y} \sim N(\gamma/n, 10^{-3})$ by discarding \mathbf{y} that lies inside the unit ℓ_p ball.

The input parameters for Algorithm 1 are set as

$$\theta = \min(\beta(\mathbf{x}^k), 1/\sqrt{k})^{1/p}, \tau = 1.1 \text{ and } M = 100.$$

Meanwhile, we initialize $\mathbf{x}^0 = \mathbf{0}$ and $\boldsymbol{\epsilon}^0 = 0.9(\frac{\gamma}{\|\boldsymbol{\nu}\|_1}\boldsymbol{\nu})^{1/p}$ such that $\|\boldsymbol{\epsilon}^0\|_p^p < \gamma$, where entries of $\boldsymbol{\nu} \in \mathbb{R}^n$ are generated from i.i.d. $U[0, 1]$. The algorithm is terminated when the following condition is satisfied

$$\max\{\bar{\alpha}(\mathbf{x}^k, \lambda^k), \beta(\mathbf{x}^k)\} \leq \delta^{\text{tol}} \max(\bar{\alpha}(\mathbf{x}^0, \lambda^0), \beta(\mathbf{x}^0), 1), \quad (30)$$

where $\bar{\alpha}(\mathbf{x}^k, \lambda^k) = \frac{1}{n}\alpha(\mathbf{x}^k, \lambda^k)$ and $\delta^{\text{tol}} = 10^{-8}$.

To our knowledge, the root-searching procedure (RS for short in the sequel) proposed in Chen et al. (2019) was reported to be the most successful algorithm for ℓ_p -ball projection so far. Therefore, we deliver a set of performance comparison experiments between the RS and the IRBP. We summarize the RS method in Algorithm 2.

On the other hand, if RS fails to find an optimal multiplier λ^* satisfying $\frac{1}{n}\|\|\mathbf{x}\|_p^p - \gamma| < 10^{-8}$, the run is considered as a failure. Similarly, any time IRBP fails to satisfy (30) within 10^3 iterations, we deem this run as a failure.

5.2 Test Results

First experiment: A 2-dimension illustrative example. In this example, we show the iteration path of IRBP.

Example 1 Given $\gamma = 1$, $p = 0.5$, and $\mathbf{y} = [0.5, 0.45]^T$, then the problem (1) takes the form

$$\min_{\mathbf{x} \in \mathbb{R}^2} \frac{1}{2} [(x_1 - 0.5)^2 + (x_2 - 0.45)^2] \quad \text{s.t.} \quad \sqrt{|x_1|} + \sqrt{|x_2|} \leq 1. \quad (31)$$

The iteration is started with $(\mathbf{x}^0, \boldsymbol{\epsilon}^0) = ([0, 0]^T, [0.072, 0.463]^T)$, where $\boldsymbol{\epsilon}^0$ is generated randomly as mentioned above. After 22 iterations, it converges to one global optimal solution $\mathbf{x}^* = [0.2972, 0.2069]^T$.

We plot the iteration path of IRBP in Figure 2 using the above initial point. In particular, some of the weighted ℓ_1 -balls in the subproblems are also plotted. Notice that the iterates always remain within the ℓ_p ball and eventually move towards the boundary of the ball.

1. We have made the code publicly available at <https://github.com/Optimizater/Lp-ball-Projection>.

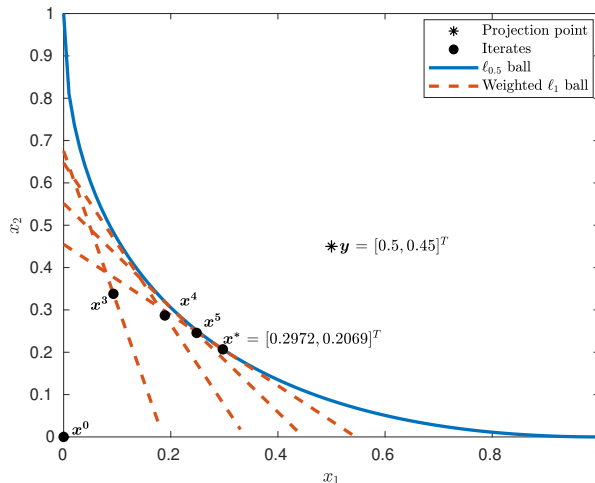


Figure 2: The iteration path of projecting $\mathbf{y} = [0.5, 0.45]^T$ onto the unit $\ell_{0.5}$ -ball. In particular, the solid circles denote the iterates with the randomly generated $\epsilon^0 = [0.072, 0.463]^T$. The dashed line represents the boundary of a weighted ℓ_1 ball constructed at each subproblem.

Second experiment: Performance comparison. In this test, we compare the elapsed wall-clock time of IRBP and RS in successful runs. The comparison is done for $n = 10^2$ and $p \in \{0.4, 0.8\}$ with 100 randomly generated test problems for each p . We compare the success ratio versus the elapsed wall-clock time and plot the performance profiles in Figure 3. We make the following observations:

- For both cases, we can see that IRBP always achieves a higher success rate compared to RS. In particular, IRBP successfully solves all testing instances for both cases, whereas RS can only handle about 25% of instances for $p = 0.4$ and about 70% for $p = 0.8$. On the other hand, both algorithms become more efficient for larger p values.
- IRBP is superior to RS in terms of elapsed wall-clock time. For most of the test examples, the computational time needed by IRBP is less than 0.1s for $p = 0.4$ and 0.01s for $p = 0.8$. Notice that the subproblem solver applied here has an observed complexity $O(n)$ in many practical contexts. We are well aware that the algorithm can be further accelerated if one can design a faster method for the weighted ℓ_1 ball.

Third experiment: IRBP solves high-dimension problems. In this experiment, we demonstrate the efficiency of IRBP for solving large-scale problems. The RS algorithm is not included in this experiment since it fails to converge within the time limit. With $p \in \{0.4, 0.8\}$ and $n \in \{10^3, 10^4, 10^5, 10^6\}$. Figure 4 reports the computation time, and the results are averaged over 20 randomly generated test problems.

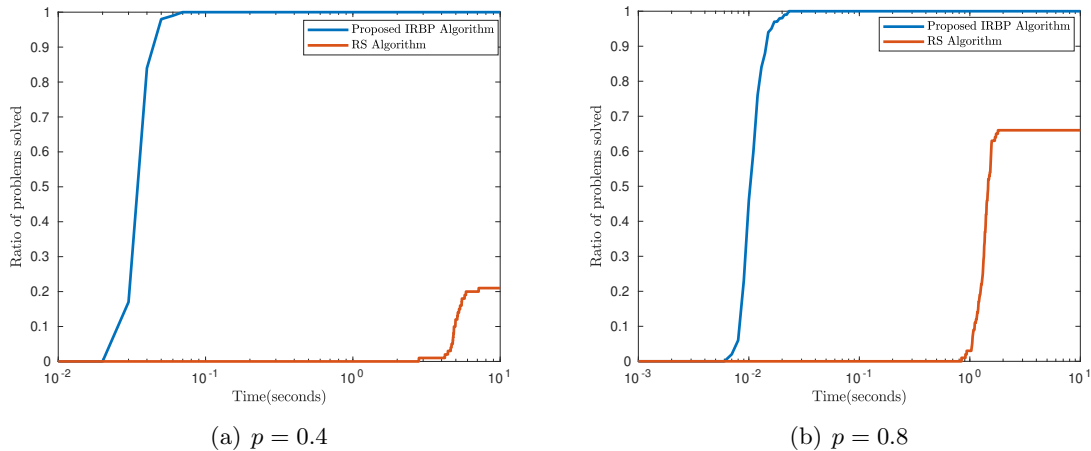


Figure 3: Performance profiles for IRBP and RS with $p \in \{0.4, 0.8\}$ and $n = 10^2$.

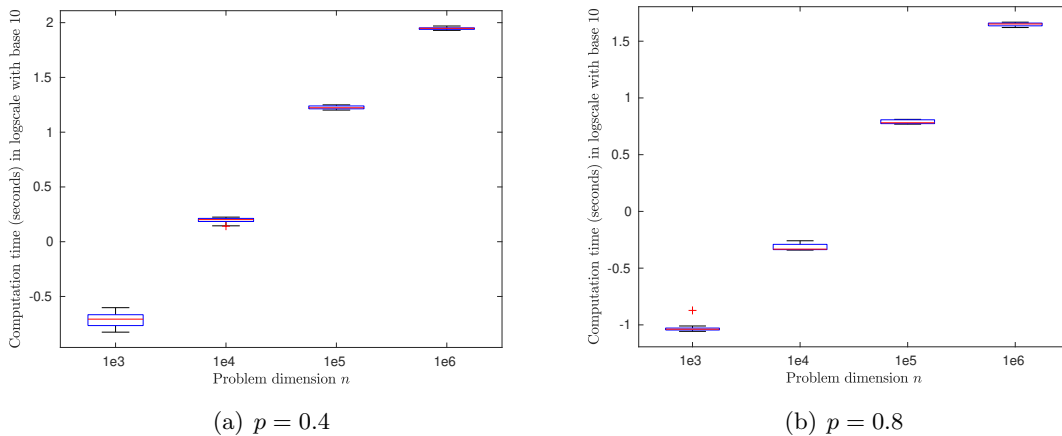


Figure 4: Box plots of the elapsed wall-clock time (seconds) for IRBP with $p \in \{0.4, 0.8\}$ and $n \in \{10^3, 10^4, 10^5, 10^6\}$. Each presented elapsed wall-clock time value is averaged over 20 runs.

Fourth experiment: Sensitivity to parameters M and τ . Following the data generation setting at the beginning of this section, in this experiment, we report the performance measures under different combinations of (M, τ) . We use “Obj” and “Time” to denote the objective function value in (\mathcal{P}) and the elapsed wall-clock time in implementing IRBP, respectively. Table 1 shows the test results, and the results are averaged over 20 randomly generated test problems.

From Table 1, we can see that the IRBP is generally not sensitive to the choice of M and τ . Based on our experience, we suggest that the choice of τ should be appropriately close to 1, and M could be determined by rough tuning depending on the problems at hand.

Table 1: Performance measures of IRBP under different combinations of (M, τ) for $p = 0.5$ and $n = 10^5$.

M	τ	Obj	$\bar{\alpha}$	β	Time(s)
10	1.01	158.145	2.0577e-11	1.0246e-10	17.19
	1.1	158.322	2.0464e-11	1.0240e-10	17.06
	1.5	158.155	2.0488e-11	1.0281e-10	17.08
	1.8	158.012	2.1386e-11	1.0258e-10	16.97
10^2	1.01	158.092	2.0793e-11	1.0298e-10	17.43
	1.1	158.130	2.0630e-11	1.0223e-10	17.70
	1.5	158.152	2.0947e-11	1.0256e-10	17.07
	1.8	157.956	2.0934e-11	1.0245e-10	17.38
10^3	1.01	158.189	2.1201e-11	1.0279e-10	17.21
	1.1	158.179	2.1532e-11	1.0245e-10	17.56
	1.5	158.244	2.0938e-11	1.0298e-10	17.43
	1.8	158.149	2.0642e-11	1.0257e-10	17.19
10^4	1.01	158.140	2.0456e-11	1.0252e-10	17.17
	1.1	158.076	2.1416e-11	1.0253e-10	17.13
	1.5	158.024	2.1136e-11	1.0233e-10	17.21
	1.8	158.245	2.0674e-11	1.0224e-10	17.06

6. Conclusion

In this paper, we have proposed, analyzed, and tested an iteratively reweighted ℓ_1 -ball projection approach for solving the projection of a given vector onto the ℓ_p ball. The central idea of the algorithm is to perform projections onto a sequence of simple and tractable weighted ℓ_1 -balls. First-order necessary optimality conditions are derived. We have established global convergence and analyzed the worst-case complexity of the proposed algorithm. Numerical experiments have demonstrated the effectiveness and efficiency of the algorithm.

We believe that our work could lead to a breakthrough movement in the research of solving ℓ_p -ball constrained optimization problems. With a practical ℓ_p -ball projection method, many intractable ℓ_p -ball constrained optimization problems can now be handled. Future work may involve analyzing the local convergence rate, accelerating the ℓ_p -ball projection algorithm, and exploring the applicability of this problem in many applications.

Acknowledgments

Hao Wang is the corresponding author. We would like to acknowledge the support for this paper from the Young Scientists Fund of the National Natural Science Foundation of China No. 12001367 and the Natural Science Foundation of Shanghai under Grant 21ZR1442800.

Appendix A. Proofs of Technical Propositions and Lemmas

A.1 Proof of (14)

Proof To compute the optimal solution $\hat{\mathbf{x}}$, we first derive the optimality condition of its associated optimization problem in the following

$$(2 + 2\nu)\hat{\mathbf{x}} = 2\nu\bar{\mathbf{y}}, \quad \nu \geq 0,$$

where ν is the defined nonnegative multiplier. It follows that $\hat{\mathbf{x}} = \frac{\nu}{1+\nu}\bar{\mathbf{y}}$. Therefore, it satisfies $\|(\frac{\nu}{1+\nu} - 1)\bar{\mathbf{y}}\|_2^2 \leq \|\mathbf{x}^{\bar{k}} - \bar{\mathbf{y}}\|_2^2$. Simple rearrangement yields that $\nu \geq \frac{\|\bar{\mathbf{y}}\|_2}{\|\mathbf{x}^{\bar{k}} - \bar{\mathbf{y}}\|_2} - 1$. Since $\nu \geq 0$ and $\mathbf{x}^{\bar{k}} \neq \mathbf{0}$, we have $0 < \|\hat{\mathbf{x}}\|_2 \leq \|\bar{\mathbf{y}}\|_2 - \|\mathbf{x}^{\bar{k}} - \bar{\mathbf{y}}\|_2$, completing the proof. \blacksquare

A.2 Proof of Lemma 8

Proof Suppose by contradiction that there exists $\hat{k} \in \mathbb{N}$ such that

$$\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 \|\text{sign}(\mathbf{x}^{k+1} - \mathbf{x}^k) \circ \mathbf{w}^k\|_2^\tau > M, \quad \text{for all } k \geq \hat{k}, \quad (32)$$

Since ϵ is never decreased after iteration \hat{k} , it holds that

$$w_i^k = p(|x_i^k| + \epsilon_i^k)^{p-1} \leq p(\epsilon_i^{\hat{k}})^{p-1}$$

for all $i \in [n]$. Thus, it follows from (32) that

$$\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 > M \|\text{sign}(\mathbf{x}^{k+1} - \mathbf{x}^k) \circ \mathbf{w}^k\|_2^{-\tau} \geq M \left(\sum_{x_i^{k+1} \neq x_i^k} p^2 (\epsilon_i^{\hat{k}})^{2(p-1)} \right)^{-\tau/2} > 0$$

for all $k \geq \hat{k}$, contradicting Lemma 6 (ii). This completes the proof. \blacksquare

A.3 Proof of Proposition 9

Proof Since $0 \notin \Lambda_{\mathcal{U}}$, there exists $\hat{\lambda}$ such that $\lambda^k \geq \hat{\lambda}$ for all $k \in \mathcal{U}$.

- (i) Consider a subsequence $\mathcal{S} \subset \mathcal{U}$. There exists $\delta_i > 0$ such that for sufficiently large $k \in \mathcal{S} \subset \mathcal{U}$, $x_i^k > \delta_i > 0$ for all $i \in \mathcal{I}^k$ (notice that $\mathcal{I}^k \neq \emptyset$ by Lemma 7(i)).

We prove this by contradiction and suppose this is not true. By Lemma 6 and Lemma 8, there exists \bar{k} such that $|x_i^k - x_i^{k-1}| < \frac{1}{4} \left(\frac{1}{\lambda_p} \|\bar{\mathbf{y}}\|_\infty \right)^{1-p}$ and $|\epsilon_i^k - \epsilon_i^{k-1}| < \frac{1}{4} \left(\frac{1}{\lambda_p} \|\bar{\mathbf{y}}\|_\infty \right)^{1-p}$ for all $k > \bar{k}$ and $i \in \mathcal{I}^k$. On the other hand, by assumption, there exists \hat{k} such that $x_i^k < \frac{1}{4} \left(\frac{1}{\lambda_p} \|\bar{\mathbf{y}}\|_\infty \right)^{1-p}$ and $\epsilon_i^k < \frac{1}{4} \left(\frac{1}{\lambda_p} \|\bar{\mathbf{y}}\|_\infty \right)^{1-p}$ for all $k > \hat{k}$ and $i \in \mathcal{I}^k$. Hence, for all $k > \max(\bar{k}, \hat{k})$, this implies

$$\begin{aligned} \bar{y}_i - x_i^k &= \lambda^k p (x_i^{k-1} + \epsilon_i^{k-1})^{p-1} \geq \hat{\lambda} p (x_i^{k-1} + \epsilon_i^{k-1})^{p-1} \\ &\geq \hat{\lambda} p (x_i^k + |x_i^k - x_i^{k-1}| + \epsilon_i^k + |\epsilon_i^k - \epsilon_i^{k-1}|)^{p-1} \\ &> \|\bar{\mathbf{y}}\|_\infty. \end{aligned}$$

This contradicts the assumption that $x_i^k > 0$. Therefore, $x_i^k > \delta_i > 0$ for sufficiently large $k \in \mathcal{U}$ and $i \in \mathcal{I}^k$.

Next, we show that \mathcal{I}^k and \mathcal{A}^k are fixed for sufficiently large k . For sufficiently large $k \in \mathcal{U}$ and $i \in \mathcal{I}^k$, we have $x_i^{k+1} > \delta_i/2 > 0$ since $x_i^{k+1} - x_i^k \rightarrow 0$ by Lemma 6. Hence

$$\mathcal{I}^k \subset \mathcal{I}^{k+1}. \quad (33)$$

It then suffices to show $\mathcal{A}^k \subset \mathcal{A}^{k+1}$. For sufficiently large $k \in \mathcal{U}$ and $i \in \mathcal{A}^k$, we have $\lambda^{k+1}w_i^k = \lambda^{k+1}p(|x_i^k| + \epsilon_i^k)^{p-1} = \lambda^{k+1}p(\epsilon_i^k)^{p-1} \geq \hat{\lambda}p(\epsilon_i^k)^{p-1} \geq \bar{y}_i$, which means $x_i^{k+1} = 0$ by (10a). In other words, we obtain $\mathcal{A}^k \subset \mathcal{A}^{k+1}$ for sufficiently large $k \in \mathcal{U}$. This, together with (33), gives

$$\mathcal{A}(\mathbf{x}^k) = \mathcal{A}(\mathbf{x}^{k+1}) \quad \text{and} \quad \mathcal{I}(\mathbf{x}^k) = \mathcal{I}(\mathbf{x}^{k+1}).$$

Now, we are ready to show the update condition in (12) is also satisfied at the $(k+1)$ th iteration. Since $\mathcal{I}^k \neq \emptyset$ (by Lemma 7 (i)) and $\lim_{\substack{k \rightarrow \infty \\ k \in \mathcal{U}}} \|\mathbf{x}_{\mathcal{I}^k}^{k+1} - \mathbf{x}_{\mathcal{I}^k}^k\|_2 = 0$, we know for sufficiently large k ,

$$\|\mathbf{w}_{\mathcal{I}^k}^k\|_2^2 = p^2 \sum_{i \in \mathcal{I}^k} (x_i^k + \epsilon_i^k)^{2(p-1)} \leq p^2 \sum_{i \in \mathcal{I}^k} \delta^{2(p-1)} = p^2 |\mathcal{I}^k| \delta^{2(p-1)}.$$

Overall, we have for sufficiently large $k \in \mathcal{U}$,

$$\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 \|\text{sign}(\mathbf{x}^{k+1} - \mathbf{x}^k) \circ \mathbf{w}^k\|_2^{\bar{\tau}} \leq \|\mathbf{x}_{\mathcal{I}^k}^{k+1} - \mathbf{x}_{\mathcal{I}^k}^k\|_2 \|\mathbf{w}_{\mathcal{I}^k}^k\|_2^{\bar{\tau}} < M, \quad (34)$$

implying $k+1 \in \mathcal{U}$. By induction, we know the condition (12) is triggered after $k > \bar{k}$.

- (ii) By statement (i), we know that 0 is not a cluster point of $\{\lambda^k\}$. Then there exists $\hat{\lambda}$ such that $\lambda^k \geq \hat{\lambda}$ for all $k \in \mathbb{N}$. Let $C := \|\bar{\mathbf{y}}\|_\infty / \hat{\lambda}$.

Suppose $\{w_i^k\}$ is unbounded, then there exists \bar{k} such that $w_i^{\bar{k}} \geq C$. By (10b), it holds true that $x_i^{\bar{k}+1} = 0$, yielding

$$w_i^{\bar{k}+1} = p(|x_i^{\bar{k}+1}| + \epsilon_i^{\bar{k}+1})^{p-1} = p(\epsilon_i^{\bar{k}+1})^{p-1} \geq p(|x_i^{\bar{k}}| + \epsilon_i^{\bar{k}})^{p-1} = w_i^{\bar{k}},$$

where the inequality holds because $(\cdot)^{p-1}$ is nonincreasing and $\epsilon_i^{\bar{k}+1} \leq |x_i^{\bar{k}}| + \epsilon_i^{\bar{k}}$. Hence, by induction, it follows that $x_i^k \equiv 0$ for any $k > \bar{k}$. On the other hand, suppose $\{w_i^k\}$ is bounded above. By statement (i), $\{(\mathbf{x}^k, \lambda^k)\}_{\mathcal{U}}$ coincides with $\{(\mathbf{x}^k, \lambda^k)\}$ after $k > \bar{k}$, we then know x_i^k is strictly bounded away from 0 for all sufficiently large k . Therefore, there exists $\delta_i > 0$ such that $x_i^k > \delta_i > 0$ for all sufficiently large k .

- (iii) This statement follows straightforwardly from (ii). ■

A.4 Proof of Lemma 11

Proof To derive (18), recalling (10b), we have

$$\begin{aligned}\alpha(\mathbf{x}^k, \lambda^k) &= \sum_{i=1}^n |(\bar{y}_i - x_i^k - \lambda^k p(x_i^k)^{p-1})x_i^k| = \lambda^k p \sum_{i \in \mathcal{I}^k} |(x_i^{k-1} + \epsilon_i^{k-1})^{p-1}(x_i^k) - (x_i^k)^p| \\ &= \lambda^k p \sum_{i \in \mathcal{I}^k} |(x_i^{k-1} + \epsilon_i^{k-1})^{p-1} - (x_i^k)^{p-1}| |x_i^k|.\end{aligned}$$

From the Lagrange's mean value theorem, we have for $i \in \mathcal{I}^k$

$$(x_i^{k-1} + \epsilon_i^{k-1})^{p-1} - (x_i^k)^{p-1} = (p-1)(\tilde{x}_i^k)^{p-2}(x_i^{k-1} + \epsilon_i^{k-1} - x_i^k) \quad (35)$$

with \tilde{x}_i^k between x_i^k and $x_i^{k-1} + \epsilon_i^{k-1}$. It then follows that

$$\begin{aligned}\alpha(\mathbf{x}^k, \lambda^k) &\leq \lambda^k(1-p)p \sum_{i \in \mathcal{I}^k} (\tilde{x}_i^k)^{p-2} x_i^k |x_i^{k-1} + \epsilon_i^{k-1} - x_i^k| \\ &\leq \lambda^k(1-p)p \left[\sum_{i \in \mathcal{I}^k} (\tilde{x}_i^k)^{p-2} x_i^k (|x_i^{k-1} - x_i^k| + \epsilon_i^k) \right] \\ &\leq \lambda^k(1-p)p \left[\|\tilde{\mathbf{x}}_{\mathcal{I}^k}^k\|_2^{p-2} \|\bar{\mathbf{y}}\|_\infty \left[\|\mathbf{x}_{\mathcal{I}^k}^{k-1} - \mathbf{x}_{\mathcal{I}^k}^k\|_2 + \|\boldsymbol{\epsilon}_{\mathcal{I}^k}^k\|_2 \right] \right],\end{aligned}$$

completing the proof for (18).

To derive (19), we have from Lemma 6 (i)

$$\begin{aligned}\beta(\mathbf{x}^k) &= \left| \sum_{i=1}^n (x_i^k)^p - \gamma \right| \\ &\leq \left| \sum_{i=1}^n (x_i^k)^p - \sum_{i=1}^n [(x_i^{k-1} + \epsilon_i^{k-1})^p + w_i^{k-1}(x_i^k - x_i^{k-1})] \right| \\ &\leq \sum_{i \in \mathcal{I}^k} w_i^{k-1} |x_i^{k-1} - x_i^k| + \sum_{i \in \mathcal{I}^k} |(x_i^{k-1} + \epsilon_i^{k-1})^p - (x_i^k)^p| + \sum_{i \in \mathcal{A}^k} (x_i^{k-1} + \epsilon_i^{k-1})^p.\end{aligned}$$

From the Lagrange's mean value theorem, we have for $i \in \mathcal{I}^k$,

$$(x_i^{k-1} + \epsilon_i^{k-1})^p - (x_i^k)^p = p(\hat{x}_i^k)^{p-1}(x_i^{k-1} + \epsilon_i^{k-1} - x_i^k) \quad (36)$$

with \hat{x}_i^k between x_i^k and $x_i^{k-1} + \epsilon_i^{k-1}$. It then follows that

$$\begin{aligned}\beta(\mathbf{x}^k) &\leq \|\mathbf{w}_{\mathcal{I}^k}^{k-1}\|_2 \|\mathbf{x}_{\mathcal{I}^k}^{k-1} - \mathbf{x}_{\mathcal{I}^k}^k\|_2 + p \sum_{i \in \mathcal{I}^k} (\hat{x}_i^k)^{p-1} (|x_i^{k-1} - x_i^k| + |\epsilon_i^{k-1}|) \\ &\quad + \sum_{i \in \mathcal{A}^k} (x_i^{k-1} + \epsilon_i^{k-1})^{p-1} (x_i^{k-1} + \epsilon_i^{k-1}) \\ &\leq \|\mathbf{w}_{\mathcal{I}^k}^{k-1}\|_2 \|\mathbf{x}_{\mathcal{I}^k}^{k-1} - \mathbf{x}_{\mathcal{I}^k}^k\|_2 + p \sum_{i \in \mathcal{I}^k} (\hat{x}_i^k)^{p-1} (|x_i^{k-1} - x_i^k| + |\epsilon_i^{k-1}|) \\ &\quad + \sum_{i \in \mathcal{A}^k} (\epsilon_i^{k-1})^{p-1} (x_i^{k-1} + \epsilon_i^{k-1}) \\ &\leq (\|\mathbf{w}_{\mathcal{I}^k}^{k-1}\|_2 + p \|\hat{\mathbf{x}}_{\mathcal{I}^k}^k\|_2^{p-1}) \|\mathbf{x}_{\mathcal{I}^k}^{k-1} - \mathbf{x}_{\mathcal{I}^k}^k\|_2 + p \|\hat{\mathbf{x}}_{\mathcal{I}^k}^k\|_\infty^{p-1} \|\boldsymbol{\epsilon}_{\mathcal{I}^k}^{k-1}\|_1 \\ &\quad + \|\boldsymbol{\epsilon}_{\mathcal{A}^k}^{k-1}\|_1 + \|\boldsymbol{\epsilon}_{\mathcal{A}^k}^{k-1}\|_1^{p-1} \circ \|\mathbf{x}_{\mathcal{A}^k}^{k-1}\|_1,\end{aligned}$$

completing the proof for (19). ■

A.5 Proof of Theorem 14

Proof Suppose $\{\mathbf{x}^k\}$ has multiple cluster points, then Γ contains more than one element. Lemma 7 (ii) implies $\mathbf{x}^* \neq 0$. Hence $\mathcal{I}^* \neq \emptyset$.

By Proposition 9, for any $(\mathbf{x}^*, \lambda^*) \in \Gamma$ satisfying (7a) implies that $\mathbf{x}_{\mathcal{I}^*}^*$ is a first-order stationary solution to the smoothed ℓ_p regularization problem in the reduced subspace $\mathbb{R}^{|\mathcal{I}^*|}$. Note that the Lagrangian for (\mathcal{P}) in $\mathbb{R}^{|\mathcal{I}^*|}$ is written as

$$\mathcal{L}(\mathbf{x}_{\mathcal{I}^*}, \lambda^*) = \frac{1}{2} \|\mathbf{x}_{\mathcal{I}^*} - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 + \lambda^* (\|\mathbf{x}_{\mathcal{I}^*}\|_p^p - \gamma). \quad (37)$$

Therefore, we have $[\nabla_{\mathbf{x}_{\mathcal{I}^*}} \mathcal{L}(\mathbf{x}_{\mathcal{I}^*}^*, \lambda^*)]_i = x_i^* - \bar{y}_i + \lambda^* p (x_i^*)^{p-1} = 0$ for $i \in \mathcal{I}^*$ and $(\mathbf{x}^*, \lambda^*) \in \Gamma$, meaning every limit point is stationary for $\mathcal{L}(\mathbf{x}_{\mathcal{I}^*}, \lambda^*)$. On the other hand, by Golubitsky and Guillemin (2012, Proposition 6.4), if the Hessian of \mathcal{L} with respect to $\mathbf{x}_{\mathcal{I}^*}$

$$[\nabla^2 \mathcal{L}(\mathbf{x}_{\mathcal{I}^*}, \lambda^*)]_{\mathcal{I}^* \mathcal{I}^*} = \mathbf{I}_{\mathcal{I}^* \mathcal{I}^*} + \text{Diag}(\lambda^* p(p-1)(\mathbf{x}_{\mathcal{I}^*}^*)^{p-2})$$

is nonsingular, or equivalently, $x_i^* \neq [\lambda^* p(1-p)]^{\frac{1}{2-p}}$ for $i \in \mathcal{I}^*$. Then \mathbf{x}^* is an isolated critical point. This contradicts Lemma 13 and the assumption. Hence, \mathbf{x}^* is the unique cluster point of $\{\mathbf{x}^k\}$. ■

A.6 Proof of Theorem 16

Proof By Proposition 9, we can assume the \bar{k} gives us the stable $\mathcal{I}(\mathbf{x}^k) = \mathcal{I}(\mathbf{x}^*)$ and $\mathcal{A}(x^k) = \mathcal{A}(x^*)$ for $k \geq \bar{k}$. Lemma 11 implies $\{\|[\tilde{\mathbf{x}}_{\mathcal{I}^*}^k]^{p-2}\|_2\}$ is bounded above by a constant c_1 .

Let $\bar{\lambda}$ denote an upper bound for $\{\lambda^k\}$ by Lemma 7 (iv) and define $\kappa_1 := (\bar{\lambda} p(1-p)c_1 \|\bar{\mathbf{y}}\|_\infty)$. Then, we rewrite (18) as

$$\begin{aligned} \alpha^2(\mathbf{x}^k, \lambda^k) &\leq (\kappa_1 \|\mathbf{x}_{\mathcal{I}^*}^{k-1} - \mathbf{x}_{\mathcal{I}^*}^k\|_2 + \kappa_1 \|\epsilon_{\mathcal{I}^*}^k\|_2)^2 \leq 2\kappa_1^2 (\|\mathbf{x}_{\mathcal{I}^*}^{k-1} - \mathbf{x}_{\mathcal{I}^*}^k\|_2^2 + \|\epsilon_{\mathcal{I}^*}^k\|_2^2) \\ &\leq 2\kappa_1^2 (\|\mathbf{x}_{\mathcal{I}^*}^{k-1} - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 - \|\mathbf{x}_{\mathcal{I}^*}^k - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 + \|\epsilon_{\mathcal{I}^*}^k\|_2^2), \end{aligned} \quad (38)$$

where the second inequality follows from the Cauchy-Schwarz inequality and the last inequality holds due to (13).

Replacing k by t and summing up both sides of (38) from \bar{k} to $\bar{k} + k$ yields

$$\begin{aligned} k \min_{\bar{k} \leq t \leq \bar{k}+k} \alpha^2(\mathbf{x}^t, \lambda^t) &\leq \sum_{t=\bar{k}}^{\bar{k}+k} \alpha^2(\mathbf{x}^t, \lambda^t) \leq \sum_{t=\bar{k}}^{\bar{k}+k} 2\kappa_1^2 (\|\mathbf{x}_{\mathcal{I}^*}^{t-1} - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 - \|\mathbf{x}_{\mathcal{I}^*}^t - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 + \|\epsilon_{\mathcal{I}^*}^t\|_2^2) \\ &\leq 2\kappa_1^2 (\|\mathbf{x}_{\mathcal{I}^*}^{\bar{k}-1} - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 - \|\mathbf{x}_{\mathcal{I}^*}^{\bar{k}+k} - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 + \sum_{t=\bar{k}}^{\bar{k}+k} \|\epsilon_{\mathcal{I}^*}^t\|_2^2) \\ &\leq 2\kappa_1^2 (\|\mathbf{x}_{\mathcal{I}^*}^{\bar{k}-1} - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 - \|\mathbf{x}_{\mathcal{I}^*}^* - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 + \frac{\theta^2 - \theta^{2(k+1)}}{1 - \theta^2} \|\epsilon_{\mathcal{I}^*}^{\bar{k}}\|_2^2) \\ &\leq C_\alpha := 2\kappa_1^2 (\|\mathbf{x}_{\mathcal{I}^*}^{\bar{k}-1} - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 - \|\mathbf{x}_{\mathcal{I}^*}^* - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 + \frac{\theta^2}{1 - \theta^2} \|\epsilon_{\mathcal{I}^*}^{\bar{k}}\|_2^2), \end{aligned}$$

where the second inequality follows from (13) and the last inequality holds because the update of ϵ in (12) is always triggered. Dividing both sides by k yields $\min_{\bar{k} \leq t \leq \bar{k}+k} \alpha(\mathbf{x}^t, \lambda^t) \leq \sqrt{C_\alpha/k}$.

Similar argument applied to (19) yields

$$\beta^2(\mathbf{x}) \leq \kappa_2(\|\mathbf{x}_{\mathcal{I}^*}^{k-1} - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 - \|\mathbf{x}_{\mathcal{I}^*}^k - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 + \|\epsilon_{\mathcal{I}^*}^{k-1}\|_1^2 + \|[\epsilon_{\mathcal{I}^*}^{k-1}]^p\|_1^2)$$

with $\kappa_2 := (\bar{w} + pc_2)^2 + p^2c_2^2 + 1$. Replacing k by t and summing both sides from \bar{k} to $\bar{k} + k$ yields that

$$\begin{aligned} k \min_{\bar{k} \leq t \leq \bar{k}+k} \beta^2(\mathbf{x}^t) &\leq \sum_{t=\bar{k}}^{\bar{k}+k} \beta^2(\mathbf{x}^t) \\ &\leq \kappa_2(\|\mathbf{x}_{\mathcal{I}^*}^{\bar{k}-1} - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 - \|\mathbf{x}_{\mathcal{I}^*}^* - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 + \frac{\theta^2 - \theta^{2(k+1)}}{1-\theta^2} \|\epsilon_{\mathcal{I}^*}^{\bar{k}}\|_1^2 + \frac{\theta^{2p} - (\theta^{2p})^{k+1}}{1-\theta^{2p}} \|[\epsilon_{\mathcal{A}^*}^{\bar{k}+1}]^p\|_1^2) \\ &\leq C_\beta := \kappa_2(\|\mathbf{x}_{\mathcal{I}^*}^{\bar{k}-1} - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 - \|\mathbf{x}_{\mathcal{I}^*}^* - \bar{\mathbf{y}}_{\mathcal{I}^*}\|_2^2 + \frac{\theta^2}{1-\theta^2} \|\epsilon_{\mathcal{I}^*}^{\bar{k}}\|_1^2 + \frac{\theta^{2p}}{1-\theta^{2p}} \|[\epsilon_{\mathcal{A}^*}^{\bar{k}+1}]^p\|_1^2). \end{aligned}$$

Dividing both sides by k yields $\min_{\bar{k} \leq t \leq \bar{k}+k} \beta(\mathbf{x}^t) \leq \sqrt{C_\beta/k}$. \blacksquare

A.7 Proof of Lemma 18

Proof We consider two cases. If $0 \leq x_i \leq \epsilon_i$ for $i \in [n]$, it holds that

$$(x_i + \epsilon_i)^p - x_i^p < (x_i + \epsilon_i)^p \leq (2\epsilon_i)^p = 2^p \epsilon_i^p, \quad (39)$$

and

$$\begin{aligned} &|(\bar{y}_i - x_i)x_i - \lambda p x_i^p| - |(\bar{y}_i - x_i)x_i - \lambda p(x_i + \epsilon_i)^{p-1}x_i| \\ &\leq \lambda p x_i |(x_i + \epsilon_i)^{p-1} - x_i^{p-1}| \\ &\leq \lambda p x_i \epsilon_i^{p-1} \leq \lambda p \epsilon_i^p. \end{aligned} \quad (40)$$

On the other hand, suppose $\epsilon_i < x_i \leq \bar{y}_i$, there then exists $\hat{x}_i \in [x_i, x_i + \epsilon_i]$ such that

$$(x_i + \epsilon_i)^p - x_i^p = p \hat{x}_i^{p-1} \epsilon_i \leq p x_i^{p-1} \epsilon_i < p \epsilon_i^{p-1} \epsilon_i = p \epsilon_i^p < 2^p \epsilon_i^p, \quad (41)$$

implying

$$\begin{aligned} &|(\bar{y}_i - x_i)x_i - \lambda p(x_i)^p| - |(\bar{y}_i - x_i)x_i - \lambda p(x_i + \epsilon_i)^{p-1}x_i| \\ &\leq \lambda p x_i |(x_i + \epsilon_i)^{p-1} - x_i^{p-1}| \\ &\leq \lambda p(1-p)x_i \hat{x}_i^{p-2} \epsilon_i \\ &\leq \lambda p(1-p)x_i^{p-1} \epsilon_i \leq \lambda p \epsilon_i^p. \end{aligned} \quad (42)$$

Combining these two cases and summing up both sides of (39) and (41), we have $|\beta(\mathbf{x}) - \beta_\epsilon(\mathbf{x})| < 2^p \|\epsilon\|_p^p$. Similarly, summing up both sides of (40) and (42), we have $|\alpha(\mathbf{x}, \lambda) - \alpha_\epsilon(\mathbf{x})| \leq \lambda p \sum_{i=1}^n \epsilon_i^p = \lambda p \|\epsilon\|_p^p$. This completes the proof. \blacksquare

A.8 Proof of Lemma 19

Proof From the optimality conditions (11), we have

$$\begin{aligned}
 \lambda^k &= \frac{\sum_{i \in \mathcal{I}^k} (\bar{y}_i - x_i^k)}{p \sum_{i \in \mathcal{I}^k} (x_i^{k-1} + \epsilon_i)^{p-1}} \leq \frac{\sum_{i \in \mathcal{I}^k} \bar{y}_i}{p \sum_{i \in \mathcal{I}^k} (\bar{y}_i + \epsilon_i)^{p-1}} \leq \frac{\|\bar{\mathbf{y}}\|_1}{p \min_{i \in \mathcal{I}^k} (\bar{y}_i + \epsilon_i)^{p-1}} \\
 &\leq \frac{\|\bar{\mathbf{y}}\|_1}{p \min_{i \in [n]} (\bar{y}_i + \epsilon_i)^{p-1}} \leq \frac{\|\bar{\mathbf{y}}\|_1}{p (\max_{i \in [n]} (\bar{y}_i + \epsilon_i))^{p-1}} \leq \frac{\|\bar{\mathbf{y}}\|_1}{p (\|\bar{\mathbf{y}}\|_\infty + \|\boldsymbol{\epsilon}\|_\infty)^{p-1}},
 \end{aligned} \tag{43}$$

where the first inequality to the fourth inequality hold due to Lemma 7 (i) and $(\cdot)^{p-1}$ is a monotonically decreasing function on \mathbb{R}_{++} .

Now we prove (ii). Since for any $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}_+^n$, it is true that

$$\sum_{i=1}^n a_i b_i c_i \leq \|\mathbf{a}\|_2 \|\mathbf{b}\|_2 \|\mathbf{c}\|_2. \tag{44}$$

It follows that

$$\begin{aligned}
 &\alpha_\epsilon(\mathbf{x}^k, \lambda^k) \\
 &= \sum_{i=1}^n |(\bar{y}_i - x_i^k) x_i^k - \lambda^k p (x_i^k + \epsilon_i)^{p-1} x_i^k| \\
 &= \sum_{i=1}^n |(\bar{y}_i - x_i^k) x_i^k - \lambda^k p (x_i^{k-1} + \epsilon_i)^{p-1} x_i^k + \lambda^k p x_i^k [(x_i^{k-1} + \epsilon_i)^{p-1} - (x_i^k + \epsilon_i)^{p-1}]| \\
 &= \lambda^k p \sum_{i=1}^n x_i^k |(x_i^{k-1} + \epsilon_i)^{p-1} - (x_i^k + \epsilon_i)^{p-1}| \\
 &\leq \lambda^k p (1-p) \sum_{i=1}^n x_i^k (\tilde{x}_i^{k-1} + \epsilon_i)^{p-2} |x_i^{k-1} - x_i^k| \\
 &\leq \lambda^k p (1-p) \sum_{i=1}^n \bar{y}_i \epsilon_i^{p-2} |x_i^{k-1} - x_i^k| \\
 &\leq \frac{\|\bar{\mathbf{y}}\|_1 \|\bar{\mathbf{y}}\|_2}{(\|\bar{\mathbf{y}}\|_\infty + \|\boldsymbol{\epsilon}\|_\infty)^{p-1}} (1-p) \|\boldsymbol{\epsilon}^{p-2}\|_2 \|\mathbf{x}^{k-1} - \mathbf{x}^k\|_2,
 \end{aligned}$$

where the third equality is by the optimality conditions (10), the first inequality is by Lagrange's mean value theorem with \tilde{x}_i^{k-1} between x_i^k and x_i^{k-1} , the second inequality is by Lemma 7 (i) and the last inequality is by (43) and (44).

As for (iii), we have from Lemma 7 (iii) that

$$\begin{aligned}
 \|\mathbf{x}^k + \boldsymbol{\epsilon}\|_p^p - \gamma &= \left| \sum_{i=1}^n w_i^k (x_i^{k+1} - x_i^k) \right| \leq \sum_{i=1}^n |w_i^k (x_i^{k+1} - x_i^k)| \\
 &\leq p \sum_{i=1}^n |\epsilon_i^{p-1} (x_i^{k+1} - x_i^k)| \leq p \|\boldsymbol{\epsilon}^{p-1}\|_2 \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2,
 \end{aligned}$$

where the second inequality is true because $(\cdot)^{p-1}$ is nonincreasing and the last inequality is by (44). This completes the proof. \blacksquare

A.9 Proof of Lemma 21

Proof By Lemma 19, at the k th iteration (note that this is the last iteration running in the algorithm), it follows that

$$\begin{aligned} \left(\frac{\varepsilon}{2}\right)^2 &\leq \alpha_\epsilon^2(\mathbf{x}^k, \lambda^k) = \kappa_3^2 \|\bar{\mathbf{y}}\|_2^2 (1-p)^2 \|\epsilon^{p-2}\|_2^2 \|\mathbf{x}^{k-1} - \mathbf{x}^k\|_2^2 \\ &\leq \kappa_3^2 \|\bar{\mathbf{y}}\|_2^2 (1-p)^2 \|\epsilon^{p-2}\|_2^2 (\|\mathbf{x}^{k-1} - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^k - \bar{\mathbf{y}}\|_2^2), \end{aligned} \quad (45)$$

where the second inequality holds owing to (13). Similarly,

$$\begin{aligned} \left(\frac{\varepsilon}{2}\right)^2 &\leq \beta_\epsilon^2(\mathbf{x}^k) = p^2 \|\epsilon^{p-1}\|_2^2 \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2 \\ &\leq p^2 \|\epsilon^{p-1}\|_2^2 (\|\mathbf{x}^k - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^{k+1} - \bar{\mathbf{y}}\|_2^2). \end{aligned} \quad (46)$$

Reindexing each term by t and summing up both sides of (45) from $t = 1$ to k , we have

$$\begin{aligned} \left(\frac{\varepsilon}{2}\right)^2 k &\leq \sum_{t=1}^k \alpha_\epsilon^2(\mathbf{x}^t, \lambda^t) \leq \kappa_3^2 \|\bar{\mathbf{y}}\|_2^2 (1-p)^2 \|\epsilon^{p-2}\|_2^2 \left(\sum_{t=1}^k \|\mathbf{x}^{t-1} - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^k - \bar{\mathbf{y}}\|_2^2 \right) \\ &= \kappa_3^2 \|\bar{\mathbf{y}}\|_2^2 (1-p)^2 \|\epsilon^{p-2}\|_2^2 (\|\mathbf{x}^0 - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^k - \bar{\mathbf{y}}\|_2^2) \\ &\leq \kappa_3^2 \|\bar{\mathbf{y}}\|_2^2 (1-p)^2 \|\epsilon^{p-2}\|_2^2 (\|\mathbf{x}^0 - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^* - \bar{\mathbf{y}}\|_2^2). \end{aligned}$$

We then conclude

$$k_\alpha \leq 4(1-p)^2 \kappa_3^2 \|\bar{\mathbf{y}}\|_2^2 \|\epsilon^{p-2}\|_2^2 (\|\mathbf{x}^0 - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^* - \bar{\mathbf{y}}\|_2^2) / \varepsilon^2. \quad (47)$$

Using similar proof techniques for $\beta(\mathbf{x}^k)$, we have

$$k_\beta \leq 4p^2 \|\epsilon^{p-1}\|_2^2 (\|\mathbf{x}^0 - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x}^* - \bar{\mathbf{y}}\|_2^2) / \varepsilon^2. \quad (48)$$

Combining (47) and (48), we therefore have

$$k \leq \max(k_\alpha, k_\beta).$$

This completes the proof. \blacksquare

Appendix B. Root-searching procedure description

The implementation of RS follows the details presented in Chen et al. (2019). In particular, for solving these nonlinear equations, we directly call the built-in *optimize.newton* routine in the open-source software SciPy Virtanen et al. (2020) and use its default values for all parameters (e.g., the maximum number of iterations). We terminate the bisection method in

the exterior loop whenever the maximum number of iterations ($\text{IterMax} = 10^3$) is exceeded or the interval becomes too narrow ($\lambda_{\text{high}} - \lambda_{\text{low}} < 10^{-10}$). In particular, we set initial $\lambda_{\text{low}} = 10^{-15}$ and $\lambda_{\text{high}} = \frac{\|\bar{\mathbf{y}}\|_{\infty}^{2-p}}{p(1-p)(\frac{1}{1-p}+1)^{2-p}}$ suggested by Chen et al. (2019). Moreover, an approximate solution \mathbf{x} is returned if $\frac{1}{n} \|\mathbf{x}\|_p^p - \gamma < 10^{-8}$ holds.

Algorithm 2 Projection onto the ℓ_p ball using RS method

- 1: **Input:** \mathbf{y} , p , γ , tol and IterMax. Let $\bar{\mathbf{y}} = |\mathbf{y}|$.
- 2: **Initialization:** Choose $\lambda_{\text{low}} = 10^{-15}$ and $\lambda_{\text{high}} = \frac{(\bar{y}_i^{\max})^{2-p}}{p(1-p)(\frac{1}{1-p}+1)^{2-p}}$. Set $k = 0$.
- 3: **while** $\lambda_{\text{high}} - \lambda_{\text{low}} \geq \text{tol}$ or $k \leq \text{IterMax}$ **do**
- 4: $\lambda^k = (\lambda_{\text{low}} + \lambda_{\text{high}})/2$.
- 5: $\mathbf{x} = \mathbf{0}_n$.
- 6: (Solve n nonlinear equations) Use Newton's method to solve

$$x_i^k - \bar{y}_i + p\lambda^k(x_i^k)^{p-1} = 0, \quad \forall i \in [n] \quad (49)$$

with the initial point $x_{\text{ini}} \sim U[\frac{2-2p}{2-p}\bar{y}_i, \bar{y}_i]$ is used in Newton's method.

- 7: **if** Newton's method fails to output a solution **then**
 - 8: $x_i^k = 0$
 - 9: **end if**
 - 10: Obtain $\mathbf{x}^k = [x_1^k, x_2^k, \dots, x_n^k]^\top$.
 - 11: **if** $\|\mathbf{x}^k\|_p^p - \gamma \leq \text{tol}$ **then**
 - 12: $\lambda^* = \lambda^k$.
 - 13: **else if** $\|\mathbf{x}^k\|_p^p > \gamma$ **then**
 - 14: $\lambda_{\text{low}} = \lambda^k$.
 - 15: **else if** $\|\mathbf{x}^k\|_p^p < \gamma$ **then**
 - 16: $\lambda_{\text{high}} = \lambda^k$.
 - 17: **end if**
 - 18: $k = k + 1$.
 - 19: **end while**
-

References

- Sohail Bahmani and Bhiksha Raj. A unifying analysis of projected gradient descent for ℓ_p -constrained least squares. *Applied and Computational Harmonic Analysis*, 34(3):366–378, 2013.
- Emilio Rafael Balda, Arash Behboodi, and Rudolf Mathar. Adversarial examples in deep neural networks: An overview. *Deep Learning: Algorithms and Applications*, pages 31–65, 2020.
- Heinz H Bauschke, Minh N Dao, and Walaa M Moursi. On Fejér monotone sequences and nonexpansive mappings. *arXiv preprint arXiv:1507.05585*, 2015.

- Jérôme Bolte and Edouard Pauwels. Majorization-minimization procedures and convergence of sqp methods for semi-algebraic and tame programs. *Mathematics of Operations Research*, 41(2):442–465, 2016.
- Digvijay Boob, Qi Deng, Guanghui Lan, and Yilin Wang. A feasible level proximal point method for nonconvex sparse constrained optimization. *arXiv preprint arXiv:2010.12169*, 2020.
- James V Burke, Frank E Curtis, Hao Wang, and Jiashan Wang. Iterative reweighted linear least squares for exact penalty subproblems on product sets. *SIAM Journal on Optimization*, 25(1):261–294, 2015.
- A Enis Cetin and Mohammad Tofghi. Projection-based wavelet denoising [lecture notes]. *IEEE Signal Processing Magazine*, 32(5):120–124, 2015.
- S Grace Chang, Bin Yu, and Martin Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9):1532–1546, 2000.
- Rick Chartrand and Wotao Yin. Iteratively reweighted algorithms for compressive sensing. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3869–3872. IEEE, 2008.
- Lin Chen, Xue Jiang, Xingzhao Liu, Thiagalingam Kirubarajan, and Zhixin Zhou. Outlier-robust moving object and background decomposition via structured ℓ_p -regularized low-rank representation. *IEEE Transactions on Emerging Topics in Computational Intelligence*, pages 1–19, 2019.
- Laurent Condat. Fast projection onto the simplex and the ℓ_1 ball. *Mathematical Programming*, 158(1-2):575–585, 2016.
- Mithun Das Gupta and Sanjeev Kumar. Non-convex p-norm projection for robust sparsity. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1593–1600, 2013.
- John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the ℓ_1 -ball for learning in high dimensions. In *Proceedings of the 25th International Conference on Machine Learning*, pages 272–279. ACM, 2008.
- Mário AT Figueiredo, José M Bioucas-Dias, and Robert D Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 16(12):2980–2991, 2007.
- Martin Golubitsky and Victor Guillemin. *Stable Mappings and Their Singularities*, volume 14. Springer Science & Business Media, 2012.
- Thomas Grandits and Thomas Pock. Optimizing wavelet bases for sparser representations. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 249–262. Springer, 2017.

- Prateek Jain, Purushottam Kar, et al. Non-convex optimization for machine learning. *Foundations and Trends® in Machine Learning*, 10(3-4):142–363, 2017.
- Kenneth Lange. *MM Optimization Algorithms*. SIAM, 2016.
- Hoai An Le Thi and Tao Pham Dinh. DC programming and DCA: Thirty years of developments. *Mathematical Programming*, 169(1):5–68, 2018.
- Zhizheng Liang, Shixiong Xia, Yong Zhou, Lei Zhang, and Youfu Li. Feature extraction based on ℓ_p -norm generalized principal component analysis. *Pattern Recognition Letters*, 34(9):1037–1045, 2013.
- Christian Michelot. A finite algorithm for finding the projection of a point onto the canonical simplex of \mathbb{R}^n . *Journal of Optimization Theory and Applications*, 50(1):195–200, 1986.
- Alexander N Ostrowski. *Solutions of Equations in Euclidean and Banach Spaces*. Academic Press, 1973.
- Guillaume Perez, Sebastian Ament, Carla Gomes, and Michel Barlaud. Efficient projection algorithms onto the weighted ℓ_1 ball. *arXiv preprint arXiv:2009.02980*, 2020.
- R Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*, volume 317. Springer Science & Business Media, 2009.
- Jiawei Su, Danilo Vasconcellos Vargas, and Kouichi Sakurai. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5):828–841, 2019.
- Markus Thom, Matthias Rapp, and Günther Palm. Efficient dictionary learning with sparseness-enforcing projections. *International Journal of Computer Vision*, 114(2), 2015.
- Paolo Toth and Silvano Martello. *Knapsack problems: Algorithms and computer implementations*. Wiley, 1990.
- Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020.
- Hao Wang, Yining Gao, Jiashan Wang, and Hongying Liu. Constrained optimization involving nonconvex ℓ_p norms: Optimality conditions, algorithm and convergence. *arXiv preprint arXiv:2110.14127*, 2021.
- Dominik Wojtczak. On strong NP-completeness of rational problems. In *International Computer Science Symposium in Russia*, pages 308–320. Springer, 2018.
- Xiaoyong Yuan, Pan He, Qile Zhu, and Xiaolin Li. Adversarial examples: Attacks and defenses for deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9):2805–2824, 2019.

Eduardo H Zarantonello. Projections on convex sets in Hilbert space and spectral theory: Part i. Projections on Convex Sets: Part ii. Spectral Theory. In *Contributions to Nonlinear Functional Analysis*, pages 237–424. Elsevier, 1971.

Fan Zhang, Hao Wang, Jiashan Wang, and Kai Yang. Inexact primal–dual gradient projection methods for nonlinear optimization on convex set. *Optimization*, 69(10):2339–2365, 2020.