Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 20, No. 1, pp. 163–184, 2024

CONSTRAINED OPTIMIZATION INVOLVING NONCONVEX ℓ_P NORMS: OPTIMALITY CONDITIONS, ALGORITHM AND CONVERGENCE

Hao Wang, Yining Gao, Jiashan Wang and Hongying Liu

Abstract: This paper investigates the optimality conditions for characterizing the local minimizers of the constrained optimization problems involving an ℓ_p norm ($0 < p < 1$), which may appear in either the objective or the constraint. This kind of problem has strong applicability to a wide range of areas since the ℓ_p norm can promote sparse solutions. However, the nonsmooth and non-Lipschitz nature of the ℓ_p norm makes these problems difficult to analyze and solve. We derived the first-order necessary conditions and the sequential optimality conditions under various constraint qualifications. We extend the iteratively reweighted algorithms for solving the unconstrained ℓ_p norm regularized problems to constrained cases and show the sequential optimality conditions are satisfied by this algorithm. Global convergence is derived and the performance of the proposed algorithm is exhibited by numerical experiments.

Key words: *non-smooth optimization, sparse optimization problem, ℓ_p norm regularization, optimality condition, iteratively reweighted algorithm*

Mathematics Subject Classification: *90C26, 90C46*

1 Introduction

Sparse regularization problems have attracted considerable attentions over the past decades, which have numerous applications in the areas including compressed sensing [4], biomedical engineering [26, 12], sensor selection [1] and portfolio management [6]. This is because sparse solutions usually lead to better generalization performance and robustness. A common approach of promoting sparsity in the solution is to involve a sparse-inducing term such as the ℓ_1 norm or the ℓ_0 norm of the variables either in the objective as a penalty or in the constraint. In recent years, nonconvex and/or non-Lipschitz sparsity inducing terms such as the ℓ_p (quasi-)norm ($0 < p < 1$) are shown [16] to have preferable performance in many situations. In the past decade, many works focus on designing and analyzing the algorithms for solving the unconstrained ℓ_p regularized problems [27, 9, 13, 21, 24, 10]. However, when it comes to the constrained cases, there are few works despite of its wider applicability. We list two examples of the constrained optimization problems involving an ℓ_p norm.

Example 1. Consider the cloud radio access network (Cloud-RAN) power consumption problem [18, 25], which solves a group sparse problem to induce the group sparsity for the beamformers to guide the remote radio head (RRH) selection. This group sparse problem is addressed by minimizing the mixed ℓ_p/ℓ_2-norm with $p \in (0, 1]$, yielding the following problem

$$\begin{aligned}
\min_{v} \quad & \sum_{l=1}^{L} \sqrt{\frac{\rho_l}{z_l}} \|\tilde{v}_l\|_2^p \\
\text{s.t.} \quad & \sqrt{\sum_{i \neq k} \|h_k^{\mathsf{H}} v_i\|_2^2 + \sigma_k^2} \leq \frac{1}{\gamma_k} \Re(h_k^{\mathsf{H}} v_k) \\
& \|\tilde{v}_l\|_2 \leq \sqrt{P_l}, \; l = 1, \cdots, L, k = 1, \ldots, K.
\end{aligned}$$

Here the Cloud-RAN architecture of this model has L RRHs and K single-antenna Mobile Users (MUs), where the l-th RRH is equipped with N_l antennas. $v_{lk} \in \mathbb{C}^{N_l}$ is the transmit beamforming vector from the l-th RRH to the k-th user with the group structure of transmit vectors $\tilde{v}_l = [v_{l1}^T, \cdots, v_{lK}^T]^T \in \mathbb{C}^{KN_l \times 1}$. Denote the relative fronthaul link power consumption by ρ_l, and the inefficient of drain efficiency of the radio frequency power amplifier by z_l. The channel propagation between user k and RRH l is denoted as $\mathbf{h}_{lk} \in \mathbb{C}^{N_l}$. P_l is the maximum transmit power of the l-th RRH. σ_k is the noise at MU k. $\gamma = (\gamma_1, ..., \gamma_K)^T$ is the target signal-to-interference-plus-noise ratio (SINR).

Example 2. (The ℓ_p-constrained sparse coding) In the context of sparse coding [20], the task is to reconstruct the unknown sparse code word $\bar{x} \in \mathbb{R}^n$ from the linear measurements $y = A\bar{x} + \epsilon$, where $y \in \mathbb{R}^m$ represents the data with m features, $\epsilon \in \mathbb{R}^m$ denotes the noise vector, and A corresponds to the fixed dictionary that consists of n atoms with respect to its columns. This problem can be formulated as

$$\min_{x \in \mathbb{R}^n} \; \|Ax - y\|_2^2 \qquad \text{s.t. } \|x\|_p^p \leq \theta, \tag{1.1}$$

where the ℓ_p ball constraint is to induce sparsity in the code word.

In this paper, we consider the following two general forms of constrained nonlinear optimization with ℓ_p norms. The first one is the constrained ℓ_p regularized problem, meaning the ℓ_p norm appears in the objective as a penalty,

$$\min \; F(x) := f_0(x) + \lambda \|x\|_p^p \quad \text{s.t.} \;\; f_j(x) \leq 0, \; \forall j \in \mathcal{I}; \; f_j(x) = 0, \; \forall j \in \mathcal{E}. \tag{$\mathcal{P}_1$}$$

The second one has the ℓ_p norm in the constraint and requires it to be smaller than a prescribed value $\theta > 0$,

$$\min \; f_0(x) \quad \text{s.t.} \;\; \|x\|_p^p \leq \theta; \; f_j(x) \leq 0, \; \forall j \in \mathcal{I}; \; f_j(x) = 0, \; \forall j \in \mathcal{E}. \tag{$\mathcal{P}_2$}$$

Here, $f_j : \mathbb{R}^n \to \mathbb{R}, j \in \{0\} \cup \mathcal{I} \cup \mathcal{E}$ are continuously differentiable on $\mathbb{R}^n$ and $\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$ with $p \in (0, 1]$. The positive λ is the given regularization parameter and θ is referred to as the radius of ℓ_p-ball.

Despite of the advantages of nonconvex ℓ_p norm in promoting sparse solutions, problems of the forms ($\mathcal{P}_1$) and ($\mathcal{P}_2$) are generally not easy to handle. This is largely due to the nonconvex and non-Lipschitz nature of the ℓ_p norm which makes it difficult to characterize the optimal solutions. In particular, verifiable optimality conditions are often difficult to derive, leaving it an obstacle for designing efficient numerical algorithms. For example, for ($\mathcal{P}_1$), many researchers [7, 15, 25, 24] tend to approximate the ℓ_p term by Lipschitz continuous functions and then solve for an approximate solution. As for ($\mathcal{P}_2$), to the best of our knowledge, not much has been done except the special case that only the ℓ_p ball constraint presents in the problem [28], meaning the projection onto the ℓ_p ball.

1.1 Literature review

The optimality conditions of the unconstrained and inequality constrained versions of ($\mathcal{P}_1$) were studied in [21], which is the immediate predecessor of our work. They derived the first-order and second-order necessary conditions by assuming the "extended" linear independence constraint qualification (ELICQ) is satisfied by ($\mathcal{P}_1$), meaning the LICQ is satisfied at the local minimizer in the subspace consisting of the nonzero variables. They also stated that the second-order optimality conditions can be derived by considering the reduced problems after fixing the zero components at a stationary point. In [3], Bian and Chen derived a first-order necessary optimality condition using the theory of the generalized directional derivative, which is also closely related to our work. In particular, for the case that the constraints are all linear, Gabriel Haeser et al. [10] articulated first- and second-order necessary optimality conditions for this problem based on the perturbed problem and the limits of perturbation. Sufficient conditions for the ϵ-perturbed stationary points are also presented. As for ($\mathcal{P}_2$), [28] derived optimality conditions for the special case where only ℓ_p ball constraint exists using the concept of generalized Fréchet normal cone. To the best of our knowledge, there has been no study on the optimality conditions for more general cases of this problem.

1.2 Contributions

In this paper, we are interested in deriving the optimality conditions to characterize the local solutions of ($\mathcal{P}_1$) and ($\mathcal{P}_2$) under different constraint qualifications (CQ). First of all, we analyze the basic properties of the ℓ_p norm and the ℓ_p norm ball. We derive the regular and general subgradients of the ℓ_p norm and the regular and general normal of the ℓ_p norm ball, which indicate the ℓ_p norm is subdifferentially regular and the ℓ_p ball is Clarke regular. For ($\mathcal{P}_1$) and ($\mathcal{P}_2$), we derive the Karush-Kuhn-Tucker (KKT) conditions and discuss the constraint qualifications that ensure that the KKT conditions are satisfied at a local minimizer. For ($\mathcal{P}_2$), we believe this is the first result.

Recently, Andreani et al. [2] introduced the sequential optimality conditions, namely, the approximate KKT (AKKT) conditions for constrained smooth optimization problems, which is commonly satisfied by many algorithms. They also proposed the Cone-Continuity Property (CCP), under which the AKKT conditions implies the KKT conditions. This is widely believed to be one of the weakest qualification under which KKT conditions hold at local minimizer. We also define the sequential optimality conditions for ($\mathcal{P}_1$) and ($\mathcal{P}_2$) and explore the constraint qualification under which the sequential conditions imply KKT conditions. We believe these are much stronger results than existing ones.

To demonstrate the applicability of the proposed sequential optimality conditions, we extend the well-known iteratively reweighted algorithms for solving unconstrained ℓ_p-regularized problem to general constrained cases and show that those conditions are satisfied at the limit points of the generated iterates. Therefore, under the proposed constraint qualification, the limit points satisfy the KKT conditions.

1.3 Notation and preliminary

We use 0 as the vector filled with all zeros of appropriate size. For $\mathcal{N} \subset \{1, \dots, n\}$ and $x \in \mathbb{R}^n$, let $\mathbb{R}^{|\mathcal{N}|}$ be the reduced subspace of $\mathbb{R}^n$ that consists of the components $x_i, i \in \mathcal{N}$, and denote $x_{\mathcal{N}} \in \mathbb{R}^{|\mathcal{N}|}$ as the subvector of x containing the elements $x_i, i \in \mathcal{N}$. Let $\mathcal{Z} = \{1, \dots, n\} \setminus \mathcal{N}$.

For a differentiable $f : \mathbb{R}^n \to \mathbb{R}$, let $\nabla_{\mathcal{N}} f(x)$ be the vector consisting of $\nabla_i f(x), i \in \mathcal{N}$.

In $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$, define the set

$$\Gamma = \{x \mid f_j(x) \le 0,\ \forall j \in \mathcal{I}; \quad f_j(x) = 0,\ \forall j \in \mathcal{E}\},$$

and the index set of active inequalities by $\mathcal{A}(x) = \{j \mid f_j(x) = 0,\ j \in \mathcal{I}\}$. The sets of zeros and nonzeros in $x \in \mathbb{R}^n$ are defined as

$$\mathcal{Z}(x) = \{i \mid x_i = 0\} \ \text{ and } \ \mathcal{N}(x) = \{i \mid x_i \neq 0\}.$$

For simplicity, we use shorthands $\bar{\mathcal{N}} = \mathcal{N}(\bar{x})$, $\bar{\mathcal{Z}} = \mathcal{Z}(\bar{x})$ and $\mathcal{N}^k = \mathcal{N}(x^k)$, $\mathcal{Z}^k = \mathcal{Z}(x^k)$.

For $C \subset \mathbb{R}^n$, its horizon cone is defined by

$$C^\infty = \begin{cases} \{x \mid \exists x^\nu \in C, \lambda^\nu \searrow 0, \text{ with } \lambda^\nu x^\nu \to x\} & \text{when } C \neq \emptyset, \\ \{0\} & \text{when } C = \emptyset. \end{cases}$$

Another operation on C is the smallest cone containing C, namely the positive hull of C, which is defined as pos $C = \{0\} \cup \{\lambda x \mid x \in C, \lambda > 0\}$. A vector $w \in \mathbb{R}^n$ is tangent to a set $C \subset \mathbb{R}^n$ at $\bar{x} \in C$, written $w \in T_C(\bar{x})$, if $(x^\nu - \bar{x})/\tau^\nu \to w$ for some $x^\nu \underset{C}{\to} \bar{x}, \tau^\nu \searrow 0$. The interior and the boundary of a set $C \subset \mathbb{R}^n$ is denoted as *int* C and ∂C, respectively.

Definition 1.1. [17, Definition 6.3, 6.4]

(a) Let $C \subset \mathbb{R}^n$ and $\bar{x} \in C$. A vector v is a regular normal to C at $\bar{x}$, written $v \in \widehat{N}_C(\bar{x})$, if

$$\limsup_{x \underset{C}{\to} \bar{x}, x \neq \bar{x}} \frac{\langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \le 0.$$

It is a (general) normal to C at $\bar{x}$, written $v \in N_C(\bar{x})$, if there are sequences $x^\nu \underset{C}{\to} \bar{x}$ and $v^\nu \to v$ with $v^\nu \in \widehat{N}_C(x^\nu)$. We call $\widehat{N}_C(\bar{x})$ the regular normal cone and $N_C(\bar{x})$ the normal cone to C.

(b) A set $C \subset \mathbb{R}^n$ is Clarke regular at $\bar{x} \in C$ if it is locally closed at $\bar{x}$ and $N_C(\bar{x}) = \widehat{N}_C(\bar{x})$.

For a nonempty convex $C \subseteq \mathbb{R}^n$ and $\bar{x} \in C$, $\widehat{N}_C(\bar{x}) = N_C(\bar{x}) = \{v \mid \langle v, z - \bar{x} \rangle \le 0, \text{ for all } z \in C\}$.

Definition 1.2. [17, Definition 8.3] Consider a function $f : \mathbb{R}^n \to \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ and $f(\bar{x}) < \infty$. For a vector $v \in \mathbb{R}^n$, one says that

(a) v is a regular subgradient of f at $\bar{x}$, written $v \in \widehat{\partial} f(\bar{x})$, if

$$f(x) \ge f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(|x - \bar{x}|);$$

(b) v is a (general) subgradient of f at $\bar{x}$, written $v \in \partial f(\bar{x})$, if there are sequence $x^\nu \underset{f}{\to} \bar{x}$ and $v^\nu \in \widehat{\partial} f(x^\nu)$ with $v^\nu \to v$;

(c) v is a horizon subgradient of f at $\bar{x}$, written $v \in \partial^\infty f(\bar{x})$, if there are sequence $x^\nu \underset{f}{\to} \bar{x}$ and $v^\nu \in \widehat{\partial} f(x^\nu)$ with $\lambda^\nu v^\nu \to v$ for some sequence $\lambda^\nu \searrow 0$.

For $f : \mathbb{R}^n \to \mathbb{R}$, the epigraph of f is the set epi $f := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \alpha \ge f(x)\}$.

Definition 1.3. [17, Definition 7.25] A function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is called subdifferentially regular at $\bar{x}$ if $f(\bar{x})$ is finite and epi f is Clarke regular at $(\bar{x}, f(\bar{x}))$ as a subset of $\mathbb{R}^n \times \mathbb{R}$.

2 First-Order Necessary Optimality Conditions

In this section, we present the first-order necessary optimality conditions for $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$. Before proceeding to the optimality conditions, we provide some basic properties.

2.1 Basic Properties

Denote $\phi(x) = \|x\|_p^p$ and the ℓ_p norm ball $\Theta := \{x \in \mathbb{R}^n \mid \phi(x) \leq \theta\}$. In this subsection, we provide basic properties about $\phi(x)$ and Θ. In particular, we derive regular and general subgradients of ϕ and the regular and the general normal cones of Θ, and then show that the ϕ is subdifferentially regular and Θ is Clarke regular on $\mathbb{R}^n$.

The regular, general, and horizon subgradients of ϕ can be calculated as follows.

Theorem 2.1. *For any $\bar{x} \in \mathbb{R}^n$, it holds that*

$$\partial\phi(\bar{x}) = \widehat{\partial}\phi(\bar{x}) = \{v \in \mathbb{R}^n \mid v_j = sign(\bar{x}_j)p|\bar{x}_j|^{p-1},\ j \in \bar{\mathcal{N}}\}, \tag{2.1}$$

$$[\widehat{\partial}\phi(\bar{x})]^\infty = \partial^\infty\phi(\bar{x}) = \{v \in \mathbb{R}^n \mid v_j = 0,\ j \in \bar{\mathcal{N}}\}. \tag{2.2}$$

Therefore, ϕ is subdifferentially regular at every $x \in \mathbb{R}^n$.

Proof. We first consider $|\cdot|^p$ on $\mathbb{R}$. If $\bar{x} \neq 0$, then $\partial|\bar{x}|^p = \widehat{\partial}|\bar{x}|^p = \{\nabla|\bar{x}|^p\} = \{sign(\bar{x})p|\bar{x}|^{p-1}\}$. On the other hand, $\lim\limits_{x\to 0, x\neq 0} \frac{|x|^p}{|x|} = +\infty$, implying $\liminf\limits_{x\to 0, x\neq 0} \frac{|x|^p - |0|^p - v(x-0)}{|x-0|} \geq 0$ for any $v \in \mathbb{R}$. Therefore, if $\bar{x} = 0$, it follows from the definition of regular subgradient that $\mathbb{R} \subset \widehat{\partial}|0|^p \subset \partial|0|^p \subset \mathbb{R}$. Hence, $\widehat{\partial}|0|^p = \partial|0|^p = \mathbb{R}$. By [17, Proposition 10.5] , we have

$$\widehat{\partial}\phi(\bar{x}) = \partial\phi(\bar{x}) = \partial|\bar{x}_1|^p \times \ldots \times \partial|\bar{x}_n|^p = \{v \in \mathbb{R}^n \mid v_j = sign(\bar{x}_j)p|\bar{x}_j|^{p-1},\ j \in \bar{\mathcal{N}}\}.$$

This proves (2.1).

By the definition of horizon cone and (2.1), it is obvious that $[\widehat{\partial}\phi(\bar{x})]^\infty = \{v \in \mathbb{R}^n \mid v_j = 0,\ j \in \bar{\mathcal{N}}\}$. We next prove $\partial^\infty\phi(\bar{x}) = [\widehat{\partial}\phi(\bar{x})]^\infty$.

For any $v \in [\widehat{\partial}\phi(\bar{x})]^\infty$ and $\{\lambda^\nu\} \searrow 0$, we can select sequence $x^\nu \underset{\phi}{\to} \bar{x}$ such that $\mathcal{N}(x^\nu) = \bar{\mathcal{N}}$. Let $v_j^\nu = v_j/\lambda^\nu$. From (2.1), this means $v^\nu \in \widehat{\partial}\phi(x^\nu)$ and $\lambda^\nu v^\nu \to v$. Therefore, $v \in \partial^\infty\phi(x)$.

On the other hand, for x^ν sufficiently close to $\bar{x}$, it holds that $\bar{\mathcal{N}} \subset \mathcal{N}(x^\nu)$. Therefore, by (2.1), $v_j^\nu = sign(\bar{x}_j)p|\bar{x}_j^\nu|^{p-1}, j \in \bar{\mathcal{N}}$ for any $v^\nu \in \widehat{\partial}\phi(x^\nu)$. Hence, for any sequence $\{\lambda^\nu\} \searrow 0$, $\lambda^\nu v_j^\nu \to 0,\ j \in \bar{\mathcal{N}}$, it holds that $v_j = 0, j \in \bar{\mathcal{N}}$ for any $v \in \partial^\infty\phi(\bar{x})$, or, equivalently, $\partial^\infty\phi(\bar{x}) \subset [\widehat{\partial}\phi(\bar{x})]^\infty$. Overall, we have shown that $\partial^\infty\phi(\bar{x}) = [\widehat{\partial}\phi(\bar{x})]^\infty$.

It then follows from [17, Corollary 8.11] that ϕ is subdifferentially regular at any $x \in \mathbb{R}^n$. □

The regular and general normal vectors can be calculated as follows.

Theorem 2.2. *For any $\bar{x} \in \Theta$, $N_\Theta(\bar{x}) = pos\ \partial\phi(\bar{x}) \cup \partial^\infty\phi(\bar{x})$, i.e.,*

$$N_\Theta(\bar{x}) = \begin{cases} \{v \in \mathbb{R}^n \mid v_j = \lambda sign(\bar{x}_j)p|\bar{x}_j|^{p-1}, j \in \bar{\mathcal{N}};\ \lambda \geq 0\} & \text{if } \bar{x} \in \partial\Theta, \\ \{0\} & \text{if } \bar{x} \in int\ \Theta. \end{cases} \tag{2.3}$$

Furthermore, Θ is Clarke regular at any $\bar{x} \in \Theta$, i.e., $\widehat{N}_\Theta(\bar{x}) = N_\Theta(\bar{x})$.

Proof. We only prove the case that $\bar{x} \in \partial\Theta$ since the other is trivial. We have $\bar{x} \neq 0$ and $0 \notin \partial\phi(\bar{x})$. Together with Theorem 2.1 and [17, Proposition 10.3], it holds that

$$\widehat{N}_\Theta(\bar{x}) = N_\Theta(\bar{x}) = \text{pos } \partial\phi(\bar{x}) \cup \partial^\infty\phi(\bar{x})$$

and Θ is Clarke regular at $\bar{x}$. By the definition of pos and (2.1),

$$\text{pos } \partial\phi(\bar{x}) = \{0\} \cup \{\lambda v \in \mathbb{R}^n \mid v_j = sign(\bar{x}_j)p|\bar{x}_j|^{p-1}, j \in \bar{\mathcal{N}};\ \lambda > 0\}.$$

Therefore, it holds that

$$N_\Theta(\bar{x}) = \text{pos } \partial\phi(\bar{x}) \cup \partial^\infty\phi(\bar{x}) = \{v \in \mathbb{R}^n \mid v_j = \lambda sign(\bar{x}_j)p|\bar{x}_j|^{p-1}, j \in \bar{\mathcal{N}};\ \lambda \geq 0\}.$$

□

From [17, Theorem 8.15], we have the following first-order necessary condition for $(\mathcal{P}_1)$. For $(\mathcal{P}_2)$, we only focus on the local minimizers $\bar{x}$ on the boundary of the ℓ_p ball, i.e., $\|\bar{x}\|_p^p = \theta$; otherwise, the characterization of local minimizers reverts to the case of traditional constrained nonlinear problems.

Theorem 2.3. *Suppose f_0 is differentiable over Γ. The following statements hold true.*

(a) If $\partial^\infty\phi(\bar{x})$ contains no vector $v \neq 0$ such that $-v \in N_\Gamma(\bar{x})$, then a necessary condition for $\bar{x}$ to be a local minimizer for $(\mathcal{P}_1)$ is

$$0 \in \nabla f_0(\bar{x}) + \lambda\partial\phi(\bar{x}) + N_\Gamma(\bar{x}). \tag{2.4}$$

(b) Suppose that $\bar{x}$ is a local minimizer of $(\mathcal{P}_2)$ with $\bar{x} \in \partial\Theta$. Then

$$-\nabla f_0(\bar{x}) \in \widehat{N}_{\Theta\cap\Gamma}(\bar{x}) \subset N_{\Theta\cap\Gamma}(\bar{x}). \tag{2.5}$$

Figure 1 shows a conterexample for Theorem 2.3. Obviously, to find an optimal solution, we can only focus on the top semicircle of the given ball, so that the original problem is equivalent to

$$\min\ (x_1+1)^2 + \sqrt{x_1} + \sqrt{1 - \sqrt{-x_1^2 + 2x_1}} \text{ s.t. } 0 \leq x_1 \leq 2$$

The derivative over the domain is always positive, therefore $x^* = (0, 1)$ is a global minimizer. However, there exist $v \in \{\nu \in \mathbb{R}^2 \mid \nu_1 > 0, \nu_2 = 0\}$, such that $\partial^\infty(\bar{x}) \ni v \neq 0$ and $-v \in N_\Gamma(\bar{x})$. In this case, one can see (2.4) does not hold at x^*.

2.2 Optimality conditions for $(\mathcal{P}_1)$

To make condition (2.4) for $(\mathcal{P}_1)$ informative, we need to clarify when $-v \in N_\Gamma(\bar{x})$ happens and how to calculate the elements in $N_\Gamma(\bar{x})$. For this purpose, we define the following extended Mangasarian-Fromovitz constraint qualification (EMFCQ). The EMFCQ holds at $\bar{x} \in \Gamma$ for Γ if the subvectors $\nabla_{\bar{\mathcal{N}}} f_j(\bar{x})$, $j \in \mathcal{E} \cup \bar{\mathcal{A}}$ are linearly independent and there exists $d \in \mathbb{R}^{|\bar{\mathcal{N}}|}$ such that

$$\langle \nabla_{\bar{\mathcal{N}}} f_j(\bar{x}), d\rangle = 0,\ j \in \mathcal{E} \quad \text{and} \quad \langle \nabla_{\bar{\mathcal{N}}} f_j(\bar{x}), d\rangle < 0,\ j \in \bar{\mathcal{A}}. \tag{2.6}$$

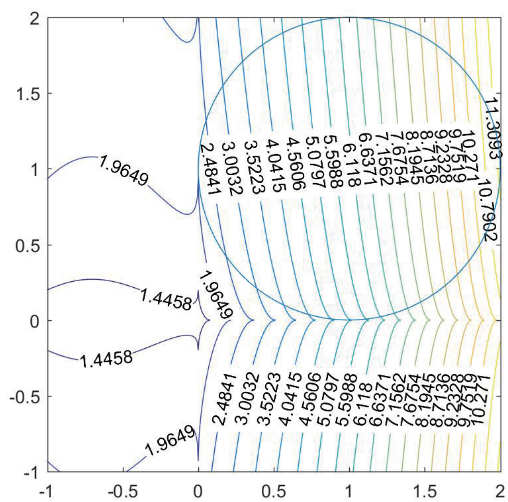

Figure 1: A conterexample for Theorem 2.3.For $(\mathcal{P}_1)$, the contour of ℓ_p $(p = 0.5)$ regularization problem with $F(x) = (x_1 + 1)^2 + \|x\|_p^p$ and $\Gamma = \{(x_1 - 1)^2 + (x_2 - 1)^2 \leq 1\}$.

Obviously, the EMFCQ is a weaker condition than the ELICQ proposed in [21]. Moreover, if the EMFCQ holds at $\bar{x} \in \Gamma$ for Γ, then the MFCQ holds naturally true at $\bar{x}$ for Γ; letting

$$\Lambda_\Gamma(\bar{x}) = \left\{ \sum_{j \in \mathcal{E} \cup \bar{\mathcal{A}}} y_j \nabla f_j(\bar{x}) \mid y_j \geq 0, j \in \bar{\mathcal{A}} \right\}, \tag{2.7}$$

we have from [17, Theorem 6.14] that Γ is Clarke regular at $\bar{x}$ and $N_\Gamma(\bar{x}) = \Lambda_\Gamma(\bar{x})$.

Theorem 2.4. (a) *Suppose the EMFCQ is satisfied at $\bar{x} \in \Gamma$ for Γ. Then $\partial^\infty \phi(\bar{x})$ contains no vector $v \neq 0$ such that $-v \in N_\Gamma(\bar{x})$. Furthermore, a necessary condition for $\bar{x}$ to be a local minimizer for $(\mathcal{P}_1)$ is that there exist $y_j,\ j \in \mathcal{E}$ and $y_j \geq 0,\ j \in \mathcal{A}$ such that*

$$\nabla_i f_0(\bar{x}) + \lambda p \text{sign}(\bar{x}_i)|\bar{x}_i|^{p-1} + \sum_{j \in \mathcal{E} \cup \bar{\mathcal{A}}} y_j \nabla_i f_j(\bar{x}) = 0,\ i \in \bar{\mathcal{N}}. \tag{2.8}$$

(b) *Suppose Γ is closed and convex in $(\mathcal{P}_1)$. If $\bar{x}$ is a local minimizer of $(\mathcal{P}_1)$ and $\partial^\infty \phi(\bar{x})$ contains no vector $v \neq 0$ such that $-v \in N_\Gamma(\bar{x})$, then it holds that*

$$\nabla_i f_0(\bar{x}) + \lambda p \text{sign}(\bar{x}_i)|\bar{x}_i|^{p-1} + v_i = 0, i \in \bar{\mathcal{N}};\ v \in N_\Gamma(\bar{x}). \tag{2.9}$$

Proof. (a) Assume by contradiction that there exists nonzero $v \in \partial^\infty \phi(\bar{x})$ such that $-v \in N_\Gamma(\bar{x})$; then it follows from Theorem 2.1 and (2.7) that

$$\sum_{j \in \mathcal{E} \cup \bar{\mathcal{A}}} y_j \nabla_{\bar{\mathcal{N}}} f_j(\bar{x}) = -v_{\bar{\mathcal{N}}} = 0; \sum_{j \in \mathcal{E} \cup \bar{\mathcal{A}}} y_j \nabla_{\bar{\mathcal{Z}}} f_j(\bar{x}) = -v_{\bar{\mathcal{Z}}} \neq 0;\ y_j \geq 0, j \in \bar{\mathcal{A}}. \tag{2.10}$$

Since EMFCQ holds true at $\bar{x} \in \Gamma$, the dual form [19] of condition (2.6) tells that $y_j = 0, j \in \mathcal{E} \cup \bar{\mathcal{A}}$ is the unique solution of the the system

$$\sum_{j \in \mathcal{E} \cup \bar{\mathcal{A}}} y_j \nabla f_j(\bar{x}) = 0,\ y_j \geq 0, j \in \bar{\mathcal{A}}.$$

It follows that $\sum_{j\in\mathcal{E}\cup\bar{\mathcal{A}}} y_j \nabla_{\bar{\mathcal{Z}}} f_j(\bar{x}) = 0$, contradicting (2.10). Therefore, for any nonzero $v \in \partial^\infty \phi(\bar{x})$, $-v \notin N_\Gamma(\bar{x})$. From Theorem 2.3, at a local minimizer $\bar{x}$ of $(\mathcal{P}_1)$, (2.8) is satisfied.

(b) This is trivially true from Theorem 2.3. □

We call the conditions (2.8) the Karush-Kuhn-Tucker (KKT) conditions for $(\mathcal{P}_1)$. Using the notation of Λ_Γ, it can also be equivalently written as

$$-\nabla f_0(\bar{x}) \in \lambda\partial\phi(\bar{x}) + \Lambda_\Gamma(\bar{x}). \tag{2.11}$$

2.3 Optimality conditions for $(\mathcal{P}_2)$

We also consider other verifiable forms of condition (2.5) if some constraint qualification is satisfied at $\bar{x}$. For $\bar{x} \in \Theta \cap \Gamma$, define the extended linearized cone $\Upsilon_{\Theta\cap\Gamma}(\bar{x})$ as:

$$\Upsilon_{\Theta\cap\Gamma}(\bar{x}) := \{d \in \mathbb{R}^n \mid \langle v, d\rangle \le 0, \forall v \in \partial\phi(\bar{x}); \langle \nabla f_j(\bar{x}), d\rangle = 0, j \in \mathcal{E};\ \langle \nabla f_j(\bar{x}), d\rangle \le 0, j \in \bar{\mathcal{A}}\}.$$

Obviously,

$$\begin{aligned} &N_\Theta(\bar{x}) + \Lambda_\Gamma(\bar{x}) = \Upsilon_{\Theta\cap\Gamma}(\bar{x})^* \\ =&\{v \in \mathbb{R}^n \mid v_i = y_0 psign(\bar{x}_i)|\bar{x}_i|^{p-1} + \sum_{j\in\bar{\mathcal{A}}\cup\mathcal{E}} y_j \nabla_i f_j(\bar{x}), i \in \bar{\mathcal{N}};\ y_j \ge 0, j \in \{0\}\cup\bar{\mathcal{A}}\}. \end{aligned}$$

It follows from [17, Theorem 6.14] that $\Upsilon_{\Theta\cap\Gamma}(\bar{x})^* \subset \widehat{N}_{\Theta\cap\Gamma}(\bar{x})$. Hence, we have the following result.

Proposition 2.5. *For $\bar{x} \in \Gamma \cap \Theta$ with $\bar{x} \in \partial\Theta$, $\Upsilon_{\Theta\cap\Gamma}(\bar{x})^* \subset \widehat{N}_{\Theta\cap\Gamma}(\bar{x})$. Therefore, if $-\nabla f_0(\bar{x}) \in \Upsilon_{\Theta\cap\Gamma}(\bar{x})^*$, meaning that there exist $y_j \ge 0, j \in \{0\}\cup\bar{\mathcal{A}}$ such that*

$$\nabla f_0(\bar{x}) + y_0 p\mathrm{sign}(\bar{x}_i)|\bar{x}_i|^{p-1} + \sum_{j\in\bar{\mathcal{A}}\cup\mathcal{E}} y_j \nabla_i f_j(\bar{x}) = 0, i \in \bar{\mathcal{N}},$$

then the first-order necessary condition (2.5) is satisfied at $\bar{x}$.

The EMFCQ for $(\mathcal{P}_2)$ holds at $\bar{x} \in \partial\Theta\cap\Gamma$ if the subvectors $sign(x_{\bar{\mathcal{N}}})|x_{\bar{\mathcal{N}}}|^{p-1}$, $\nabla_{\bar{\mathcal{N}}} f_j(\bar{x})$, $j \in \mathcal{E}\cup\bar{\mathcal{A}}$ are linearly independent and there exists $d \in \mathbb{R}^{|\bar{\mathcal{N}}|}$ such that

$$\langle psign(x_{\bar{\mathcal{N}}})|x_{\bar{\mathcal{N}}}|^{p-1}, d\rangle < 0, \langle \nabla_{\bar{\mathcal{N}}} f_j(\bar{x}), d\rangle = 0,\ j \in \mathcal{E} \text{ and } \langle \nabla_{\bar{\mathcal{N}}} f_j(\bar{x}), d\rangle < 0,\ j \in \bar{\mathcal{A}}.$$

Equivalently, the dual form of EMFCQ for $(\mathcal{P}_2)$ holds at $\bar{x} \in \Theta\cap\Gamma$ if $y_j = 0, j \in \{0\}\cup\bar{\mathcal{A}}$ is the unique solution of

$$\left\{ y_0 psign(\bar{x}_i)|\bar{x}_i|^{p-1} + \sum_{j\in\bar{\mathcal{A}}\cup\mathcal{E}} y_j \nabla_i f_j(\bar{x}) = 0, i \in \bar{\mathcal{A}};\ y_j \ge 0, j \in \{0\}\cup\bar{\mathcal{A}} \right\}.$$

We now state the necessary optimality conditions for $(\mathcal{P}_2)$.

Theorem 2.6. *Suppose $\bar{x} \in \Gamma\cap\Theta$ with $\bar{x} \in \partial\Theta$ is a local minimizer for $(\mathcal{P}_2)$.*

(a) *If the EMFCQ holds at $\bar{x}$, then there exist y_j, $j \in \mathcal{E}$ and $y_j \ge 0$, $j \in \{0\}\cup\bar{\mathcal{A}}$, such that*

$$\nabla_i f_0(\bar{x}) + y_0 psign(\bar{x}_i)|\bar{x}_i|^{p-1} + \sum_{j\in\bar{\mathcal{A}}\cup\mathcal{E}} y_j \nabla_i f_j(\bar{x}) = 0, i \in \bar{\mathcal{N}}. \tag{2.12}$$

(b) *Suppose* Γ *is closed and convex. If* $v = 0$ *is the only vector such that* $f\,v \in N_\Theta(\bar{x})$ *and* $-v \in N_\Gamma(\bar{x})$*, then* $N_{\Theta\cap\Gamma}(\bar{x}) = N_\Theta(\bar{x}) + N_\Gamma(\bar{x})$*. Therefore, if* $\bar{x}$ *is a local minimizer for* $(\mathcal{P}_2)$*, then*

$$-\nabla_i f(\bar{x}) + y_0 p\text{sign}(\bar{x}_i)|\bar{x}_i|^{p-1} + v_i = 0, i \in \bar{\mathcal{N}};\ \ y_0 \geq 0;\ \ v \in N_\Gamma(\bar{x}).$$

(c) *Suppose* $\Gamma = \mathbb{R}^n$*. If* $\bar{x}$ *is a local minimizer for* $(\mathcal{P}_2)$*, then there exists* $y_0 \geq 0$ *such that*

$$-\nabla_i f(\bar{x}) + y_0 p\text{sign}(\bar{x}_i)|\bar{x}_i|^{p-1} = 0, i \in \bar{\mathcal{N}}.$$

Proof. (a) From [17, Theorem 6.14], if the EMFCQ holds for $(\mathcal{P}_2)$ at $\bar{x} \in \Gamma\cap\Theta$ with $\bar{x} \in \partial\Theta$, then $\Theta \cap \Gamma$ is regular at $\bar{x}$ and

$$\widehat{N}_{\Theta\cap\Gamma}(\bar{x}) = \Upsilon_{\Theta\cap\Gamma}(\bar{x})^*. \tag{2.13}$$

By Theorem 2.3, (a) is true.

(b) By [17, Theorem 6.42], $N_{\Theta\cap\Gamma}(\bar{x}) = N_\Theta(\bar{x}) + N_\Gamma(\bar{x})$. Therefore, if $\bar{x}$ is a local minimizer, then $\bar{x} \in \widehat{N}_{\Theta\cap\Gamma}(\bar{x}) \subset N_{\Theta\cap\Gamma}(\bar{x}) = N_\Theta(\bar{x}) + N_\Gamma(\bar{x})$.

(c) Trivial by (b). ☐

We call the conditions (2.12) the Karush-Kuhn-Tucker (KKT) conditions for $(\mathcal{P}_2)$. Using the notation of Λ_Γ, it can also be equivalently written as

$$-\nabla f_0(\bar{x}) \in N_\Theta(\bar{x}) + \Lambda_\Gamma(\bar{x}). \tag{2.14}$$

3 First-Order Sequential Optimality Condition

In this section, we study the sequential optimality conditions under the approximate Karush-Kuhn-Tucker (AKKT) conditions, which are defined as follows.

Definition 3.1. (i) For $(\mathcal{P}_1)$, we say that $\bar{x} \in \Gamma$ satisfies the AKKT if there exist $\{x^\nu\} \subset \mathbb{R}^n$, $\{y_j^\nu\} \subset \mathbb{R}$, $j \in \mathcal{E} \cup \bar{\mathcal{A}}$ such that $\lim_{\nu\to\infty} x^\nu = \bar{x}$, $y_j^\nu \geq 0, j \in \bar{\mathcal{A}}$ and

$$\lim_{\nu\to\infty} \nabla_i f(x^\nu) + \lambda psign(x_i^\nu)|x_i^\nu|^{p-1} + \sum_{j\in\mathcal{E}\cup\bar{\mathcal{A}}} y_j^\nu \nabla_i f_j(x^\nu) = 0,\ \forall i \in \bar{\mathcal{N}}.$$

(ii) For $(\mathcal{P}_2)$, we say that $\bar{x} \in \Theta \cap \Gamma$ with $\bar{x} \in \partial\Theta$ satisfies the AKKT if there exist $\{x^\nu\} \subset \mathbb{R}^n$, $\{y_j^\nu\} \subset \mathbb{R}$, $j \in \{0\} \cup \mathcal{E} \cup \bar{\mathcal{A}}$ such that $\lim_{\nu\to\infty} x^\nu = \bar{x}$, $y_j^\nu \geq 0, j \in \{0\} \cup \bar{\mathcal{A}}$ and

$$\lim_{\nu\to\infty} \nabla_i f(x^\nu) + y_0^\nu psign(x_i^\nu)|x_i^\nu|^{p-1} + \sum_{j\in\mathcal{E}\cup\bar{\mathcal{A}}} y_j^\nu \nabla_i f_j(x^\nu) = 0,\ \forall i \in \bar{\mathcal{N}}.$$

Next we provide properties of $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$ under which the AKKT implies the KKT. This property is named the extended cone-continuity property (ECCP), which is defined as follows.

Definition 3.2. (a) We say that $\bar{x} \in \Gamma$ satisfies the ECCP for $(\mathcal{P}_1)$ if the set-valued mapping $\lambda\partial\phi(x) + \Lambda_\Gamma(x)$ is outer semicontinuous at $\bar{x}$, that is,

$$\limsup_{x^\nu\to\bar{x}} [\lambda\partial\phi(x^\nu) + \Lambda_\Gamma(x^\nu)] \subset \lambda\partial\phi(\bar{x}) + \Lambda_\Gamma(\bar{x}).$$

If Γ is a closed and convex set, Λ_Γ can be replaced by N_Γ.

(b) Similarly, we say $\bar{x} \in \Theta \cap \Gamma$ satisfies the ECCP for $(\mathcal{P}_2)$ if the set-valued mapping $\Lambda_{\Theta\cap\Gamma}(x)$ is outer semicontinuous at $\bar{x}$, that is,

$$\limsup_{x^\nu \to \bar{x}} [N_\Theta(x^\nu) + \Lambda_\Gamma(x^\nu)] \subset N_\Theta(\bar{x}) + \Lambda_\Gamma(\bar{x}).$$

If Γ is a closed and convex set, Λ_Γ can be replaced by N_Γ.

Example 3 (ECCP is easy to be satisfied). In $\mathbb{R}^2$, consider $x^* = (1,1)^T$ and the inequality constraints defined by $g_1(x_1,x_2) = (x_1-1)^3,\ g_2(x_1,x_2) = (x_1-1)\exp(x_2-1)$. Therefore, Γ is not a convex set. Clearly, x^* is feasible and both constraints are active at x^*.

In the next, we show ECCP is satisfied at $x^* = (1,1) \in \Gamma$ for $(\mathcal{P}_1)$. First, consider a sequence $\{x^\nu\} \to x^*$. It holds that

$$\partial\phi(x^\nu) = \{u^\nu\} \text{ and } \partial\phi(x^*) = \{u^*\}$$

with $u^\nu = (p(x_1^\nu)^{p-1}, p(x_2^\nu)^{p-1})^T$ and $u^* = (p,p)^T$. Since

$$\nabla g_1(x_1^\nu, x_2^\nu) = \begin{bmatrix} 3(x_1^\nu - 1)^2 \\ 0 \end{bmatrix} \text{ and } \nabla g_2(x_1^\nu, x_2^\nu) = \begin{bmatrix} \exp(x_2^\nu - 1) \\ (x_1^\nu - 1)\exp(x_2^\nu - 1) \end{bmatrix},$$

it holds that $\Lambda_\Gamma(x^\nu) = \{w^\nu\}$ and $\Lambda_\Gamma(x^*) = \{y_1(1,0)^T \mid y_1 \geq 0\}$ with

$$w^\nu = (w_1^\nu, w_2^\nu)^T = \begin{bmatrix} \mu_1^\nu 3(x_1^\nu - 1)^2 + \mu_2^\nu \exp(x_2^\nu - 1) \\ \mu_2^\nu (x_1^\nu - 1)\exp(x_2^\nu - 1) \end{bmatrix}.$$

Now assume $u^\nu + w^\nu \to u^* + w^*$, it suffices to show $u^* + w^* \in \partial\phi(x^*) + \Lambda_\Gamma(x^*)$, which is equivalent to $w^* \in \Lambda_\Gamma(x^*)$. Suppose by contradiction that $w^* = (w_1^*, w_2^*)^T$ does not belong to $\Lambda_\Gamma(x^*)$, so w_2^* must be nonzero. There exists $\rho > 0$ such that for k large enough

$$|w_2^\nu| = \mu_2^\nu |(x_1^\nu - 1)\exp(x_2^\nu - 1)| > \rho > 0. \tag{3.1}$$

In particular $x_1^\nu \neq 0$. Using $\mu_1^\nu \geq 0$ and (3.1), we get

$$w_1^\nu = 3\mu_1^\nu (x_1^\nu - 1)^2 + \mu_2^\nu \exp(x_2^\nu - 1) \geq \mu_2^\nu \exp(x_2^\nu - 1) \geq \frac{|w_2^\nu|}{|x_1^\nu - 1|} > \frac{\rho}{|x_1^\nu - 1|} > 0. \tag{3.2}$$

Taking limits in (3.2) we obtain $w_1^\nu \to \infty$, a contradiction with its convergence. Hence, w^* must be in $\Lambda_\Gamma(x^*)$ and thus

$$\limsup_{x^\nu \to \bar{x}} [\lambda\partial\phi(x^\nu) + \Lambda_\Gamma(x^\nu)] \subset \lambda\partial\phi(x^*) + \Lambda_\Gamma(x^*).$$

Now we show ECCP is satisfied at $x^* = (1,1) \in \Gamma$ for $(\mathcal{P}_2)$. If $x^* \in \operatorname{int}\Theta$, then for sufficiently large ν, $N_\Theta(x^\nu) = N_\Theta(x^*) \equiv \{0\}$. The proof is trivially the same as above. We only consider the case that $x^* \in \partial\Theta$. In this case, $N_\Theta(x^\nu) = \{\lambda u^\nu : \lambda \geq 0\}$ and $N_\Theta(x^*) = \{\lambda u^* : \lambda \geq 0\}$. Suppose $h^\nu = \lambda^\nu u^\nu \in N_\Theta(x^\nu)$ with $\lambda^\nu \geq 0$ and $h^\nu \to h^*$ with $\lambda^\nu \to \lambda^* \geq 0$. It suffices to show that $h^* + w^* \in N_\Theta(x^*) + \Lambda_\Gamma(x^*)$, which is equivalent to $w^* \in \Lambda_\Gamma(x^*)$. This reverts to the proof for $(\mathcal{P}_1)$. Hence,

$$\limsup_{x^\nu \to x^*} [N_\Theta(x^\nu) + \Lambda_\Gamma(x^\nu)] \subset N_\Theta(x^*) + \Lambda_\Gamma(x^*).$$

In the following theorem we show that if ECCP holds, the AKKT implies the KKT.

Theorem 3.3. (a) *For $(\mathcal{P}_1)$, suppose the AKKT holds at $\bar{x} \in \Gamma$. If the ECCP holds at $\bar{x}$, then the KKT condition* (2.11) *is satisfied at $\bar{x}$.*

(b) *For $(\mathcal{P}_2)$, suppose the AKKT holds at $\bar{x} \in \partial\Theta \cap \Gamma$. If the ECCP holds at $\bar{x}$, then the KKT condition* (2.14) *is satisfied at $\bar{x}$.*

Proof. (a) First of all, $\bar{\mathcal{N}} \subset \mathcal{N}(x^\nu)$ for x^ν sufficiently close to $\bar{x}$. Since the AKKT condition holds at $\bar{x}$, there exist $\{x^\nu\} \to \bar{x}$ and $\{w^\nu\} \subset \mathbb{R}^n$ such that $\nabla f(x^\nu) + w^\nu \to 0$, where $w^\nu \in \lambda\partial\phi(x^\nu) + \Lambda_\Gamma(x^\nu)$. Taking limits and using the continuity of the gradient of f near $\bar{x}$, we obtain

$$\begin{aligned}-\nabla f(\bar{x}) = \lim_{\nu\to\infty} w^\nu &\in \limsup_{\nu\to\infty}[\lambda\partial\phi(x^\nu) + \Lambda_\Gamma(x^\nu)] \\ &\subset \limsup_{x\to\bar{x}}[\lambda\partial\phi(x) + \Lambda_\Gamma(x)] \subset \lambda\partial\phi(\bar{x}) + \Lambda_\Gamma(\bar{x}),\end{aligned}$$

where the last inclusion follows from the ECCP. Therefore, $-\nabla f(\bar{x}) \in \lambda\partial\phi(\bar{x}) + \Lambda_\Gamma(\bar{x})$.

(b) Similarly, since the AKKT condition holds at $\bar{x}$, there exist $\{x^\nu\} \to \bar{x}$ and $\{w^\nu\} \subset \mathbb{R}^n$ such that $\nabla f(x^\nu) + w^\nu \to 0$, where $w^\nu \in N_\Theta(x^\nu) + \Lambda_\Gamma(x^\nu)$. Taking limits and using the continuity of the gradient of f near $\bar{x}$, we obtain

$$-\nabla f(\bar{x}) = \lim_{\nu\to\infty} w^\nu \in \limsup_{\nu\to\infty} N_\Theta(x^\nu) + \Lambda_\Gamma(x^\nu) \subset \lim_{x\to\bar{x}} N_\Theta(x) + \Lambda_\Gamma(x) \subset N_\Theta(\bar{x}) + \Lambda_\Gamma(\bar{x}),$$

where the last inclusion follows from the ECCP. Therefore, $-\nabla f(\bar{x}) \in N_\Theta(\bar{x}) + \Lambda_\Gamma(\bar{x})$. □

We discuss the cases when the ECCP holds true. For $(\mathcal{P}_1)$, we have the following results.

Proposition 3.4. *For $(\mathcal{P}_1)$, the ECCP holds true for any of the following cases.*

(a) *Λ_Γ is outer semicontinuous at $\bar{x}$ and $\partial^\infty\phi(\bar{x})$ contains no vector $v \neq 0$ such that $-v \in \Lambda_\Gamma(\bar{x})$.*

(b) *The EMFCQ is satisfied at $\bar{x}$.*

(c) *Γ is a closed and convex set and $\partial^\infty\phi(\bar{x})$ contains no vector $v \neq 0$ such that $-v \in N_\Gamma(\bar{x})$.*

Proof. (a) Let $\bar{w}$ be an element of $\limsup_{x\to\bar{x}}[\lambda\partial\phi(x) + \Lambda_\Gamma(x)]$, so there are sequences $\{x^\nu\}$, $\{w^\nu\}$, $\{u^\nu\}$ such that $x^\nu \to \bar{x}$, $w^\nu \to \bar{w}$ and

$$w_i^\nu = u_i^\nu + \sum_{j\in\mathcal{E}\cup\bar{\mathcal{A}}} y_j^\nu \nabla_i f_j(x^\nu) \tag{3.3}$$

with $u^\nu \in \partial\lambda\phi(x^\nu)$ and $y_j^\nu \in \mathbb{R}_+, j \in \bar{\mathcal{A}}$.

If $\{y_j^\nu\}, j \in \mathcal{E} \cup \bar{\mathcal{A}}$ are bounded, then they all have limits $\bar{y}_j, j \in \mathcal{E} \cup \bar{\mathcal{A}}$; moreover, $\{u^\nu\}$ is also bounded and $\bar{u} := \lim_{\nu\to\infty} u^\nu \in \lambda\partial\phi(\bar{x})$ (taking limis on a convergent subsequence if necessary) due to the outer semicontinuity of $\partial\phi$. By possibly extracting an convergent subsequence, we have

$$u^\nu + \sum_{j\in\mathcal{E}\cup\bar{\mathcal{A}}} y_j^\nu \nabla_i f_j(x^\nu) \to \bar{u} + \sum_{j\in\mathcal{E}\cup\bar{\mathcal{A}}} \bar{y}_j \nabla_i f_j(\bar{x}) \in \lambda\partial\phi(\bar{x}) + \Lambda_\Gamma(\bar{x}),$$

meaning $\lambda\partial\phi + \Lambda_\Gamma$ is outer semicontinuous at $\bar{x}$.

If $\{y_j^\nu\}, j \in \mathcal{E} \cup \bar{\mathcal{A}}$ are unbounded, letting $M^\nu = \max\{|y_j^\nu|, j \in \mathcal{E} \cup \bar{\mathcal{A}}\}$. Dividing (3.3) by M^ν, we arrive at

$$\frac{w^\nu}{M^\nu} = \frac{u^\nu}{M^\nu} + \sum_{j \in \mathcal{E} \cup \bar{\mathcal{A}}} \frac{y_j^\nu}{M^\nu} \nabla_i f_j(x^\nu)$$

Since $\max\{y_j^\nu/M^\nu, j \in \{0\} \cup \mathcal{E} \cup \bar{\mathcal{A}}\} = 1$ for all ν, they have nonzero limit point $\tilde{y}_j, j \in \mathcal{E} \cup \bar{\mathcal{A}}$. Moreover, $\bar{u} := \lim_{\nu \to \infty} \frac{u^\nu}{M^\nu} \in \lambda \partial^\infty \phi(\bar{x})$, we can extract a convergent subsequence. Thus, taking limits above, we get

$$\lambda \partial^\infty \phi(\bar{x}) \ni \bar{u} = -\sum_{j \in \mathcal{E} \cup \bar{\mathcal{A}}} \tilde{y}_j \nabla_i f_j(\bar{x}) \in \Lambda_\Gamma(\bar{x}),$$

a contradiction.

(b) If the EMFCQ holds at $\bar{x}$, then $\partial^\infty \phi(\bar{x})$ contains no vector $v \neq 0$ such that $-v \in \Lambda_\Gamma(\bar{x})$ and $\Lambda_\Gamma = N_\Gamma$ by Theorem 2.4(a). This reverts to (a).

(c) Let $\bar{w}$ be an element of $\limsup\limits_{x \to \bar{x}} [\lambda \partial \phi(x) + N_\Gamma(x)]$, so there are sequences $\{x^\nu\}$, $\{w^\nu\}$, $\{u^\nu\}$, $\{v^\nu\}$ such that $x^\nu \to \bar{x}$, $w^\nu \to \bar{w}$ and

$$w^\nu = u^\nu + v^\nu \tag{3.4}$$

with $u^\nu \in \partial \lambda \phi(x^\nu)$ and $v^\nu \in N_\Gamma(x^\nu)$.

If $\{v^\nu\}$ are bounded, then $\{u^\nu\}$ and $\{v^\nu\}$ all have limits $\bar{u}$ and $\bar{v}$. Moreover, $\bar{u} := \lim\limits_{\nu \to \infty} u^\nu \in \lambda \partial \phi(\bar{x})$ and $\bar{v} \in N_\Gamma(\bar{x})$ (possibly taking limits on a convergent subsequence) due to the outer semicontinuity of $\partial \phi$ and N_Γ. By possibly extracting an convergent subsequence, we have

$$u^\nu + v^\nu \to \bar{u} + \bar{v} \in \lambda \partial \phi(\bar{x}) + \Lambda_\Gamma(\bar{x}),$$

meaning $\lambda \partial \phi + \Lambda_\Gamma$ is outer semicontinuous at $\bar{x}$.

If $\{v^\nu\}$ are unbounded, letting $M^\nu = \|v^\nu\|$. Dividing (3.3) by M^ν, we arrive at

$$\frac{w^\nu}{M^\nu} = \frac{u^\nu}{M^\nu} + \frac{v^\nu}{M^\nu}.$$

Since $\frac{v^\nu}{M^\nu} = 1$ for all ν, it has nonzero limit point $\tilde{v} \in N_\Gamma(\bar{x})$ due to the outer semicontinuity of N_Γ. Moreover, $\bar{u} := \lim\limits_{\nu \to \infty} \frac{u^\nu}{M^\nu} \in \lambda \partial^\infty \phi(\bar{x})$, we can extract a convergent subsequence. Thus, taking limits above, we get

$$\lambda \partial^\infty \phi(\bar{x}) \ni \bar{u} = -\bar{v} \in N_\Gamma(\bar{x}),$$

a contradiction. □

As for $(\mathcal{P}_2)$, we have the following results.

Proposition 3.5. *For $(\mathcal{P}_2)$, the ECCP holds true for any of the following cases.*

(a) *Λ_Γ is outer semicontinuous at $\bar{x}$ and the only solution for $v_1 + v_2 = 0$ with $v_1 \in N_\Theta(\bar{x})$ and $v_2 \in \Lambda_\Gamma(\bar{x})$ is $v_1 = v_2 = 0$.*

(b) *The EMFCQ is satisfied at $\bar{x}$.*

(c) *Γ is a closed and convex set and the only solution for $v_1 + v_2 = 0$ with $v_1 \in N_\Theta(\bar{x})$ and $v_2 \in N_\Gamma(\bar{x})$ is $v_1 = v_2 = 0$.*

Proof. (a) The proof is similar to the argument for Proposition 3.4(a) by replacing the role of $u^\nu \in \lambda\partial\phi(\bar{x})$ with $y_0 psign(x^\nu_{\bar{\mathcal{N}}})|x^\nu_{\bar{\mathcal{N}}}|^{p-1} \in N_\Theta(x^\nu)$ (x^ν is selected sufficiently close to $\bar{x}$ so that $\bar{\mathcal{N}} \subset \mathcal{N}(x^\nu)$) and considering the boundedness of $\{y^\nu_j\}, j \in \{0\} \cup \mathcal{E} \cup \bar{\mathcal{A}}$ instead of $\{y^\nu_j\}, j \in \mathcal{E} \cup \bar{\mathcal{A}}$. Therefore, we skip the details of the proof.

(b) The EMFCQ for $(\mathcal{P}_2)$ is equivalent to saying that the only solution for $v_1 + v_2 = 0$ with $v_1 \in N_\Theta(\bar{x})$ and $v_2 \in \Lambda_\Gamma(\bar{x})$ is $v_1 = v_2 = 0$. Moreover, notice that the EMFCQ for $(\mathcal{P}_2)$ implies the EMFCQ for $(\mathcal{P}_1)$. Therefore, we have from [17, Theorem 6.14] that Γ is regular at $\bar{x}$ and $N_\Gamma(\bar{x}) = \Lambda_\Gamma(\bar{x})$. This case then reverts to (a).

(c) The proof is similar to the argument for (3.4)(c) by replacing the role of $\lambda\partial\phi$ with N_Θ; therefore it is skipped. □

4 Convergence Analysis Using AKKT

The research on algorithms for solving general constrained problems involving ℓ_p norms is so far limited. To the best of our knowledge, only simple cases such as linearly constrained or convex set constrained cases have been studied. When solving for general constrained problems with ℓ_p norm, many works focus on the reformulations where the constraint violation is penalized in the objective [15, 5]. We now extend the existing algorithms for solving unconstrained cases to general constrained cases and prove the global convergence by showing that AKKT conditions discussed in the previous section are satisfied at the limit point. In fact, the iteratively reweighted proposed here represents a common and popular class of methods for handling the nonconvex regularization problems. For example, in [5, 13, 22, 23] iteratively reweighted algorithms for solving the unconstrained nonconvex ℓ_p norm problems were proposed. In [8, 11], relaxation parameters were used to transform the nonconvex and nonsmooth sparsity-inducing terms into smooth approximate functions and in [25] Wang et al. proposed a general formulation of nonconvex and nonsmooth group-sparse optimization problems with convex set constraint. Our main purpose here is to show how the proposed sequential optimality conditions is used in the convergence analysis.

We extend the framework of iteratively reweighted ℓ_1 methods for solving unconstrained ℓ_p regularization problems [24] to constrained cases $(\mathcal{P}_1)$,

$$\min\ F(x) := f_0(x) + \lambda\|x\|_p^p \quad \text{s.t.}\quad f_j(x) \le 0,\ \forall j \in \mathcal{I};\ f_j(x) = 0,\ \forall j \in \mathcal{E}. \tag{4.1}$$

where $f_0 : \mathbb{R}^n \to \mathbb{R}$ is continuously Lipschitz differentiable with constant $L_f \ge 0$. Here Γ is assumed to be closed and convex.

We first formulated a smooth approximation $F(x;\epsilon)$ of $F(x)$

$$F(x;\epsilon) := f_0(x) + \lambda\sum_{i=1}^n (|x_i| + \epsilon_i)^p,$$

where $\epsilon \in \mathbb{R}^n_{++}$. At kth iteration, a convex local model to approximate $F(x;\epsilon)$ is constructed

$$G(x; x^k, \epsilon^k) := Q_k(x) + \lambda\sum_{i=1}^n w(x_i^k, \epsilon_i^k)|x_i|$$

where the weights are given by $w(x_i^k, \epsilon_i^k) = p(|x_i^k| + \epsilon_i^k)^{p-1}$ and $Q_k(x)$ represents a local approximation mode to f at x^k and is generally assumed to be smooth and convex.

As in [24], we make the following assumptions about the choice of (x^0, ϵ^0) and $Q_k(\cdot)$.

Algorithm 1 General framework of iteratively reweighted ℓ_1 approach

Require: $\alpha \in (0,1)$, $\epsilon^0 \in \mathbb{R}^n_{++}$ and x^0.

1: Initialization: Set $k = 0$.
2: **repeat**
3: (Reweighing) Compute $w_i^k = p(|x_i^k| + \epsilon_i^k)^{p-1}$, $i = 1, \cdots, n$.
4: (Solving the Subproblem) $x^{k+1} \leftarrow \arg\min_{x\in\Gamma}\{Q_k(x) + \lambda \sum_{i=1}^{n} w_i^k |x_i|\}$.
5: (Update ϵ) Set $\epsilon^{k+1} \in (0, \alpha\epsilon^k)$.
6: Set $k \leftarrow k+1$ and go to step 3.
7: **until** Convergence

Assumption 4.1. The initial point (x^0, ϵ^0) and local model $Q_k(\cdot)$ are such that
(i) The level set $\mathcal{L}(F^0) := \{x \mid F(x) \le F^0\}$ is bounded where $F^0 := F(x^0; \epsilon^0)$.
(ii) For all $k \in \mathbb{N}, \nabla Q_k(x^k) = \nabla f(x^k)$, $Q_k(\cdot)$ is strongly convex with constant $M > L_f/2 > 0$.

First, we show that $F(x, \epsilon)$ is monotonically decreasing over the iterates (x^k, ϵ^k). Define the following two terms

$$\Delta F(x^{k+1}, \epsilon^{k+1}) := F(x^k, \epsilon^k) - F(x^{k+1}, \epsilon^{k+1})$$
$$\Delta G(x^{k+1}; x^k, \epsilon^k) := G(x^k; x^k, \epsilon^k) - G(x^{k+1}; x^k, \epsilon^k),$$

and use the shorthands $W^k := \operatorname{diag}\left(w_1^k, \ldots, w_n^k\right)$.

Proposition 4.1. *Suppose Assumption 4.1 holds. Let $\{(x^k, \epsilon^k)\}$ be the sequence generated by Algorithm 1 . It follows that $F(x, \epsilon)$ is monotonically decreasing over $\{(x^k, \epsilon^k)\}$ and*

$$(M - \frac{L_f}{2}) \sum_{t=0}^{k-1} \|x^{t+1} - x^t\|_2^2 \le F(x^0, \epsilon^0) - F(x^k, \epsilon^k). \tag{4.2}$$

Hence, $\lim_{k\to\infty} \|x^{k+1} - x^k\|_2 = 0.$

Proof. From the same argument as the proof for [24, Proposition 1], we have

$$\Delta F(x^{k+1}, \epsilon^{k+1}) \ge \Delta G(x^{k+1}; x^k, \epsilon^k) + \frac{M - L_f}{2}\|x^k - x^{k+1}\|_2^2. \tag{4.3}$$

Assumption 4.1 implies the subproblem solution x^{k+1} satisfies the optimality condition

$$\nabla Q_k(x^{k+1}) + \lambda W^k y^{k+1} + z^{k+1} = 0 \tag{4.4}$$

with $z^{k+1} \in N_\Gamma(x^{k+1})$ and $y^{k+1} \in \partial\|x^{k+1}\|_1$. Hence,

$$\begin{aligned}
&\Delta G(x^{k+1}; x^k, \epsilon^k) \\
&= Q_k(x^k) - Q_k(x^{k+1}) + \lambda \sum_{i=1}^n w_i^k(|x_i^k| - |x_i^{k+1}|) \\
&\geq \langle \nabla Q_k(x^{k+1}), x^k - x^{k+1}\rangle + \frac{M}{2}\|x^{k+1} - x^k\|_2^2 + \lambda \sum_{i=1}^n w_i^k y_i^{k+1}(x_i^k - x_i^{k+1}) \\
&\quad + z_i^{k+1}(x_i^k - x_i^{k+1}) \\
&= \langle \nabla Q_k(x^{k+1}) + \lambda W^k y^{k+1} + z^{k+1}, x^k - x^{k+1}\rangle + \frac{M}{2}\|x^{k+1} - x^k\|_2^2 \\
&= \frac{M}{2}\|x^{k+1} - x^k\|_2^2,
\end{aligned} \tag{4.5}$$

where the inequality is by Assumption 4.1, the convexity of $|\cdot|$ and the definition of normal cone, and the last equality is by (4.4). Combining (4.3) and (4.5), we have

$$\Delta F(x^{k+1}, \epsilon^{k+1}) \geq (M - \frac{L_f}{2})\|x^k - x^{k+1}\|_2^2. \tag{4.6}$$

Replacing k with t and summing up from $t = 0$ to $k - 1$, we have

$$\sum_{t=0}^{k-1}(F(x^t, \epsilon^t) - F(x^{t+1}, \epsilon^{t+1})) \geq (M - \frac{L_f}{2}) \sum_{t=0}^{k-1} \|x^t - x^{t+1}\|_2^2,$$

completing the proof of (4.2). □

Now we prove that every cluster point $\bar{x}$ of $\{x^k\}$ generated by Algorithm 1 satisfied the AKKT. As a result, it satisfies the first-order necessary optimality condition of $(\mathcal{P}_1)$ by Theorem 3.3 if the CCP condition holds at $\bar{x}$.

Theorem 4.2. *Suppose $\{x^k\}$ is the sequence generated by Algorithm 1 for solving $(\mathcal{P}_1)$. It holds that every cluster point of $\{x^k\}$ satisfies the AKKT condition.*

Proof. Let $\bar{x}$ be a cluster point of $\{x^k\}$ with subsequence $\{x^k\}_S \to \bar{x}$. By Proposition 4.1, $\{x^{k+1}\}_S \to \bar{x}$. From the optimality condition of the subproblem of Algorithm 1 , we have

$$\nabla_i f(x^k) + M(x_i^{k+1} - x_i^k) + \lambda p(|x_i^k| + \epsilon_i^k)^{p-1} sign(x_i^{k+1}) + z_i^{k+1} = 0, \quad \forall i \in \mathcal{N}^{k+1}, \tag{4.7}$$

where $z^{k+1} \in N_\Gamma(x^{k+1})$. For any $i \in \mathcal{N}^{k+1}$,

$$\begin{aligned}
&|\nabla_i f(x^{k+1}) + \lambda p|x_i^{k+1}|^{p-1} sign(x_i^{k+1}) + z_i^{k+1}| \\
&= |(\nabla_i f(x^{k+1}) - \nabla_i f(x^k)) + \lambda p sign(x_i^{k+1})(|x_i^{k+1}|^{p-1} - (|x_i^k| + \epsilon_i^k)^{p-1}) - M(x_i^{k+1} - x_i^k)| \\
&\leq (L_f - M)|x_i^{k+1} - x_i^k| + \lambda p(1-p)|\hat{x}_i^k|^{p-2}||x_i^{k+1}| - |x_i^k| - \epsilon_i^k|
\end{aligned}$$

where $\hat{x}_i^k$ is between $|x_i^{k+1}|^{p-1}$ and $(|x_i^k| + \epsilon_i^k)^{p-1}$. For sufficiently large k, $|\hat{x}_i^k|^{p-2} > \delta > 0$, $\mathcal{N}(\bar{x}) \subset \mathcal{N}(x^k)$ and $\mathcal{N}(\bar{x}) \subset \mathcal{N}(x^{k+1})$. Therefore, for any $i \in \mathcal{N}(\bar{x})$,

$$\begin{aligned}
&|\nabla_i f(x^{k+1}) + \lambda p|x_i^{k+1}|^{p-1} sign(x_i^{k+1}) + z_i^{k+1}| \\
&\leq (L_f - M)|x_i^{k+1} - x_i^k| + \lambda p(1-p)\delta^{p-2}|x_i^{k+1} - x_i^k| + \lambda p(1-p)\delta^{p-2}\epsilon_i^k \to 0
\end{aligned}$$

as $k \to \infty$. This completes the proof. □

5 Numerical Experiment

In this section, we design numerical experiments to demonstrate the performance of Algorithm 1 for solving the $(\mathcal{P}_1)$ with application in portfolio management, where f can include various loss functions such as the variance of portfolio or tracking error of index tracking. Here we use the commonly seen Markowitz mean-variance model to predict an optimal portfolio. Specifically, we only consider the shorting-prohibited Markowitz model and assume the optimal Lagrangian multiplier associated with the mean constraint is known as η which can be chosen accordingly to a reasonable expected return rate.

In the experiment, we collected historical daily stock price data to obtain R and μ in $S\&P$ 500 index from Yahoo finance, which spans from 01/01/2013 to 31/12/2013. We do not include any company unless it is traded on the market at least 90% of the trading days during the data period, nor do any company not listed on the market for the entire timescale. The total list has 471 companies by 251 trading days. We recast the Markowitz model with no-shorting constraint as a linear equality constrained optimization problem with $\ell_{1/2}$ regularizer,

$$\begin{aligned} \min \quad & \frac{1}{2}x^T Rx - \eta\mu^T x + \lambda\|x\|_{1/2}^{1/2} \\ \text{s.t.} \quad & e^T x = 1, \\ & x \geq 0 \end{aligned} \tag{5.1}$$

where $R \in \mathbb{R}^{n\times n}$ is the estimated covariance matrix of the portfolio, and $\eta\mu$ is the estimated return vector with $\eta > 0$ and $\mu \in \mathbb{R}^n$. We test over a grid of regularization parameter values and choose the best μ and λ corresponding to a solution that performs well in both in-sample and out-of-sample Sharpe ratios. By doing this, we choose $\eta = 0.001$ and $\lambda = 0.01$ in the model.

In Algorithm 1, we use $Q_k(x) = (Rx^k - \eta\mu)^T x + \frac{\beta}{2}\|x - x^k\|_2^2$. The parameters are selected as $\alpha = 0.998$, $\beta = 1.1L_f$, $\epsilon^0 = 0.001$ and the initial point is $x^0 = \frac{1}{n}e$ so that it is feasible. As mentioned in the last section, AKKT conditions are satisfied at the clustering point and the ECCP holds true by Proposition 3.4(c) since Γ here is a closed and convex set. Therefore, from Theorem 3.3(a), the KKT condition of (5.1) is satisfied at the clustering point generated by the algorithm.

After solving each subproblem, we obtain a primal feasible iterate x^{k+1} and dual iterate ν^k satisfying the optimality condition for the subproblem

$$(R_i x^k - \eta\mu)^T x^{k+1} + \beta(x_i^{k+1} - x_i^k) + \lambda w_i^k + \nu^k = 0,\ i \in \mathcal{N}(x^{k+1}), \tag{5.2}$$

where R_i is the ith row of R. For any primal-dual pair (x, ν) with $x \geq 0$, we can define the following metric to measure the optimality residual at (x, ν)

$$\alpha(x,\nu) := \sum_{i\in\mathcal{N}(x)} |x^T R_i - \eta\mu_i + \lambda p x_i^{p-1} + \nu|, \tag{5.3}$$

where ν is the dual variable associated with the simplex constraint.

We plot the evolution of $\alpha(x^k, \nu^k)$ over iterations in Figure 2, which steadily decreases to 0. Figure 3 shows the number of nonzero components of the portfolio versus the regularization parameter λ by fixing $\eta = 0.001$. We can see that larger λ can yield sparser solutions.

finance.yahoo.com

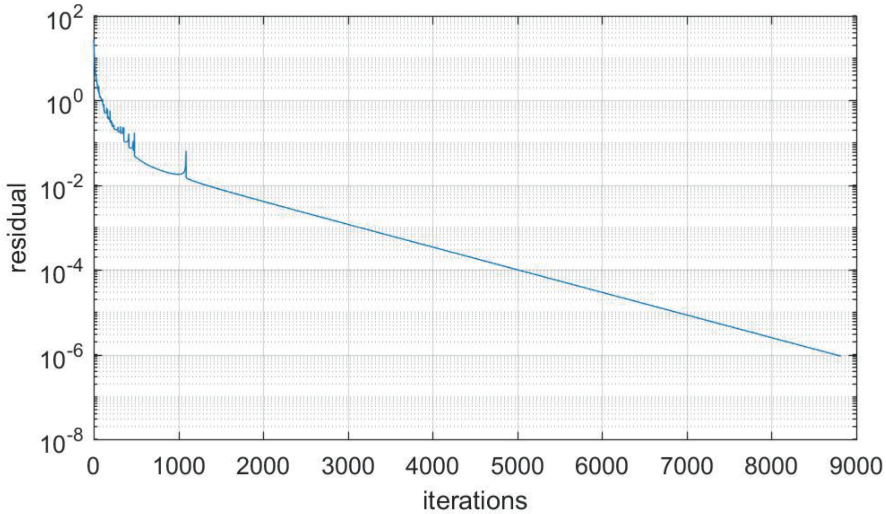

Figure 2: The residual $\alpha(x^k, \nu^k)$ generated by Algorithm 1

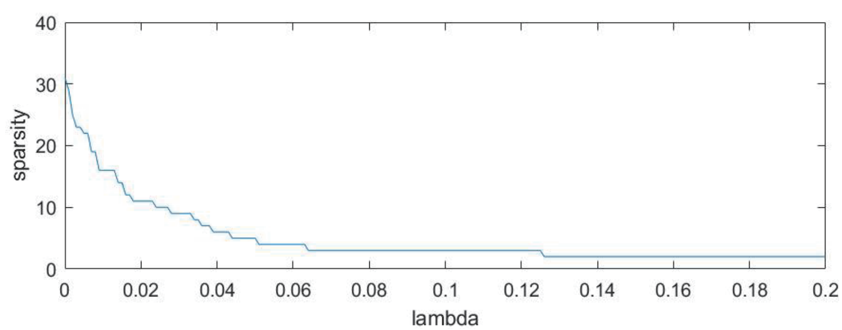

Figure 3: Portfolio sparsity (nonzero components) for different λ

For out-of-sample testing, we collected historical daily stock price data in $S\&P$ 500 index from Yahoo finance, which spans from 01/01/2014 to 31/03/2014.

Figure 4 shows the Sharpe ratios of our ℓ_p-norm regularized portfolio. The sparse portfolios are more implementable due to the transaction costs or physical limitations reasons. Our results indicate that an intermediate sparsity (around 10) portfolio may have the best Sharpe ratio in both in-sample or out-of-sample performance.

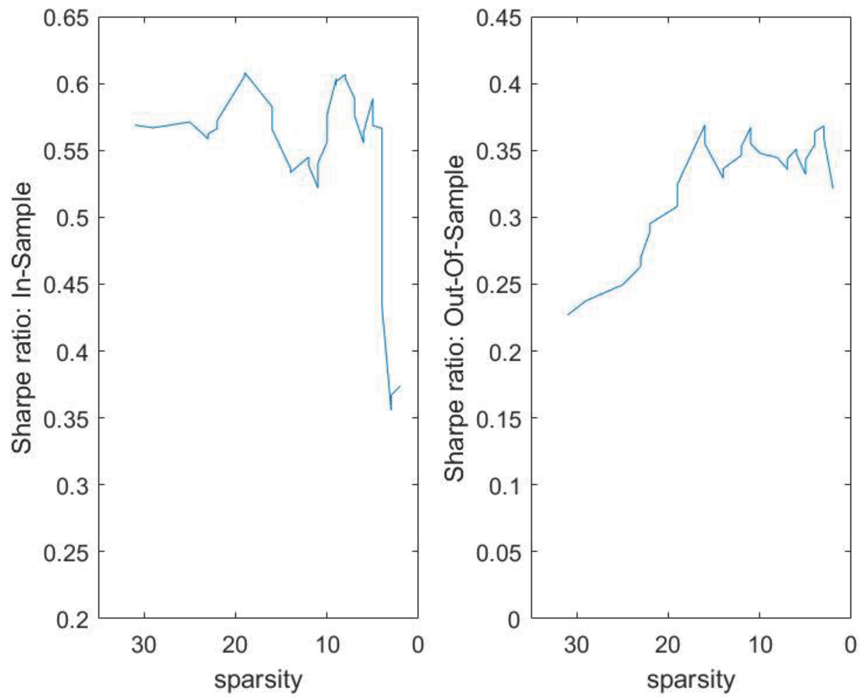

Figure 4: Portfolio Sharpe ratios for different sparsity: in-sample and out-of-sample

We also compare the performance of the proposed algorithm with different type of existing method, which is the Successive Difference of Convex Approximation Method (SDCAM) in [14]. This method makes use of the Moreau envelope as a smoothing technique for the nonsmooth terms and the implementation details are provided in the appendix. With $\lambda = 0.01$, Figure 5 depicts the evolution of objective values versus CPU time for both algorithms, which implies that Algorithm 1 can converge more rapidly compared with SDCAM.

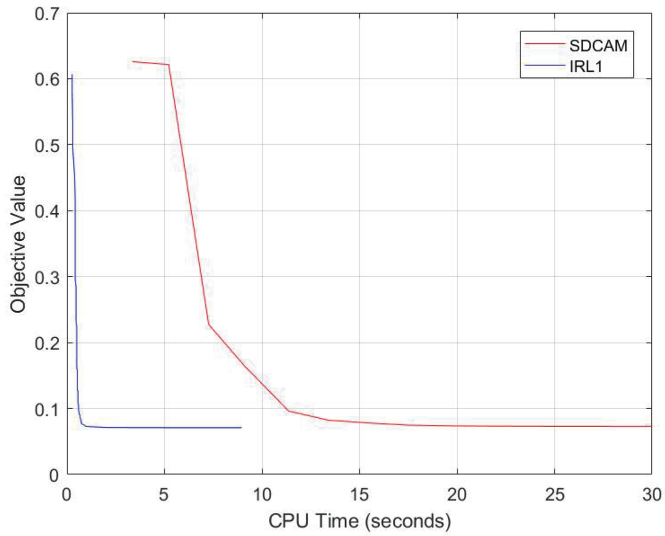

Figure 5: Evolution of objective versus CPU time

Acknowledgments

Hao Wang was supported by the Natural Science Foundation of Shanghai under Grant 21ZR1442800 and the National Natural Science Foundation of China No. 12001367.

A Implementation Details of SDCAM

In this section, we provide the details of the implementation of SDCAM [14]. We can write problem (5.1) as

$$\min_{x\in C} \quad F(x) := f(x) + v(x),$$

where $f(x) = \frac{1}{2}x^T Rx - \eta\mu^T x$, $v(x) = \lambda\|x\|_{1/2}^{1/2}$ and $C = \{x \mid e^T x = 1, x \geq 0\}$. Then this problem can be solved by the SDCAM [14], which approximates F by its Moreau envelope at each iteration. At the kth iterate, this method solves the following approximate problem

$$\min_{x\in C} \quad F_{\lambda_k}(x) := f(x) + e_{\lambda_k} v(x).$$

Here $e_{\lambda_k} v(x)$ is the Moreau envelope of $v(x)$ with parameter λ_k, which takes the form

$$e_{\lambda_k} v(x) =:= \inf_{\mathbf{x}}\{\frac{1}{2\lambda_k}\|x-y\|_2^2 + v(y)\}.$$

The SDCAM then drives the parameter λ_k to 0 and solves each corresponding subproblem F_{λ_k} iteratively. We can reformulate F_{λ_k} as a DC problem by taking advantage of the equivalently formulation of $e_{\lambda_k} v(x)$, i.e.,

$$e_{\lambda_k} v(x) = \frac{1}{2\lambda}\|x\|_2^2 - \underbrace{\sup_{y\in\text{dom } v}\{\frac{1}{\lambda}x^T y - \frac{1}{2\lambda}\|y\|_2^2 - v(y)\}}_{D_{\lambda,v}(x)}.$$

This DC subproblem is solved by the Nonmonotone Proximal Gradient method with majorization.

References

[1] A.A. Ababneh, Density-based sensor selection for RSS target localization, *IEEE Sens. J.* 18 (2018) 8532–8540.

[2] R. Andreani, J.M. MartÍnez, A. Ramos and P.J. Silva, A cone-continuity constraint qualification and algorithmic consequences, *SIAM J. Optim.* 26 (2016) 96–110.

[3] W. Bian and X. Chen, Optimality and complexity for constrained optimization problems with nonconvex regularization, *Math. Oper. Res.* 42 (2017) 1063–1084,

[4] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, *IEEE Signal Process. Lett.* 14 (2007) 707–710.

[5] R. Chartrand and W. Yin, Iteratively reweighted algorithms for compressive sensing, in: *Proceedings of Int. Conf. on Acoustics, Speech, Signal Processing (ICASSP)*, IEEE, 2008, pp. 3869–3872.

[6] C. Chen, X. Li, C. Tolman, S. Wang and Y. Ye, Sparse portfolio selection via quasi-norm regularization, *arXiv preprint arXiv:1312.6350*, 2013.

[7] X. Chen, F. Xu and Y. Ye, Lower bound theory of nonzero entries in solutions of $\ell_2 - \ell_p$ minimization, *SIAM J. Sci. Comput.* 32 (2010), 2832–2852.

[8] X. Chen and W. Zhou, Convergence of reweighted ℓ_1 minimization algorithms and unique solution of truncated ℓ_p minimization, *Department of Applied Mathematics, The Hong Kong Polytechnic University*, 2010.

[9] D. Ge, X. Jiang and Y. Ye, A note on the complexity of ℓ_p minimization, *Math. Program.* 129 (2011) 285–299.

[10] G. Haeser, H. Liu and Y. Ye, Optimality condition and complexity analysis for linearly-constrained optimization without differentiability on the boundary. *Math. Program.* 178 (2019) 263–299.

[11] M.-J. Lai and J. Wang, An unconstrained ℓ_q minimization with $0 \leq q \leq 1$ for sparse solution of underdetermined linear systems, *SIAM J. Optim.* 21 (2011) 82–101.

[12] Z. Lin, Z. Huijuan, M. Wenjuan, J. Jingying, Z. Limin, L. Jiao, G. Feng and Z. Zhongxing, Resolution and noise performance of sparse view X-ray CT reconstruction via ℓ_p-norm regularization, *Phys. Med.* 52 (2018) 72–80.

[13] Q. Ling, Z. Wen and W. Yin, Decentralized jointly sparse optimization by reweighted minimization, *IEEE Trans. Image Process.* 61 (2013) 1165–1170.

[14] T. Liu, T.K. Pong and A. Takeda, A successive difference-of-convex approximation method for a class of nonconvex nonsmooth optimization problems, *Math. Program.* 176 (2019) 339–367.

[15] Z. Lu, Iterative reweighted minimization methods for ℓ_p regularized unconstrained nonlinear programming, *Math. Program.* 147 (2014) 277–307.

[16] S. Oymak, B. Recht and M. Soltanolkotabi, Sharp time–data tradeoffs for linear inverse problems. *IEEE Trans. Inform. Theory* 64 (2017) 4129–4158.

[17] R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis*, vol. 317, Springer Science & Business Media, 2009.

[18] Y. Shi, J. Zhang and K. B. Letaief, Group sparse beamforming for green Cloud-RAN, *IEEE Trans. Wirel. Commun.* 13 (2914) 2809–2823.

[19] M.V. Solodov, *Constraint Qualifications*, Wiley Encyclopedia of Operations Research and Management Science, 2010.

[20] M. Thom, M. Rapp and G. Palm, Efficient dictionary learning with sparseness-enforcing projections, *Internat. J. Comput. Vision* 114 (2015) 168–194.

[21] H. Wang, D.-H. Li, X.-J. Zhang and L. Wu, Optimality conditions for the constrained ℓ_p-regularization, *Optimization* 64 (2015) 2183–2197.

[22] H. Wang, H. Zeng and J. Wang, Convergence rate analysis of proximal iteratively reweighted ℓ_1 methods for ℓ_p regularization problems, *Optim. Lett.* 17 (2022) 413–435.

[23] H. Wang, H. Zeng and J. Wang, An extrapolated iteratively reweighted ℓ_1 method with complexity analysis, *Comput. Optim. Appl.* 83 (2022) 967–997.

[24] H. Wang, H. Zeng, J. Wang and Q. Wu, Relating ℓ_p regularization and reweighted ℓ_1 regularization, *Optim. Lett.* 15 (2021) 2639–2660.

[25] H. Wang, F. Zhang, Y. Shi and Y. Hu, Nonconvex and nonsmooth sparse optimization via adaptively iterative reweighted methods. *J. Glob. Optim.* 81 (2021) 717–748.

[26] P. Xu, Y. Tian, H. Chen and D. Yao, ℓ_p norm iterative sparse solution for EEG source localization, *IEEE. Trans. Biomed. Eng.* 54 (2007) 400–409.

[27] Z.B. Xu, Z. Hai, W. Yao, X.Y. Chang and L. Yong, $\ell_{1/2}$ regularization, *Sci. China Ser. F* 53 (2010) 1159–1169.

[28] X. Yang, J. Wang and H. Wang, Towards an efficient approach for the nonconvex ℓ_p-ball projection: algorithm and analysis, *J. Mach. Learn. Res.* 23 (2022) 1–31.

Manuscript received 15 May 2022
revised 14 October 2022
accepted for publication 10 November 2022

Hao Wang
School of Information Science and Technology,
ShanghaiTech University, Shanghai, China
E-mail address: wanghao1@shanghaitech.edu.cn

Yining Gao
School of Information Science and Technology,
ShanghaiTech University, Shanghai, China
E-mail address: gaoyn@shanghaitech.edu.cn

Jiashan Wang
Department of Mathematics,
University of Washington, Seattle, WA, USA
E-mail address: jsw1119@gmail.com

Hongying Liu
School of Mathematical Sciences,
Beihang University, Beijing, China
E-mail address: liuhongying@buaa.edu.cn