

Convergence of the Safeguarded Augmented Lagrangian Method under the Polyak-Łojasiewicz constraint qualification for Constrained Composite Optimization

Christian Kanzow and Jannis Krüger

June 22, 2026

Abstract

In this work we provide theoretical and practical results of the Safeguarded Augmented Lagrangian Method (SALM) for constrained composite optimization problems whose objective is the sum of a smooth and a nonsmooth function. We obtain global convergence results to an M-stationary point for SALM under the Polyak-Łojasiewicz constraint qualification (PLCQ). For this result the boundedness of the Lagrange multipliers is crucial, and this is shown under the assumption that the nonsmooth part of the objective function is locally Lipschitz continuous. A counterexample shows that one cannot expect to get bounded multipliers without such an assumption. The performance of the algorithm is evaluated numerically on a set of sparse portfolio optimization problems with two different regularization terms, one being Lipschitz and the other one being non-Lipschitz. The results are significantly better for the Lipschitz sparsity term, whereas the underlying method generates seemingly unbounded multipliers in many instances when using the non-Lipschitz sparsity function.

Keywords. Safeguarded augmented Lagrangian method, global convergence, M-stationarity, Polyak-Łojasiewicz constraint qualification, locally Lipschitz regularization function, composite optimization.

1 Introduction

We consider constrained optimization problems of the form

$$\min_{x \in \mathbb{R}^n} f(x) + \varphi(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0 \quad (1.1)$$

under the following assumptions:

Assumption 1.1. (i) *The functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ are continuously differentiable.*

(ii) *φ is a locally Lipschitz continuous function.*

Problems of this kind arise in a wide range of topics such as engineering, machine learning, optimal control or signal processing, with tasks like image restoration [6], low-rank matrix completion [8], compressed sensing [41], clustering [39], principal component analysis [29] and model predictive control [13].

The theoretical difficulty for solving problems of such composite nature is the non-smooth function in the objective, as this requires an extension of differentiability and stationarity concepts that may lack the nice properties of smooth problems.

In this paper we consider the Safeguarded Augmented Lagrangian Method (SALM) as a suitable method for numerically solving composite constrained problems. Building on the classical Augmented Lagrangian Method by Powell [35], Hestenes [22], and Rockafellar [36], the safeguarded version introduced for the ALGENCAN solver [1] is a powerful and versatile choice for smooth constrained optimization due to its algorithmic simplicity and extensive global convergence theory, see also [7].

As Augmented Lagrangian methods require solving an unconstrained minimization problem in each step, a practically useful extension of SALM to composite problems necessitates the existence of subproblem solvers which can at least find subdifferentially stationary points of unconstrained composite problems satisfying Assumption 1.1. More recently, suitable results for the Proximal Gradient Method [24] and Proximal Quasi Newton Methods [17, 12] have been established. This gave the opportunity to develop a global convergence theory for SALM and constrained composite problems in [16], using even a more general problem setting via geometric constraints, and only requiring φ to be proper and lower semicontinuous (lsc). In the latter paper, the authors use AM-regularity as a constraint qualification in order to show subsequential convergence of the SALM sequence to a M-stationary point for the more general problem.

We want to follow a different approach here. As we consider a more specialized setting with explicit equality and inequality constraints, we are looking to use the least strict constraint qualification tailored to this constraint setting that also still ensures a global convergence result of the iterates. For this approach, we mainly follow the considerations of Andreani et al. in [2]. In that paper, the authors introduce a constraint qualification (PŁCQ) using a Polyak-Łojasiewicz inequality, show that it is in fact equivalent to an Error Bound Condition (EB), and also show that this is the least restrictive constraint qualification for smooth problems that still ensures boundedness of the multiplier sequences, which is in return pivotal in establishing convergence of the SALM iterates to an M-stationary point.

In view of these results cited above, φ being locally Lipschitz continuous may seem like a quite restrictive assumption for the regularization term. However, this restriction is necessary for the required multiplier boundedness, as illustrated in Section 4.1.

The rest of this work is organized as follows: Section 2 summarizes some basic facts from variational analysis. In Section 3, we extend the global convergence theory of SALM under the PŁCQ from [2] to our composite optimization setting (1.1) under Assumption 1.1. We then provide a counterexample that demonstrates the necessity of the local Lipschitz assumption for φ in Section 4. We present some numerical results of SALM applied to sparse portfolio optimization comparing Lipschitz and non-Lipschitz regularization functions in Section 5, and conclude the paper with some final remarks in Section 6.

Notation: Let Ω denote the feasible set of the optimization problem (1.1). The active

set for a feasible point is written as

$$I(x^*) := \{i \in \{1, \dots, m\} : g_i(x^*) = 0\}.$$

The sets of non-negative and positive real numbers are denoted by \mathbb{R}_+ and \mathbb{R}_{++} , respectively. For $v \in \mathbb{R}^n$ we write

$$v_+ := (\max\{0, v_1\}, \dots, \max\{0, v_n\})^T$$

for the projection of v onto the nonnegative orthant \mathbb{R}_+^n . The symbol $\|\cdot\|$ always denotes the Euclidean norm, and $B_r(x)$ and $\overline{B}_r(x)$ the corresponding open and closed balls with radius r centered at x , respectively. Finally, $\text{dist}(x, S)$ denotes the Euclidean distance $\inf_{w \in S} \|x - w\|$ from $x \in \mathbb{R}^n$ to $S \subseteq \mathbb{R}^n$.

2 Mathematical Background

The set of extended real numbers is denoted by $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. We write $x \xrightarrow{f} x^*$ if $x \rightarrow x^*$ and $f(x) \rightarrow f(x^*)$, also called *f-attentive convergence*. For $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ we write $\text{dom}(f) := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$ and say that f is proper, if $\text{dom}(f) \neq \emptyset$ and $f(x) > -\infty$ for all $x \in \mathbb{R}^n$.

For our convergence analysis we recall some definitions and classical results from nonsmooth analysis as presented in the standard sources [30] and [37].

Definition 2.1. Let $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be proper, lsc and $x \in \text{dom}(f)$.

(a) The set

$$\hat{\partial}f(x) := \left\{ v \in \mathbb{R}^n \mid \liminf_{x \rightarrow \bar{x}} \frac{f(x) - f(\bar{x}) - v^T(x - \bar{x})}{\|x - \bar{x}\|} \geq 0 \right\}$$

is called the Fréchet subdifferential of f at x .

(b) The set

$$\partial f(x) := \{v \in \mathbb{R}^n \mid \exists \{x^k\}, \{v^k\} \subseteq \mathbb{R}^n : x^k \xrightarrow{f} x, v^k \rightarrow v, v^k \in \hat{\partial}f(x^k) \forall k \in \mathbb{N}\}$$

is called the limiting/Mordukhovich subdifferential of f at x .

By the definitions of the subdifferentials it is apparent that

$$\hat{\partial}f(x) \subseteq \partial f(x) \tag{2.1}$$

holds for all $x \in \text{dom}(f)$.

These subdifferentials admit a generalized Fermat rule [30, Proposition 1.30 (i)], which leads to a natural extension of stationarity for an unconstrained non-smooth setting.

Lemma 2.2 (Subdifferential Fermat rule). Let f be proper and lsc and let f have a local minimum in x^* . Then

$$0 \in \hat{\partial}f(x^*) \tag{2.2}$$

and

$$0 \in \partial f(x^*). \tag{2.3}$$

Points satisfying these conditions are called *F- or S-stationary* (2.2) (S for strong) and *M-stationary* (2.3) respectively. We can relax the stationarity condition by introducing ε -stationarity. Here, we only introduce this terminology for the Mordukhovich subdifferential.

Definition 2.3. *Let $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be proper and lsc. A vector $x^* \in \text{dom}(f)$ is called an ε -stationary point, if there exists an $\eta \in \partial f(x^*)$ with $\|\eta\| \leq \varepsilon$.*

In the following we will list the properties of the Mordukhovich subdifferential used in the convergence analysis. The first result is the robustness of the Mordukhovich subdifferential, which serves as a non-smooth substitute for gradient continuity.

Theorem 2.4 (Robustness of the Limiting Subdifferential). *Let $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be proper, lsc, and $\bar{x} \in \text{dom}(f)$ be given. Then*

$$\partial f(\bar{x}) = \text{Limsup}_{x \rightarrow^f \bar{x}} \partial f(x) := \{v \in \mathbb{R}^n \mid \exists \{x^k\} \rightarrow^f \bar{x}, \exists \{v^k\} \rightarrow v : v^k \in \partial f(x^k) \forall k \in \mathbb{N}\}.$$

The next important result is the sum rule, which essentially allows us to split our composite problem and deal with the differentiable and non-differentiable part separately.

Theorem 2.5 (Sum Rule). *Let $f, g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be proper, lsc, and let $\bar{x} \in \text{dom}(f) \cap \text{dom}(g)$ be given. Then*

- (a) *If f is differentiable in \bar{x} , then $\hat{\partial}(f + g)(\bar{x}) = \nabla f(\bar{x}) + \hat{\partial}g(\bar{x})$.*
- (b) *If f is differentiable in a neighborhood of \bar{x} and continuously differentiable in \bar{x} , then $\partial(f + g)(\bar{x}) = \nabla f(\bar{x}) + \partial g(\bar{x})$.*

If we have local Lipschitz continuity around a given point, the Mordukhovich subdifferential has the following useful boundedness property, cf. [37, Theorem 9.13].

Proposition 2.6. *Let $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be locally Lipschitz around some point $\bar{x} \in \text{dom}(f)$ with constant $L = L(\bar{x}) > 0$. Then $\|v\| \leq L$ for all $v \in \partial f(\bar{x})$.*

3 The Safeguarded Augmented Lagrangian Method

Recall that the *Lagrangian function* of the optimization problem (1.1) is given by

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \varphi(x) + \lambda^T g(x) + \mu^T h(x),$$

whereas the (Hestenes-Powell-Rockafellar) *augmented Lagrangian function* with penalty parameter $\rho > 0$ is defined by

$$\mathcal{L}_\rho(x, \lambda, \mu) = f(x) + \varphi(x) + \frac{\rho}{2} \left(\left\| \left(g(x) + \frac{\lambda}{\rho} \right)_+ \right\|^2 + \left\| h(x) + \frac{\mu}{\rho} \right\|^2 \right)$$

for some $x \in \mathbb{R}^n$ and certain vectors $\lambda \in \mathbb{R}_+^m$, $\mu \in \mathbb{R}^p$, usually called the *Lagrange multipliers*.

We define the Safeguarded Augmented Lagrangian Method (SALM) similar to [2] as follows.

Algorithm 3.1. (*Safeguarded Augmented Lagrangian Method*)

(S.0) Let $u_{\max} > 0$, $v_{\min} < v_{\max}$, $\{u_k\} \subseteq [0, u_{\max}]^m$ and $\{v_k\} \subseteq [v_{\min}, v_{\max}]^p$ be bounded sequences. Let $\{\varepsilon_k\} \subseteq \mathbb{R}_{++}$ with $\varepsilon_k \rightarrow 0$. Choose a starting point $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^p$ and a $\rho_0 > 0$ as well as parameters $\tau \in (0, 1)$, $\gamma > 1$ and set $k := 0$.

(S.1) If a suitable termination criterion holds for (x^k, λ^k, μ^k) : STOP.

(S.2) Compute x^{k+1} as an ε_k -stationary point of

$$\min_{x \in \mathbb{R}^n} \mathcal{L}_{\rho_k}(x, u^k, v^k).$$

(S.3) Set

$$\begin{aligned} \lambda_i^{k+1} &:= (u_i^k + \rho_k g_i(x^{k+1}))_+ \quad \forall i = 1, \dots, m, \\ \mu_j^{k+1} &:= v_j^k + \rho_k h_j(x^{k+1}) \quad \forall j = 1, \dots, p. \end{aligned} \quad (3.1)$$

(S.4) If either $k = 0$ or the two conditions

$$\begin{aligned} \|h(x^{k+1})\| &\leq \tau \|h(x^k)\| \quad \text{and} \\ \left\| \min \left\{ -g(x^{k+1}), \frac{u^k}{\rho_k} \right\} \right\| &\leq \tau \left\| \min \left\{ -g(x^k), \frac{u^{k-1}}{\rho_{k-1}} \right\} \right\| \end{aligned} \quad (3.2)$$

hold, set $\rho_{k+1} := \rho_k$, otherwise set $\rho_{k+1} := \gamma \rho_k$.

(S.5) Set $k \leftarrow k + 1$ and go to (S.1)

The safeguarding mechanism, which is to use the bounded sequences $\{u^k\}$, $\{v^k\}$ in the augmented Lagrangian for unconstrained subproblems, yields better global convergence properties, as discussed in [25].

The multiplier update in (3.1) is done such that the following relation between Lagrangian and augmented Lagrangian holds in each step.

Lemma 3.2. *Let $k \in \mathbb{N}$. It holds*

$$\partial_x \mathcal{L}(x^k, \lambda^k, \mu^k) = \partial_x \mathcal{L}_{\rho_{k-1}}(x^k, u^{k-1}, v^{k-1})$$

for the iterates of the safeguarded augmented Lagrangian method.

Proof. By the sum rule for the Mordukhovich subdifferential and the continuous differentiability of f, g, h , we have

$$\begin{aligned} \partial_x^M \mathcal{L}_{\rho_{k-1}}(x^k, u^{k-1}, v^{k-1}) &= \nabla f(x^k) + \partial_x \varphi(x^k) \\ &\quad + \sum_{i=1}^m (u_i^{k-1} + \rho_{k-1} g_i(x^k))_+ \nabla g_i(x^k) \\ &\quad + \sum_{j=1}^p (v_j^{k-1} + \rho_{k-1} h_j(x^k)) \nabla h_j(x^k) \\ &= \nabla f(x^k) + \partial_x \varphi(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla h_j(x^k) \\ &= \partial_x \mathcal{L}(x^k, \lambda^k, \mu^k), \end{aligned}$$

where the penultimate identity follows from the definitions of the updated multipliers λ_i^k and μ_j^k in Algorithm 3.1. \square

In smooth constrained minimization problems, we are aiming for convergence of our algorithm to a point satisfying the KKT conditions. These conditions are necessary for local minima under a suitable constraint qualification, and are also algorithmically verifiable in a reasonable way. For our potentially non-smooth setting, the counterpart of the KKT conditions is the notion of M-stationarity that is applicable to our case. We do this by adapting the generalized definition of composite M-stationarity from [16] to our explicit constraint setting, which leads to the following definition.

Definition 3.3. *The triple $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^p$ is called an M-stationary point of problem (1.1) if the following conditions hold:*

$$\begin{aligned} 0 &\in \nabla f(x^*) + \partial\varphi(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^*) \quad (\text{stationarity}), \\ h(x^*) &= 0 \quad (\text{equality feasibility}), \\ g(x^*) &\leq 0 \quad (\text{inequality feasibility}), \\ (\lambda^*)^T g(x^*) &= 0 \quad (\text{complementarity}). \end{aligned}$$

In our subsequent convergence analysis, we only consider the case that the SALM Algorithm runs indefinitely and produces an infinite series of iterates. In practice, we want to employ a termination criterion in (S.1) that checks for (approximate) stationarity, for more details we refer to Section 5.

Ensuring the boundedness of the sequences generated by SALM is already sufficient for convergence to an M-stationary point on a subsequence, as stated in the following theorem.

Theorem 3.4. *Let $\{(x^k, \lambda^k, \mu^k)\}$ be an infinite sequence generated by Algorithm 3.1. Then every accumulation point of this sequence is an M-stationary point of the program (1.1).*

Proof. Let (x^*, λ, μ) be an accumulation point of the sequence $\{(x^k, \lambda^k, \mu^k)\}$, and let $\{(x^k, \lambda^k, \mu^k)\}_K$ be a subsequence converging to this accumulation point. We proceed by showing that the M-stationarity conditions hold in (x^*, λ, μ) .

To this end, we begin with the feasibility of x^* . Taking the multiplier update (3.1) and rearranging, we obtain

$$\rho_{k-1} g_i(x^k)_+ \leq (u_i^{k-1} + \rho_{k-1} g_i(x^k))_+ = \lambda_i^k$$

for $i = 1, \dots, m$ and

$$\begin{aligned} \rho_{k-1} |h_j(x^k)| &= \text{sign}(h_j(x^k)) \rho_{k-1} h_j(x^k) \\ &= \text{sign}(h_j(x^k)) (\mu_j^k - v_j^{k-1}) \\ &\leq |\mu_j^k| + \max\{|v_{\min}|, |v_{\max}|\} \end{aligned}$$

for $j = 1, \dots, p$. By the boundedness of the multiplier sequences on K , the right-hand sides of the respective equations are bounded for $k \in K$, hence there exists a constant $c > 0$ such that

$$\rho_{k-1} g_i(x^k)_+ \leq c, \quad \rho_{k-1} |h_j(x^k)| \leq c \quad (3.3)$$

for all $i = 1, \dots, m$, $j = 1, \dots, p$, $k \in K$.

We first assume that the sequence $\{\rho_k\}$ is unbounded. Due to the updating of the penalty parameter, this already implies $\{\rho_k\} \rightarrow \infty$. In particular, we therefore have $\lim_{k \in K} \rho_{k-1} = \infty$. Then (3.3) yields

$$\lim_{k \in K} g_i(x^k)_+ = 0 \quad \text{and} \quad \lim_{k \in K} |h_j(x^k)| = 0$$

for all $i = 1, \dots, m$ and $j = 1, \dots, p$. Since $\lim_{k \in K} x^k = x^*$ and g, h are continuous, we therefore obtain $g(x^*) \leq 0$ and $h(x^*) = 0$, which immediately shows feasibility of x^* .

On the other hand, if $\{\rho_k\}$ is bounded, this sequence is eventually constant. Thus the condition (3.2) cannot be violated infinitely many times, and in infinitely many iterations we have a decrease of at least $\tau \in (0, 1)$. This immediately yields $\lim_{k \rightarrow \infty} \|h(x^k)\| = 0$ and thus $h(x^*) = 0$. Similarly, we obtain

$$\lim_{k \rightarrow \infty} \min \left\{ -g(x^k), \frac{u^{k-1}}{\rho_{k-1}} \right\} = 0. \quad (3.4)$$

Assume there exists an index $i \in \{1, \dots, m\}$ with $g_i(x^*) > 0$. Then it holds

$$\min \left\{ -g_i(x^k), \frac{u_i^{k-1}}{\rho_{k-1}} \right\} \leq -g_i(x^k) \leq -\frac{1}{2}g_i(x^*) < 0$$

for k sufficiently large, which produces a contradiction by taking the limit $k \rightarrow \infty$. Thus we must have $g_i(x^*) \leq 0$ for all $i = 1, \dots, m$, and feasibility of x^* holds also in this case.

We next verify the stationarity of the limit (x^*, λ, μ) . In Step 2 of the SALM, we compute x^k as an ε_k -stationary point of the augmented Lagrangian, hence there exists an $\eta_k \in \partial_x \mathcal{L}_{\rho_{k-1}}(x^k, u^{k-1}, v^{k-1})$ with $\|\eta_k\| \leq \varepsilon_k$. By Lemma 3.2 we have $\eta_k \in \partial_x \mathcal{L}(x^k, \lambda^k, \mu^k)$. As all functions except φ in the Lagrangian function are continuously differentiable, this is equivalent to the existence of a $z^k \in \partial_x \varphi(x^k)$ with

$$\left\| \nabla f(x^k) + z^k + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla h_j(x^k) \right\| \leq \varepsilon_k.$$

We now define

$$E^k := \nabla f(x^k) + z^k + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla h_j(x^k) \quad (3.5)$$

and $z^* := -\nabla f(x^*) - \sum_{i=1}^m \lambda_i \nabla g_i(x^*) - \sum_{j=1}^p \mu_j \nabla h_j(x^*)$. Then

$$\begin{aligned}
\|z^k - z^*\| &= \left\| z^k + \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla h_j(x^k) \right. \\
&\quad \left. - \nabla f(x^k) - \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) - \sum_{j=1}^p \mu_j^k \nabla h_j(x^k) \right. \\
&\quad \left. + \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{j=1}^p \mu_j \nabla h_j(x^*) \right\| \\
&\leq \|E^k\| + \|\nabla f(x^*) - \nabla f(x^k)\| \\
&\quad + \left\| \sum_{i=1}^m \lambda_i \nabla g_i(x^*) - \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) \right\| + \left\| \sum_{j=1}^p \mu_j \nabla h_j(x^*) - \sum_{j=1}^p \mu_j^k \nabla h_j(x^k) \right\| \\
&\leq \varepsilon_k + \|\nabla f(x^*) - \nabla f(x^k)\| \\
&\quad + \left\| \sum_{i=1}^m \lambda_i \nabla g_i(x^*) - \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) \right\| + \left\| \sum_{j=1}^p \mu_j \nabla h_j(x^*) - \sum_{j=1}^p \mu_j^k \nabla h_j(x^k) \right\|
\end{aligned}$$

Taking the limit $k \rightarrow \infty$, $k \in K$ and using $\varepsilon_k \searrow 0$ as well as the continuous differentiability of f, g, h , the right-hand side vanishes, and we have $\lim_{k \rightarrow \infty, k \in K} z^k = z^*$. By the robustness of the Mordukhovich subdifferential, this shows $z^* \in \partial\varphi(x^*)$ and thus

$$0 \in \nabla f(x^*) + \partial\varphi(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{j=1}^p \mu_j \nabla h_j(x^*),$$

so that the triple (x^*, λ, μ) satisfies the stationarity condition.

We finally establish satisfaction of the complementarity property. To this end, consider an arbitrary index $i \in \{1, \dots, m\}$. If $g_i(x^*) = 0$, we are already finished. Hence, let $g_i(x^*) < 0$. If $\lim_{k \rightarrow \infty} \rho_{k-1} = \infty$, we obtain by the boundedness of $\{u^k\}$

$$\lim_{k \in K} u_i^{k-1} + \rho_{k-1} g_i(x^k) = -\infty.$$

In particular, for $k \in K$ large enough, we have

$$\lambda_i^k = (u_i^{k-1} + \rho_{k-1} g_i(x^k))_+ = 0,$$

and thus $\lambda_i = 0$ in the limit $k \in K$. We now assume that ρ_k is bounded, hence, eventually constant. By the same reasoning that led to (3.4) and $g_i(x^*) < 0$, we must have

$$\min \left\{ -g_i(x^k), \frac{u_i^{k-1}}{\rho_{k-1}} \right\} = \frac{u_i^{k-1}}{\rho_{k-1}}$$

for sufficiently large $k \in K$. As $\{\rho_{k-1}\}$ is eventually constant, taking the limit on K yields $\lim_{k \in K} u_i^{k-1} = 0$. For all $k \in K$ sufficiently large, we therefore have

$$u_i^{k-1} + \rho_{k-1} g_i(x^k) < 0$$

and thus by (3.1)

$$\lambda_i^k = (u_i^{k-1} + \rho_{k-1} g_i(x^k))_+ = 0.$$

In the limit $k \in K$, we obtain $\lambda_i = 0$. Hence, for all $i = 1, \dots, m$, we have shown $\lambda_i g_i(x^*) = 0$. \square

Note that in the proof we did not have to rely on the Lipschitz-continuity of φ , hence this result is also valid if φ is merely continuous. However, we need the Lipschitz property of φ to establish sufficient conditions for multiplier boundedness, and this is also necessary in the sense that applications of SALM on non-Lipschitz functions can lead to unbounded multipliers, as discussed in Section 4.1.

We can ensure multiplier boundedness if certain constraint qualifications hold in the limit point x^* , which are taken from [2] and defined as follows.

Definition 3.5. *We define the infeasibility measure $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}_+$ as*

$$\Phi(x) := \frac{1}{2}(\|g(x)_+\|^2 + \|h(x)\|^2). \quad (3.6)$$

We say that Φ meets the Polyak-Łojasiewicz (PŁ) inequality at $x^ \in \Omega$ when there are constants $\nu > 0$ and $\delta > 0$ such that*

$$\sqrt{\Phi(x)} \leq \nu \|\nabla \Phi(x)\| \quad \forall x \in B_\delta(x^*). \quad (3.7)$$

We call ν a Polyak-Łojasiewicz constant at x^ . If Φ satisfies the PŁ inequality at x^* we say that the Polyak-Łojasiewicz constraint qualification (PŁCQ) holds at x^* .*

This type of inequality was first studied by Polyak in [34], and was identified as a special instance of the Łojasiewicz inequality [28]. Different to our application here, it was formulated for an optimality measure $\Theta(x) := f(x) - f^*$, where f^* is the optimal value of an unconstrained minimization problem with objective f . This is why, since its introduction, the inequality has been mainly used to prove convergence rate results for unconstrained optimization methods like gradient descent and proximal gradient methods, see also [5, 26].

In the approach by Andreani et al. [2], the PŁ inequality for the infeasibility measure Φ as a constraint qualification yields a sufficient condition for multiplier boundedness. Additionally, Andreani et al. show in [2, Theorem 3.2, Theorem 4.1] that, in the case of the constraint functions g, h being also locally Lipschitz continuous, the PŁCQ is equivalent to their own Relaxed Quasi-Normality (RQN) constraint qualification from [3] as well as the commonly used error bound constraint qualification (EB). The latter is defined as follows.

Definition 3.6. *We say that a feasible $x^* \in \Omega$ satisfies the Error Bound (EB) constraint qualification or Error Bound condition if there exist constants $\kappa > 0$ and $\delta > 0$ such that*

$$\text{dist}(x, \Omega) \leq \kappa \sqrt{\Phi(x)} \quad \forall x \in B_\delta(x^*)$$

where κ is called the local error bound constant for Ω at x^ relative to the infeasibility measure Φ .*

The error bound condition was first introduced in [23] for systems of linear inequalities. Similar to the PŁCQ, it has often been used in unconstrained minimization to connect an optimality measure in a point x to the distance of x to a set of minimizers S in order to obtain results about convergence rates (cf. [18, 31]). However, there are also instances in the literature where the EB condition is applied to an infeasibility measure and the distance to the feasible set, like in [32]. In the latter, the EB condition is used to derive an exact penalty result for a constrained problem.

In the following, we show that if the PŁCQ holds in an accumulation point of a SALM sequence, we obtain multiplier boundedness. In the case of locally Lipschitz gradients of the constraint functions, this result can also be extended to the equivalent EB and RQN constraint qualifications. In the latter setting, we also note that on the contrary, if any of the three equivalent CQs is violated, SALM can produce unbounded multiplier sequences even for completely smooth objective functions, as shown in [2, Theorem 3.3], which is why these CQs can be seen as the minimally restrictive ones for ensuring multiplier boundedness.

Theorem 3.7. *Let x^* be feasible and assume that g, h are continuously differentiable. Let additionally φ be locally Lipschitz continuous in a neighborhood of x^* . If the PŁCQ holds at x^* and there exists an infinite subset $K \subseteq \mathbb{N}$ such that the sequence $\{x^k\}$ generated by the SALM converges on K to x^* , then the associated dual subsequences $\{\lambda^k\}_{k \in K}$ and $\{\mu^k\}_{k \in K}$ are bounded.*

Proof. Let $\{x^k\} \subseteq \mathbb{R}^n$ and $\{\rho_k\} \subseteq \mathbb{R}_{++}$ be infinite sequences generated by the safeguarded augmented Lagrangian method. By the definition of the algorithm and similar reasoning as in the proof of Theorem 3.4, we have that there exists $z^k \in \partial\varphi(x^k)$ and $\{E^k\}$ as defined in (3.5) such that

$$\|E^k\| \leq \varepsilon_k \quad (3.8)$$

for all $k \in K$. By $\varepsilon_k \searrow 0$, we obtain $\lim_{k \rightarrow \infty} E^k = 0$. As φ is locally Lipschitz at x^* , there exist $L > 0$ and $\delta > 0$ such that

$$|\varphi(x) - \varphi(y)| \leq L\|x - y\|$$

for all $x, y \in B_\delta(x^*)$. Proposition 2.6 now implies $\|z^k\| \leq L$ for all $k \in \mathbb{K}$ sufficiently large. This again implies by the above results and the continuous differentiability of f that

$$\|\nabla f(x^k) + z^k - E^k\| \leq \|\nabla f(x^k)\| + \|z^k\| + \|E^k\| \leq \|\nabla f(x^*)\| + L + 1 \quad (3.9)$$

for $k \in K$ sufficiently large. We define $M_1 := \max\{u_{\max}, |v_{\min}|, |v_{\max}|\}$. By continuous differentiability of g, h , there exists a constant $M_2 > 0$ such that

$$\|\nabla g_i(x^k)\| \leq M_2 \text{ and } \|\nabla h_j(x^k)\| \leq M_2$$

for all $i = 1, \dots, m$ and $j = 1, \dots, p$ as well as $k \in K$ sufficiently large. Altogether we

have

$$\begin{aligned}
\rho_{k-1} \|\nabla \Phi(x^k)\| &= \left\| \sum_{i=1}^m \rho_{k-1} (g_i(x^k))_+ \nabla g_i(x^k) + \sum_{j=1}^p \rho_{k-1} h_j(x^k) \nabla h_j(x^k) \right\| \\
&= \left\| \sum_{i=1}^m [\rho_{k-1} (g_i(x^k))_+ - \lambda_i^k] \nabla g_i(x^k) \right. \\
&\quad \left. + \sum_{j=1}^p [\rho_{k-1} h_j(x^k) - \mu_j^k] \nabla h_j(x^k) + E^k - \nabla f(x^k) - z^k \right\| \\
&= \left\| \sum_{i=1}^m [\rho_{k-1} (g_i(x^k))_+ - (\rho_{k-1} g_i(x^k) + u_i^{k-1})_+] \nabla g_i(x^k) \right. \\
&\quad \left. + \sum_{j=1}^p [\rho_{k-1} h_j(x^k) - (\rho_{k-1} h_j(x^k) + v_j^{k-1})] \nabla h_j(x^k) + E^k - \nabla f(x^k) - z^k \right\| \\
&\leq \sum_{i=1}^m u_i^{k-1} \|\nabla g_i(x^k)\| + \sum_{j=1}^p |v_j^{k-1}| \|\nabla h_j(x^k)\| + \|E^k - \nabla f(x^k) - z^k\| \\
&\leq \sum_{i=1}^m M_1 M_2 + \sum_{j=1}^p M_1 M_2 + \|\nabla f(x^*)\| + L + 1 \\
&= (m + p) M_1 M_2 + \|\nabla f(x^*)\| + L + 1 =: S
\end{aligned}$$

for sufficiently large $k \in K$, where the first inequality follows from the nonexpansiveness of the projection operator. By assumption, the PL-inequality holds in x^* . By combining it with the above inequality, we obtain

$$\rho_{k-1} \sqrt{\Phi(x^k)} \leq \nu \rho_{k-1} \|\nabla \Phi(x^k)\| \leq \nu S$$

for $k \in K$ sufficiently large, where $\nu > 0$ denotes the PL constant. For all $k \in K$ large enough, we have

$$\begin{aligned}
\|\lambda^k\| &= \|(u^{k-1} + \rho_{k-1} g(x^k))_+\| \\
&\leq \|(u^{k-1} + \rho_{k-1} g(x^k))_+ - \rho_{k-1} g(x^k)_+\| + \rho_{k-1} \|g(x^k)_+\| \\
&\leq \|u^{k-1}\| + \rho_{k-1} \sqrt{2\Phi(x^k)} \\
&\leq \sqrt{m} M_1 + \sqrt{2} \nu S
\end{aligned}$$

and, analogously,

$$\begin{aligned}
\|\mu^k\| &= \|v^{k-1} + \rho_{k-1} h(x^k)\| \\
&\leq \|v^{k-1}\| + \rho_{k-1} \|h(x^k)\| \\
&\leq \|v^{k-1}\| + \rho_{k-1} \sqrt{2\Phi(x^k)} \\
&\leq \sqrt{p} M_1 + \sqrt{2} \nu S.
\end{aligned}$$

This shows the boundedness of the multiplier sequences $\{\lambda^k\}$ and $\{\mu^k\}$ on K . \square

Through our convergence analysis, we implicitly also proved another property of the SALM iterates. Andreani et al. define the CAKKT property (Complementary Approximate KKT) in [4] as a necessary optimality condition for smooth problems, which can be readily extended to non-smooth problems in the following way.

Definition 3.8. $x^* \in \mathbb{R}^n$ fulfills CAM-stationarity (Complementarity Approximate M-stationarity), if x^* is feasible and there exists sequences $\{x^k\}, \{\eta^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}_+^m$ and $\{\mu^k\} \subseteq \mathbb{R}^p$ with $\lambda_i^k \geq 0$ for all $i \in I(x^*)$ and $\lambda_i^k = 0$ for all $i \notin I(x^*)$ such that the following conditions are fulfilled.

- (a) $x^k \rightarrow x^*$.
- (b) For all $k \in \mathbb{N}$ we have $\eta^k \in \nabla f(x^k) + \partial\varphi(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla h_j(x^k)$.
- (c) $\|\eta^k\| \rightarrow 0$.
- (d) $\lim_{k \rightarrow \infty} \sum_{i \in I(x^*)} |\lambda_i^k g_i(x^k)| + \sum_{j=1}^p |\mu_j^k h_j(x^k)| = 0$.

A proof that CAM-stationarity is in fact a necessary optimality condition for our non-smooth program (1.1) is given in the appendix, Theorem A.1. Combining the previous results from this section, we obtain the following corollary.

Corollary 3.9. Let x^* be feasible, fulfilling PŁCQ and there exists an infinite subset $K \subseteq \mathbb{N}$ such that the sequence $\{x^k\}$ generated by the SALM converges on \mathbb{K} to x^* . Then x^* is CAM-stationary.

Proof. If we consider $\{x^k\}_K$, property (a) holds by assumption. By the SALM algorithm, we solve the subproblems as ε_k -stationary, which in turn implies the existence of a sequence $\{\eta^k\} \subseteq \mathbb{R}^n$ satisfying properties (b) and (c). Lastly, by Theorem 3.7, the PŁCQ in x^* guarantees that the corresponding Lagrange multiplier subsequences generated by SALM stay bounded, which already implies (d). Also, by the reasoning from the complementarity section of the proof of Theorem 3.4, we have $\lambda_i^k = 0$ for $i \notin I(x^*)$ and $k \in K$ sufficiently large. Hence all requirements of the sequences $\{\lambda^k\} \subseteq \mathbb{R}_+^m$ and $\{\mu^k\} \subseteq \mathbb{R}^p$ are fulfilled by the SALM-multipliers. \square

4 On the Local Lipschitz Assumption of φ

For some applications of the SALM it would be desirable, not to demand the local Lipschitz property of φ in Theorem 3.7. This issue arises for example when considering sparsity optimization: In order to obtain sparse solutions of a smooth optimization problem, it is popular to set φ to a ℓ^0 -regularization-term and consider the composite problem. The ℓ^0 -pseudonorm is defined by first setting

$$|a|_0 := \begin{cases} 0 & \text{if } a = 0, \\ 1 & \text{otherwise} \end{cases} \quad (4.1)$$

for $a \in \mathbb{R}$ and then setting

$$\|x\|_0 := \sum_{i=1}^n |x_i|_0$$

which is essentially counting the number of non-zero entries of the vector $x \in \mathbb{R}^n$. Note that despite the notation, this does not define a norm on \mathbb{R}^n . The function is also obviously not continuous in zero, and hence also not locally Lipschitz continuous there. However, $\|\cdot\|_0$ is proper and lsc.

The following example shows that requiring φ to be only proper and lsc does not necessarily lead to bounded Lagrange multipliers under the PLCQ which would be required for the applicability of Theorem 3.4. The core observation is that without local Lipschitz continuity, we cannot rely on Lemma 2.6 and may get arbitrarily large Mordukhovich subgradients, which in turn impact the Lagrange multipliers.

4.1 Non-Lipschitz Example With Unbounded Multiplier Sequences

Let $f \equiv 0$ and $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be defined as (cf. Figure 1)

$$\varphi(x) := \begin{cases} 0 & \text{if } x = 0 \text{ or } x = \frac{1}{n} \text{ for } n \in \mathbb{N}, \\ 1 & \text{else.} \end{cases}$$

Then φ is proper and lsc. We also impose two equality constraints with $h_1(x) = x$ and

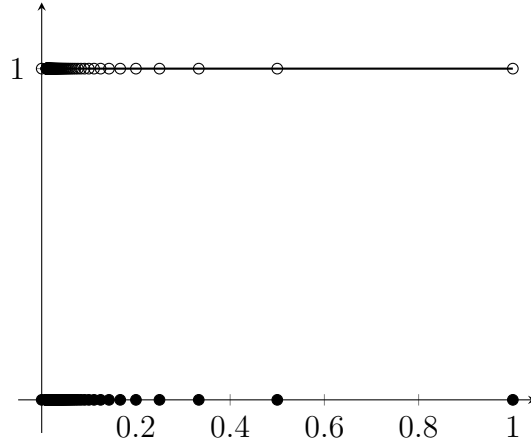


Figure 1: Plot of φ

$h_2(x) = x^2$. The resulting optimization problem has the following properties.

Lemma 4.1. *Consider the optimization problem*

$$\min_{x \in \mathbb{R}} \varphi(x) \quad \text{s.t.} \quad x = 0, x^2 = 0. \quad (4.2)$$

Then the following statements hold.

- (a) For all $n \in \mathbb{N}$, we have $\partial\varphi\left(\frac{1}{n}\right) = \mathbb{R}$.
- (b) PLCQ holds in $x^* = 0$.

Proof. (a) Let $n \in \mathbb{N}$. Then we have $\varphi\left(\frac{1}{n}\right) = 0$ and

$$\varphi(x) = 1 \text{ for all } x \in \left(\frac{1}{n} - \frac{1}{n+1}, \frac{1}{n} + \frac{1}{n+1}\right) \setminus \left\{\frac{1}{n}\right\} \quad (4.3)$$

Now let $a \in \mathbb{R}$. Let $\{x^k\} \subseteq \mathbb{R}$ be arbitrary with $x^k \rightarrow \bar{x} := \frac{1}{n}$ and $x^k \neq \bar{x}$. Then by (4.3), we have for $k \in \mathbb{N}$ sufficiently large that $\varphi(x^k) = 1$ and thus

$$\frac{\varphi(x^k) - \varphi(\bar{x}) - a(x^k - \bar{x})}{|x^k - \bar{x}|} \geq \frac{1 - |a||x^k - \bar{x}|}{|x^k - \bar{x}|} = \frac{1}{|x^k - \bar{x}|} - |a|.$$

We obtain

$$\liminf_{k \rightarrow \infty} \frac{\varphi(x^k) - \varphi(\bar{x}) - a(x^k - \bar{x})}{|x^k - \bar{x}|} \geq \liminf_{k \rightarrow \infty} \frac{1}{|x^k - \bar{x}|} - |a| = +\infty \geq 0$$

This shows

$$a \in \hat{\partial}\varphi\left(\frac{1}{n}\right) \subseteq \partial\varphi\left(\frac{1}{n}\right)$$

and, as $a \in \mathbb{R}$ was chosen as arbitrary,

$$\partial\varphi\left(\frac{1}{n}\right) = \mathbb{R}$$

which verifies statement (a).

(b) We define $\delta := 1$ and $\nu := \frac{1}{\sqrt{2}}$. For all $x \in \mathbb{R}$ with $|x| < \delta = 1$ we have

$$\begin{aligned} \Phi(x) &= \frac{1}{2} \|h(x)\|^2 \\ &= \frac{1}{2} (x^2 + x^4) \\ &\leq \frac{1}{2} (x^2 + 4x^4 + 4x^6) \\ &= \frac{1}{2} (x + 2x^3)^2 \\ &= \frac{1}{2} \left(\begin{pmatrix} x \\ x^2 \end{pmatrix}^T \begin{pmatrix} 1 \\ 2x \end{pmatrix} \right)^2 \\ &= \frac{1}{2} (h(x)^T \nabla h(x))^2 \\ &= \nu^2 (\Phi'(x))^2. \end{aligned}$$

This shows that PŁCQ holds in $x^* = 0$. □

We now show that the analogon of Theorem 3.7 does not hold, i.e., there exists a sequence in \mathbb{R} converging φ -attentively to the solution and only feasible point $x^* = 0$ of (4.2) fulfilling PŁCQ, which has unbounded corresponding Lagrange multipliers $\{\mu^k\}$.

We define the sequence $\{x^k\} := \frac{1}{k}$ and show that it is generated by SALM. We note that in our case the subdifferential of the augmented Lagrangian function at x is given by

$$\partial\mathcal{L}_\rho(x, u, v) = \partial\varphi(x) + (v + \rho h(x))^T \nabla h(x).$$

By Lemma 4.1 (a) we have $\partial\varphi(x^k) = \mathbb{R}$ for all $k \in \mathbb{N}$. Starting in $x^1 = 1$ we can in each step choose $z^k \in \partial\varphi(x^k)$ such that

$$z^k := -(v_{k-1} + \rho_{k-1} h(x^k))^T \nabla h(x^k)$$

and we get $0 \in \partial\mathcal{L}_{\rho_{k-1}}(x^k, u_{k-1}, v_{k-1})$ for all $k \in \mathbb{N}$, so x^k is indeed a ε_k -stationary point of the augmented Lagrangian function in each step and for each choice of $\varepsilon_k \searrow 0$. Moreover, x^k also converges φ -attentively to the feasible point $x^* = 0$, as $\varphi(x^k) = 0$ for all $k \in \mathbb{N}$.

We set the parameters $v^k = 0$ for all $k \in \mathbb{N}$, $\rho_0 = 1$, $\tau = \frac{1}{2}$, $\gamma = 2$. In our specific case, we then have the multiplier update

$$\mu^k = \rho_{k-1} \begin{pmatrix} x^k \\ (x^k)^2 \end{pmatrix} = \rho_{k-1} \begin{pmatrix} \frac{1}{k} \\ \frac{1}{k^2} \end{pmatrix}. \quad (4.4)$$

Additionally we have for $k \in \mathbb{N}$

$$\|h(x^k)\|^2 = \frac{1}{k^2} + \frac{1}{k^4}$$

and thus

$$\frac{\|h(x^{k+1})\|^2}{\|h(x^k)\|^2} = \frac{\frac{1}{(k+1)^2} + \frac{1}{(k+1)^4}}{\frac{1}{k^2} + \frac{1}{k^4}} = \frac{\frac{k^2}{(k+1)^2} + \frac{k^2}{(k+1)^4}}{1 + \frac{1}{k^2}} \rightarrow 1$$

for $k \rightarrow \infty$. This implies the existence of a $K \in \mathbb{N}$ with

$$\|h(x^{k+1})\|^2 > \frac{1}{4} \|h(x^k)\|^2$$

for all $k \geq K$. The penalty update rule implies $\rho_k \geq 2^{k-K}$ for all $k \geq K$ and we obtain with (4.4)

$$\|\mu^k\| = \rho_{k-1} \left\| \begin{pmatrix} \frac{1}{k} \\ \frac{1}{k^2} \end{pmatrix} \right\| \geq 2^{k-1-K} \frac{1}{k} \rightarrow \infty$$

by taking k sufficiently large. Thus, this example results in an unbounded multiplier sequence $\{\mu^k\}$.

4.2 Locally Lipschitz Continuous Regularization Functions

The above example illustrates that the convergence theory we established in this paper is not directly applicable to optimization problems with ℓ_0 -sparsity-regularization. However, there exist ways of circumventing this issue. One of them is to introduce surrogate sparsity functions with similar behavior as the ℓ_0 -pseudonorm, but with nicer analytical properties. A popular approach presented in [11, Section 3.1.4] is to approximate the expression in (4.1) by a function $\psi_\delta : \mathbb{R} \rightarrow [0, 1]$, parameterized by a $\delta > 0$, and then define the surrogate ℓ_0 -function by

$$\varphi(x) := \sum_{i=1}^n \psi_{\delta_i}(x_i).$$

The approximation is done such that ψ_δ satisfies the following three conditions.

- (R1) $\lim_{\delta \searrow 0} \psi_\delta(t) = |t|_0$ for any $t \in \mathbb{R}$ (pointwise).
- (R2) For all $\delta > 0$, ψ_δ is symmetric on \mathbb{R} (i.e., $\psi_\delta(-t) = \psi_\delta(t)$ for all $t \geq 0$) as well as concave and non-decreasing on $[0, \infty)$.
- (R3) $\psi_\delta(t) = 1$ for all t with $|t| \geq \delta$.

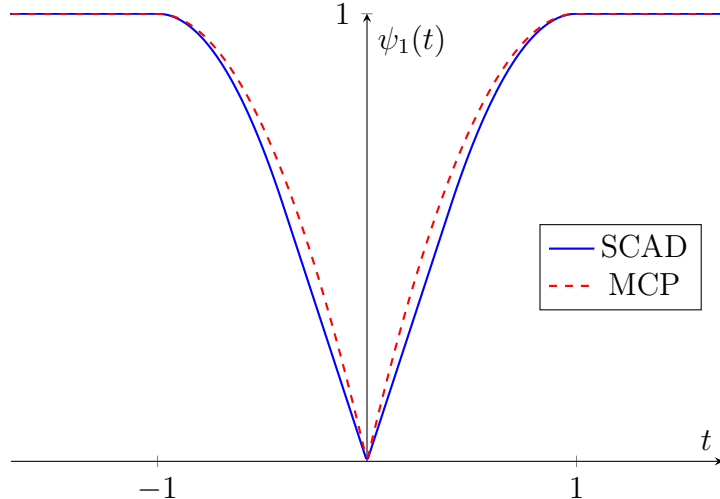


Figure 2: Two surrogate sparsity functions: MCP and SCAD (with $\delta = 1$ and $a = 2.5$)

In the following, we recall the definition of two popular surrogate functions which satisfy the above three properties and which are, in addition, Lipschitz continuous (even globally). They are visualized in Figure 2 and described in the following.

Example 4.2. *The minmax concave penalty (MCP) family [42] is defined by*

$$\psi_{\delta}^{\text{MCP}}(t) := \begin{cases} \frac{2|t|}{\delta} - \frac{t^2}{\delta^2} & \text{if } |t| \leq \delta, \\ 1 & \text{if } |t| > \delta. \end{cases}$$

Another possible choice, which was first introduced in [19], is the SCAD family. The idea is similar to MCP, but the function is now split into three parts, namely a (piecewise) linear section in a neighborhood of zero, the constant part with value one for $|t|$ sufficiently large, and a quadratic part connecting the linear and constant sections while preserving continuous differentiability.

Example 4.3. *The Smoothly Clipped Absolute Deviation (SCAD) family parameterized by $a > 2$ and $\delta > 0$ is defined by*

$$\psi_{\delta,a}^{\text{SCAD}}(t) := \begin{cases} \frac{2a}{(a+1)\delta}|t| & \text{if } |t| \leq \frac{\delta}{a}, \\ 1 - \frac{(\delta-|t|)^2}{(1-\frac{1}{a^2})\delta^2} & \text{if } \frac{\delta}{a} < |t| \leq \delta, \\ 1 & \text{if } |t| \geq \delta. \end{cases}$$

We note that both functions are designed in a way such that they are continuously differentiable outside of zero, and even globally Lipschitz continuous on the whole space \mathbb{R} . Thus, both are a suitable choice for ℓ_0 -surrogates that fulfill Assumption 1.1 for our problem setting.

5 Numerical Results

In this section we are testing the algorithm performance on a model problem and evaluate the relevant metrics. In order to create a flexible test environment tailored to our problem setting that is able to deal with a large range of problems, we implemented our own Python library.

5.1 Implementation Details

The implementation details of the SALM Algorithm 3.1 are as follows: The sequences $\{u^k\}$ and $\{v^k\}$ are obtained in each step by projecting λ^k and μ^k onto $[0, u_{\max}]^m$ and $[v_{\min}, v_{\max}]^p$ respectively. The tolerance reduction for the sequence ε_k is done in a similar way as in *bazinga.jl*¹, namely setting a global tolerance TOL for termination checks, calculating an initial inner tolerance as $\sqrt[3]{\text{TOL}}$ for the subproblems, and for each iteration in which the algorithm does not stop, multiplying the inner tolerance by a decrease factor $\kappa \in (0, 1)$. As solvers of the subproblems, both the proximal gradient method and the quasi-Newton method PANOC⁺ [40, 17] are implemented. For the proximal gradient algorithm, the python library `copt`² is used. The PANOC⁺ implementation was done by adapting the code from the julia Library³ by Alberto de Marchi into Python.

As a termination criterion for SALM, we check the following two conditions modeled after the M-stationarity conditions, we stop the algorithm if both conditions are fulfilled.

- The last subproblem has been solved with a tolerance of TOL_{dual} , indicating approximate stationarity of the augmented Lagrangian.
- We have $\|h(x^k)\|_{\infty} < \text{TOL}_{\text{prim}}$ and $\|\min\{-g(x^k), \lambda^k\}\|_{\infty} < \text{TOL}_{\text{prim}}$, indicating approximate feasibility and complementarity.

For parameter values, we chose $u_{\max} = v_{\max} = -v_{\min} = 10^8$, $\rho_0 = 1$, $\kappa = 0.1$, $\gamma = 2$ and $\tau = 0.8$. We set the maximum iterations for SALM to 200, and the ones for the subproblem solvers to 500 per outer iteration k . In all instances of SALM in this section, we chose a global tolerance of

$$\text{TOL} = \text{TOL}_{\text{prim}} = \text{TOL}_{\text{dual}} = 10^{-6}.$$

5.2 Sparse Portfolio Optimization

The test problem from the field of mathematical finance is the sparse formulation of a portfolio optimization problem, specifically the Markowitz mean-variance model (cf. [10]). The problem in its original form is formulated as

$$\min_x \frac{1}{2} x^T Q x \quad \text{s.t.} \quad \mu^T x \geq \varrho, \quad \mathbf{1}_n^T x = 1, \quad x \geq 0.$$

The vector $x \in \mathbb{R}^n$ is seen as the amount invested in a collection of n financial assets. Expected risk $x^T Q x$ and return $\mu^T x$ are modeled by determining the covariance matrix $Q \in \mathbb{R}^{n \times n}$ and a vector μ of mean values of each asset, which have been calculated from past stock market performance. The value $\varrho > 0$ indicates a minimal return threshold. We assume to have a budget of 1. The equality constraint then indicates that the whole budget must be spent on assets. Finally, we want to invest money, hence only allowing positive spending.

Splitting the investment onto a broad range of assets with potentially small amounts per position may not be feasible in practice due to for example fixed transaction costs

¹Julia library for composite optimization, <https://github.com/aldma/Bazinga.jl>, [16][14]

²<https://github.com/openopt/copt>, [33]

³<https://github.com/JuliaFirstOrder/ProximalAlgorithms.jl>

for each purchase [9]. Thus, we modify the original problem in order to obtain a sparse solution to the portfolio optimization:

$$\min_x \frac{1}{2}x^T Qx + \beta\varphi(x) \quad \text{s.t.} \quad \mu^T x \geq \varrho, \quad \mathbf{1}_n^T x = 1, \quad x \geq 0, \quad (5.1)$$

with a sparsity promoting function φ and regularization factor $\beta > 0$, we obtain a composite formulation this way. A natural choice would be $\varphi = \|\cdot\|_0$. However, in view of the presented convergence theory, we chose the ℓ_0 -norm surrogate $\varphi = \varphi_\delta^{\text{MCP}}$ from Example 4.2 with $\delta = 0.1$ for testing our SALM algorithm. Determining the proximal operator of the MCP function is an elementary, but quite arduous task, we refer to [38] for more details, where the prox-operator is determined for a slightly modified version of MCP.

We additionally ran the problem with ℓ_p^p -regularization for $0 < p < 1$. This function is in particular not locally Lipschitz in zero, which we use to demonstrate the numerical differences in the algorithmic performance of SALM, especially in view of the Lipschitz properties discussed in Section 4. The ℓ_p^p -quasinorm admits analytical expressions for its proximal operator if $p = \frac{1}{2}$ or $p = \frac{2}{3}$ [27]. We chose $p = \frac{1}{2}$ for our numerical experiments.

We obtained the problem data from [21]⁴. It contains three sets P_1, P_2, P_3 consisting each of 10 problems with randomly generated data and dimension $n = 200$. Each set was generated by the authors according to a desired diagonal dominance ratio of the matrices Q , for more details we refer to the documentation.

On each of the 30 problems we ran the SALM algorithm with the proximal gradient method (PG) and PANOC⁺ as subproblem solvers. As a reference solver, we ran the Gurobi solver⁵ on an equivalent MIQP⁶ formulation of the ℓ_0 -regularized problem (5.1) obtained from [20]. For the regularization, we found a good working setup by choosing a common regularization factor for each of the three problem classes as indicated in Table 5.2. We compiled the results in two tables, namely Table 2 for MCP regularization and Table 3 for ℓ_p^p regularization. We use the following abbreviations in reference to the terminal point x^* of the algorithm:

- Method: The SALM implementations are labeled by the subproblem solver used, PG standing for SALM with proximal gradient, PANOC+ for the SALM with PANOC⁺.
- S: Status, 0 when the termination criterion was met, 1 if the algorithm failed to find an acceptable solution before the maximum number of iterations was reached.
- OI: Outer iterations of the SALM Loop.
- II: Cumulated inner iterations of the subproblem solver over all SALM steps.
- Obj.: Value of the objective function $\Psi(x^*) = \frac{1}{2}(x^*)^T Qx^* + \beta\varphi(x^*)$.
- NNZ: Number of non-zero entries in x^* .
- RE: relative Error calculated by

$$\text{RE} = \frac{\text{obj}_{\text{SALM}} - \text{obj}_{\text{Gurobi}}}{\text{obj}_{\text{Gurobi}}},$$

⁴<https://commalab.di.unipi.it/datasets/mv/>

⁵<https://www.gurobi.com>

⁶Mixed Integer Quadratic Programming

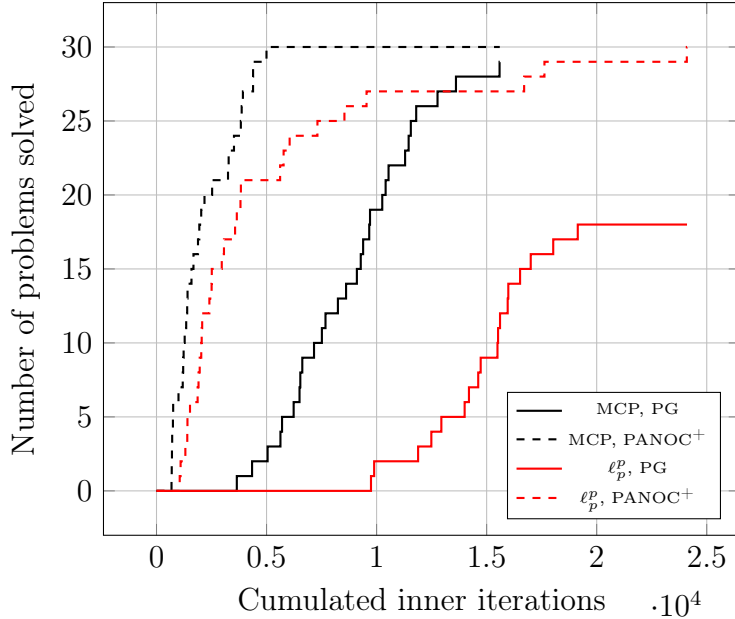


Figure 3: Performance profile of cumulated inner iterations

- Risk: expected risk term $\frac{1}{2}(x^*)^T Q x^*$.
- Return: expected return $\mu^T x^*$.
- λ : Multiplier for the minimum return constraint $\mu^T x \geq \varrho$.
- μ : Multiplier for the budget constraint $\mathbf{1}_n^T x = 1$.

It is worth to note that we incorporated the bound constraint $x \geq 0$ directly into the formulations of the proximal operators. As a starting vector we chose $x_0 = \frac{1}{n} \cdot \mathbf{1}_n$.

Problem Class	Problem Class Name	β
P_1	or1200-005	25
P_2	or1200-05	100
P_3	pard200	50

Table 1: Regularization parameters for the portfolio problem classes

With MCP regularization, SALM was able solve all problems except the instance of `pard200_d` with the PG subproblem solver. For this particular problem, we also see that the multipliers grow quite large, probably indicating unbounded behavior, as in all other instances the multipliers are very tame, being zero in all except one other problem instance. We also see that all successfully obtained solutions have a sparsity of 2 to 6 non-zero entries, which are all quite comparable to the reference Gurobi values.

Another, albeit somewhat expected, effect we can observe is that the quasi-Newton approach of PANOC⁺ generally leads to convergence in fewer iterations in the subproblems compared to the proximal gradient method. This is especially visible when we look at the performance profile for the inner iterations in Figure 3.

However, when looking at computation time, we see that PG performs slightly better, indicating that the PANOC⁺-iterations require a significantly higher computational effort than the ones for proximal gradient, cf. Figure 4.

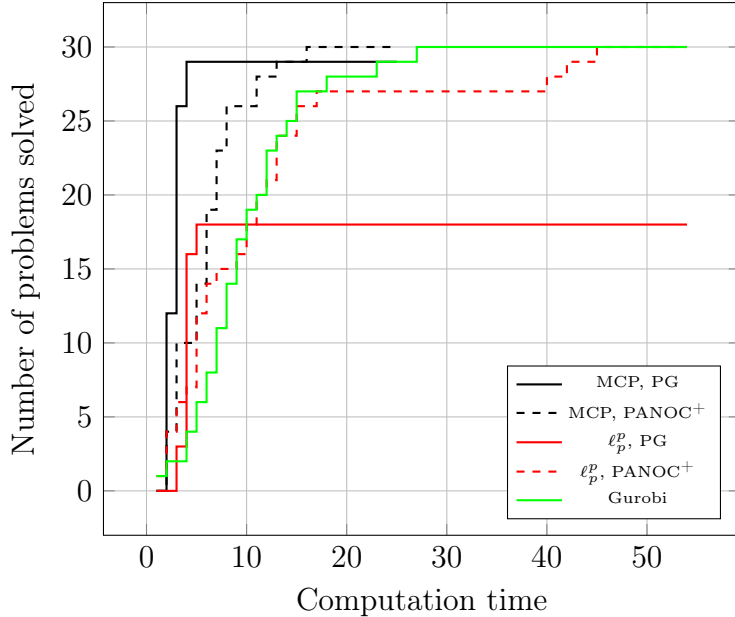


Figure 4: Performance profile of for computation time in seconds

The same behavior can also be observed when comparing the subproblem solvers for the ℓ_p^p -regularized problems also displayed in the same performance profiles. But in contrast to MCP, we here have a total number of 12 problem instances (all using the proximal gradient subproblem solver), in which the algorithm fails to find an acceptable solutions in time. Otherwise, the comparability of the results to the ones from Gurobi and MCP is limited, as the latter ones always contribute a constant value outside (a neighborhood) of zero, whereas the ℓ_p^p -regularization scales infinitely with the distance from zero. This property also leads to a significantly lower sparsity in the result for this regularization method. We also see that especially in the case of failure, but also for successful instances, the multipliers take on higher values than the ones for MCP regularization.

We interpret these results as evidence for our theoretical necessity of local Lipschitzness of φ in Theorems 3.4, 3.7, Section 4.1 in order to ensure sensible behavior of the multiplier sequences, and consequently better convergence guarantees of SALM.

6 Conclusions

We presented the Safeguarded Augmented Lagrangian Method as an algorithm capable of performing optimization on composite constrained problems with a locally Lipschitz continuous regularization function. We obtained a global convergence result to an M-stationary point under the minimally restrictive PLCQ, equivalently under the error bound condition. The boundedness of the Lagrange multipliers played a central role in the convergence analysis, and not imposing a Lipschitz property onto φ may lead to unbounded multipliers, as demonstrated in the counterexample. Our numerical findings underline our theoretical considerations about the choice of a regularization function.

Recently, De Marchi et al. introduced the Elastic SALM in [15], a modification of the SALM algorithm that allows the safeguards to grow in a controlled way. This approach has the advantage of unifying the good global convergence guarantees of SALM with some

desirable properties of the classical ALM when applied to convex problems. As to the best of our knowledge, it is not yet clear, if Elastic SALM also has a global convergence property for potentially non-convex problems under EB/PŁCQ, which might be a topic for further research.

A Appendix

This proof for CAM-stationarity as a necessary optimality condition is a direct extension of [4, Theorem 3.3] to the composite setting.

Theorem A.1. *Let x^* be a local minimizer of (1.1). Then x^* is CAM-stationary.*

Proof. We denote $\Psi := f + \varphi$. As x^* is a local minimizer, there exists a $\delta > 0$ such that $\Psi(x^*) \leq \Psi(x)$ for all feasible $x \in \overline{B}_\delta(x^*)$. We consider the auxiliary problem

$$\min_{x \in \mathbb{R}^n} \Psi(x) + \frac{1}{2} \|x - x^*\|^2 \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0, \quad x \in \overline{B}_\delta(x^*). \quad (\text{A.1})$$

By construction, x^* is the unique solution of Problem (A.1). For each $k \in \mathbb{N}$, we define another auxiliary problem as follows:

$$\min_{x \in \mathbb{R}^n} \Psi(x) + \frac{1}{2} \|x - x^*\|^2 + \frac{k}{2} [\|g(x)_+\|^2 + \|h(x)\|^2] \quad \text{s.t.} \quad x \in \overline{B}_\delta(x^*). \quad (\text{A.2})$$

As we have a continuous objective function on a compact set, there exists a solution x^k to (A.2) for each $k \in \mathbb{N}$. By the feasibility of x^* and the optimality of each x^k with respect to (A.2), we have the property

$$\Psi(x^k) + \frac{1}{2} \|x^k - x^*\|^2 + \frac{k}{2} [\|g(x^k)_+\|^2 + \|h(x^k)\|^2] \leq \Psi(x^*). \quad (\text{A.3})$$

Furthermore, as the sequence $\{x^k\} \subseteq \overline{B}_\delta(x^*)$ is bounded, there exists an infinite set $K \subseteq \mathbb{N}$ and $z^* \in \overline{B}_\delta(x^*)$ such that

$$\lim_{k \rightarrow \infty, k \in K} x^k = z^*.$$

We can now divide expression (A.3) by k and take the limit $k \rightarrow \infty$ for $k \in K$. This yields

$$\|g(z^*)_+\|^2 + \|h(z^*)\|^2 = 0,$$

showing feasibility of z^* for problem (A.1). Additionally, we have

$$\Psi(x^k) + \frac{1}{2} \|x^k - x^*\|^2 \leq \Psi(x^k) + \frac{1}{2} \|x^k - x^*\|^2 + \frac{k}{2} [\|g(x^k)_+\|^2 + \|h(x^k)\|^2] \leq \Psi(x^*),$$

and taking the limit $k \rightarrow \infty, k \in K$ yields

$$\Psi(z^*) + \frac{1}{2} \|z^* - x^*\|^2 \leq \Psi(x^*),$$

which implies $z^* = x^*$ by the optimality of x^* for Problem (A.1), thus showing

$$\lim_{k \rightarrow \infty, k \in K} x^k = x^*. \quad (\text{A.4})$$

We now define the sequences $\lambda^k := kg(x^k)_+$, $\mu^k := kh(x^k)$ and $\eta^k := x^* - x^k$ for $k \in \mathbb{K}$. Clearly, $\lambda^k \geq 0$ and for $i \notin I(x^*)$, by the continuity of g we have $g_i(x^k) \leq 0$ and hence $\lambda_i^k = 0$ for k sufficiently large. We take $K_1 \subseteq K$ such that $\lambda_i^k = 0$ for all $i \notin I(x^*)$ and $k \in K_1$ and show that all properties of CAM-stationarity hold for the above defined sequences on the index set K_1 .

In (A.4) we showed property (a) of CAM-stationarity for the sequence $\{x_k\}_{k \in K_1}$. By definition, $\{\eta^k\}_{k \in K_1}$ fulfills (c).

By local optimality with respect to the auxiliary problem (A.2), the generalized Fermat rule from Lemma 2.2 and the Mordukhovich subdifferential sum rule yields

$$0 \in \nabla f(x^k) + \partial\varphi(x^k) + x^k - x^* + \sum_{i=1}^m kg_i(x^k)_+ \nabla g_i(x^k) + \sum_{j=1}^p kh_j(x^k) \nabla h_j(x^k),$$

which is equivalent to

$$\eta^k = x^* - x^k \in \nabla f(x^k) + \partial\varphi(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla h_j(x^k),$$

which shows property (b), whereas (c) holds by definition of η_k and (a).

Once again considering (A.3) and taking the limit $k \rightarrow \infty$ for $k \in K_1$ yields

$$\lim_{k \rightarrow \infty, k \in K_1} \sum_{i=1}^m k(g_i(x^k)_+)^2 + \sum_{j=1}^p kh_j(x^k)^2 = 0.$$

As

$$|\mu_j^k h_j(x^k)| = |kh_j(x^k)^2| = kh_j(x^k)^2$$

and

$$|\lambda_i^k g_i(x^k)| = |kg_i(x^k)_+ g_i(x^k)| = |k(g_i(x^k)_+)^2| = k(g_i(x^k)_+)^2,$$

as well as $\lambda_i^k = 0$ for all $k \in K_1$ and $i \notin I(x^*)$, we have

$$\lim_{k \rightarrow \infty, k \in K_1} \sum_{i \in I(x^*)} |\lambda_i^k g_i(x^k)| + \sum_{j=1}^p |\mu_j^k h_j(x^k)| = 0,$$

which shows the last property (d). □

Problem	Method	S	OI	II	Obj.	NNZ	RE	Risk	Return	λ	μ
orl200-005-a	Gurobi	-	-	-	189.779	4	-	89.779	0.00692	-	-
	PG	0	49	11,555	199.71	3	0.052	124.71	0.00897	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	49	1,212	200.358	5	0.056	75.358	0.00877	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-b	Gurobi	-	-	-	189.779	4	-	89.779	0.00593	-	-
	PG	0	50	7,157	229.368	2	0.209	179.368	0.00913	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	50	2,180	229.368	2	0.209	179.368	0.00913	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-c	Gurobi	-	-	-	190.122	4	-	90.122	0.00692	-	-
	PG	0	47	9,280	196.343	3	0.033	121.343	0.00934	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	48	1,394	201.427	5	0.059	76.427	0.00957	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-d	Gurobi	-	-	-	190.576	4	-	90.576	0.00631	-	-
	PG	0	48	10,410	198.446	3	0.041	123.446	0.00837	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	47	1,232	213.705	6	0.121	63.705	0.00827	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-e	Gurobi	-	-	-	189.812	4	-	89.812	0.00696	-	-
	PG	0	49	8,242	192.464	4	0.014	92.464	0.00926	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	49	1,280	199.117	5	0.049	74.117	0.00929	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-f	Gurobi	-	-	-	190.905	4	-	90.905	0.00661	-	-
	PG	0	47	6,231	203.993	3	0.069	128.993	0.00917	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	47	2,031	203.993	3	0.069	128.993	0.00917	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-g	Gurobi	-	-	-	189.281	4	-	89.281	0.00677	-	-
	PG	0	50	10,543	233.423	2	0.233	183.423	0.00966	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	50	1,885	233.423	2	0.233	183.423	0.00966	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-h	Gurobi	-	-	-	189.613	4	-	89.613	0.00751	-	-
	PG	0	49	9,104	233.458	2	0.231	183.458	0.00963	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	47	999	192.661	4	0.016	92.661	0.00941	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-i	Gurobi	-	-	-	190.713	4	-	90.713	0.00549	-	-
	PG	0	51	4,342	229.499	2	0.203	179.499	0.00685	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	48	702	197.199	3	0.034	122.199	0.00778	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-j	Gurobi	-	-	-	189.619	4	-	89.619	0.00493	-	-
	PG	0	49	11,799	200.128	3	0.055	125.128	0.00768	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	49	1,349	201.55	5	0.063	76.55	0.00802	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-a	Gurobi	-	-	-	306.125	2	-	126.125	0.00481	-	-
	PG	0	46	7,678	356.813	3	0.166	86.813	0.00922	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	44	712	503.799	5	0.646	53.799	0.00904	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-b	Gurobi	-	-	-	305.5	2	-	125.5	0.00646	-	-
	PG	0	46	8,613	357.792	3	0.171	87.792	0.00775	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	44	692	504.382	5	0.651	54.382	0.00724	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-c	Gurobi	-	-	-	306.625	2	-	126.625	0.00682	-	-
	PG	0	46	9,701	358.32	3	0.169	88.32	0.00824	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	44	746	585.712	6	0.91	45.712	0.00833	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-d	Gurobi	-	-	-	308.374	2	-	128.374	0.00923	-	-
	PG	0	44	7,514	356.92	3	0.157	86.92	0.00899	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	42	3,278	588.228	6	0.908	48.228	0.00921	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-e	Gurobi	-	-	-	312.934	2	-	132.934	0.00944	-	-
	PG	0	46	11,456	359.815	3	0.15	89.815	0.00919	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	45	4,387	510.839	5	0.632	60.839	0.00919	$7.73 \cdot 10^4$	$5.88 \cdot 10^2$
orl200-05-f	Gurobi	-	-	-	310.232	2	-	130.232	0.00854	-	-
	PG	0	45	3,649	360.444	3	0.162	90.444	0.00903	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	43	1,679	506.499	5	0.633	56.499	0.00896	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-g	Gurobi	-	-	-	306.999	2	-	126.999	0.00862	-	-
	PG	0	46	10,257	357.584	3	0.165	87.584	0.00776	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	44	682	503.98	5	0.642	53.98	0.00682	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-h	Gurobi	-	-	-	305.5	2	-	125.5	0.00606	-	-
	PG	0	46	6,534	357.106	3	0.169	87.106	0.00876	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	44	1,192	504.091	5	0.65	54.091	0.00848	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-i	Gurobi	-	-	-	306.248	2	-	126.248	0.00532	-	-
	PG	0	46	6,498	356.936	3	0.166	86.936	0.00788	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	46	757	427.509	4	0.396	67.509	0.00813	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-j	Gurobi	-	-	-	308.353	2	-	128.353	0.00716	-	-
	PG	0	46	11,295	360.91	3	0.17	90.91	0.00949	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	46	1,410	429.789	4	0.394	69.789	0.0094	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_a	Gurobi	-	-	-	455.601	5	-	205.601	0.00639	-	-
	PG	0	50	5,631	494.432	3	0.085	344.432	0.00821	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	49	3,832	464.039	4	0.019	264.039	0.00856	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_b	Gurobi	-	-	-	463.791	5	-	213.791	0.00892	-	-
	PG	0	50	9,388	504.996	3	0.089	354.996	0.00952	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	50	3,926	468.509	4	0.01	268.509	0.00951	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_c	Gurobi	-	-	-	458.203	5	-	208.203	0.00867	-	-
	PG	0	51	9,665	500.99	3	0.093	350.99	0.00959	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	51	4,997	500.99	3	0.093	350.99	0.00959	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_d	Gurobi	-	-	-	459.211	5	-	209.211	0.00898	-	-
	PG	1	200	91,026	514.45	20	0.12	334.006	0.00783	$5.65 \cdot 10^{16}$	$5.23 \cdot 10^{14}$
	PANOC+	0	48	1,589	459.77	5	0.001	209.77	0.00911	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_e	Gurobi	-	-	-	455.404	4	-	255.404	0.00567	-	-
	PG	0	51	12,768	502.874	3	0.104	352.874	0.00908	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	51	4,388	499.272	3	0.096	349.272	0.00932	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_f	Gurobi	-	-	-	452.966	4	-	252.966	0.00429	-	-
	PG	0	51	5,048	489.828	3	0.081	339.828	0.00902	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	51	1,945	489.828	3	0.081	339.828	0.00902	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_g	Gurobi	-	-	-	457.109	5	-	207.109	0.00827	-	-
	PG	0	48	6,624	461.692	4	0.01	261.692	0.00926	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	48	3,249	461.692	4	0.01	261.692	0.00926	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_h	Gurobi	-	-	-	455.832	4	-	255.832	0.00574	-	-
	PG	0	53	15,587	539.696	3	0.184	389.696	0.00748	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	50	3,870	460.177	4	0.01	260.177	0.00895	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_i	Gurobi	-	-	-	455.23	5	-	205.23	0.00639	-	-
	PG	0	51	5,704	494.325	3	0.086	344.325	0.00862	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	51	2,527	494.325	3	0.086	344.325	0.00862	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_j	Gurobi	-	-	-	453.958	4	-	253.958	0.00661	-	-
	PG	0	52	13,606	513.207	3	0.131	363.207	0.00776	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	52	3,520	494.062	3	0.088	344.062	0.00864	$0 \cdot 10^0$	$0 \cdot 10^0$

Table 2: Sparse portfolio optimization results with MCP-regularization

Problem	Method	S	OI	II	Objective	NNZ	RE	Risk	Return	λ	μ
orl200-005-a	Gurobi	-	-	-	189.779	4	-	89.779	0.00692	-	-
	PG	0	46	15,984	126.883	17	-0.331	23.878	0.00649	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	46	2,053	125.175	16	-0.34	25.237	0.00674	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-b	Gurobi	-	-	-	189.779	4	-	89.779	0.00593	-	-
	PG	0	44	14,617	126.867	17	-0.332	23.903	0.00566	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	44	3,065	125.071	16	-0.341	25.15	0.00565	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-c	Gurobi	-	-	-	190.122	4	-	90.122	0.00692	-	-
	PG	1	200	94,001	143.25	56	-0.247	22.326	0.0065	$8.13 \cdot 10^{15}$	$6.56 \cdot 10^{13}$
	PANOC+	0	43	3,649	125.191	16	-0.342	25.233	0.00663	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-d	Gurobi	-	-	-	190.576	4	-	90.576	0.00631	-	-
	PG	0	45	14,729	127.194	17	-0.333	24.159	0.00581	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	45	2,402	123.694	15	-0.351	26.901	0.00584	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-e	Gurobi	-	-	-	189.812	4	-	89.812	0.00696	-	-
	PG	0	46	15,956	128.741	18	-0.322	22.753	0.00642	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	46	1,963	123.374	15	-0.35	26.583	0.0065	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-f	Gurobi	-	-	-	190.905	4	-	90.905	0.00661	-	-
	PG	1	200	94,001	161.702	68	-0.153	20.91	0.00597	$3.35 \cdot 10^{16}$	$2.17 \cdot 10^{14}$
	PANOC+	0	36	6,049	122.166	14	-0.36	28.709	0.0063	$1.64 \cdot 10^3$	$0 \cdot 10^0$
orl200-005-g	Gurobi	-	-	-	189.281	4	-	89.281	0.00677	-	-
	PG	1	200	94,001	148.868	59	-0.214	22.717	0.00632	$1.77 \cdot 10^{16}$	$1.44 \cdot 10^{14}$
	PANOC+	0	38	5,616	121.624	14	-0.357	28.244	0.00649	$1.15 \cdot 10^3$	$0 \cdot 10^0$
orl200-005-h	Gurobi	-	-	-	189.613	4	-	89.613	0.00751	-	-
	PG	1	200	94,001	150.345	56	-0.207	22.548	0.00703	$2.5 \cdot 10^{16}$	$2.14 \cdot 10^{14}$
	PANOC+	0	44	9,543	125.238	16	-0.34	25.47	0.00728	$8.57 \cdot 10^2$	$0 \cdot 10^0$
orl200-005-i	Gurobi	-	-	-	190.713	4	-	90.713	0.00549	-	-
	PG	0	49	18,025	129.25	18	-0.322	23.351	0.00641	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	45	1,095	125.66	16	-0.341	25.752	0.00617	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-005-j	Gurobi	-	-	-	189.619	4	-	89.619	0.00493	-	-
	PG	0	54	19,140	128.414	18	-0.323	22.386	0.00477	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	40	3,815	124.822	16	-0.342	24.866	0.00468	$3.96 \cdot 10^2$	$0 \cdot 10^0$
orl200-05-a	Gurobi	-	-	-	306.125	2	-	126.125	0.00481	-	-
	PG	0	48	17,003	264.951	6	-0.135	44.612	0.00751	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	45	1,406	254.294	5	-0.169	53.082	0.00767	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-b	Gurobi	-	-	-	305.5	2	-	125.5	0.00646	-	-
	PG	0	48	16,519	264.977	6	-0.133	44.689	0.00567	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	45	2,033	264.977	6	-0.133	44.685	0.00567	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-c	Gurobi	-	-	-	306.625	2	-	126.625	0.00682	-	-
	PG	0	46	15,492	266.122	6	-0.132	45.789	0.00832	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	45	1,398	266.122	6	-0.132	45.788	0.00832	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-d	Gurobi	-	-	-	308.374	2	-	128.374	0.00923	-	-
	PG	1	200	93,501	311.972	43	0.012	44.068	0.00838	$3.77 \cdot 10^{16}$	$3.62 \cdot 10^{14}$
	PANOC+	0	37	5,775	255.438	5	-0.172	54.539	0.00876	$5.58 \cdot 10^3$	$0 \cdot 10^0$
orl200-05-e	Gurobi	-	-	-	312.934	2	-	132.934	0.00944	-	-
	PG	1	200	93,501	342.4	27	0.094	40.661	0.00811	$1.08 \cdot 10^{17}$	$8.76 \cdot 10^{14}$
	PANOC+	0	56	1,909	344.153	5	0.1	212.981	0.00919	$0 \cdot 10^0$	$9.75 \cdot 10^5$
orl200-05-f	Gurobi	-	-	-	310.232	2	-	130.232	0.00854	-	-
	PG	1	200	93,501	351.154	26	0.132	42.411	0.00739	$6.01 \cdot 10^{16}$	$5.39 \cdot 10^{14}$
	PANOC+	0	38	7,312	258.872	5	-0.166	59.852	0.00799	$7.54 \cdot 10^3$	$0 \cdot 10^0$
orl200-05-g	Gurobi	-	-	-	306.999	2	-	126.999	0.00862	-	-
	PG	0	46	15,519	265.265	6	-0.136	44.948	0.00493	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	45	1,311	254.205	5	-0.172	52.96	0.00419	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-h	Gurobi	-	-	-	305.5	2	-	125.5	0.00606	-	-
	PG	0	47	15,608	264.991	6	-0.133	44.601	0.0068	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	46	1,868	264.991	6	-0.133	44.6	0.0068	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-i	Gurobi	-	-	-	306.248	2	-	126.248	0.00532	-	-
	PG	0	46	14,196	264.893	6	-0.135	44.501	0.00496	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	45	2,965	264.893	6	-0.135	44.5	0.00496	$0 \cdot 10^0$	$0 \cdot 10^0$
orl200-05-j	Gurobi	-	-	-	308.353	2	-	128.353	0.00716	-	-
	PG	1	200	93,501	295.92	39	-0.04	45.106	0.00709	$6.86 \cdot 10^{15}$	$6.09 \cdot 10^{13}$
	PANOC+	0	43	3,834	266.635	6	-0.135	46.378	0.00767	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_a	Gurobi	-	-	-	455.601	5	-	205.601	0.00639	-	-
	PG	0	46	14,000	282.481	21	-0.38	53.53	0.0055	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	45	3,570	282.481	21	-0.38	53.53	0.0055	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_b	Gurobi	-	-	-	463.791	5	-	213.791	0.00892	-	-
	PG	1	200	93,646	385.835	67	-0.168	45.323	0.0076	$1.33 \cdot 10^{17}$	$1.09 \cdot 10^{15}$
	PANOC+	0	55	16,703	292.39	8	-0.37	153.832	0.00892	$7.42 \cdot 10^4$	$2.84 \cdot 10^2$
pard200_c	Gurobi	-	-	-	458.203	5	-	208.203	0.00867	-	-
	PG	1	200	93,776	356.932	56	-0.221	49.484	0.00765	$9.66 \cdot 10^{16}$	$7.83 \cdot 10^{14}$
	PANOC+	0	70	24,095	265.34	11	-0.421	100.196	0.00862	$3.07 \cdot 10^4$	$0 \cdot 10^0$
pard200_d	Gurobi	-	-	-	459.211	5	-	209.211	0.00898	-	-
	PG	1	200	93,833	396.328	78	-0.137	44.368	0.0074	$9.9 \cdot 10^{16}$	$7.68 \cdot 10^{14}$
	PANOC+	0	57	17,626	271.128	9	-0.41	121.711	0.00839	$2.5 \cdot 10^4$	$0 \cdot 10^0$
pard200_e	Gurobi	-	-	-	455.404	4	-	255.404	0.00567	-	-
	PG	0	46	9,889	273.458	18	-0.4	61.411	0.00566	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	46	2,512	279.283	20	-0.387	55.819	0.00604	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_f	Gurobi	-	-	-	452.966	4	-	252.966	0.00429	-	-
	PG	0	46	11,886	272.075	18	-0.399	60.019	0.0057	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	46	1,054	275.089	19	-0.393	57.284	0.0058	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_g	Gurobi	-	-	-	457.109	5	-	207.109	0.00827	-	-
	PG	1	200	93,519	358.97	62	-0.215	49.524	0.00729	$7.83 \cdot 10^{16}$	$6.3 \cdot 10^{14}$
	PANOC+	0	39	8,538	266.584	13	-0.417	87.746	0.00808	$1.86 \cdot 10^4$	$0 \cdot 10^0$
pard200_h	Gurobi	-	-	-	455.832	4	-	255.832	0.00574	-	-
	PG	0	47	9,745	274.124	18	-0.399	62.101	0.00618	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	47	1,522	279.937	20	-0.386	56.497	0.00591	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_i	Gurobi	-	-	-	455.23	5	-	205.23	0.00639	-	-
	PG	0	45	12,487	278.607	20	-0.388	55.095	0.00605	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	45	2,504	278.67	20	-0.388	55.17	0.00627	$0 \cdot 10^0$	$0 \cdot 10^0$
pard200_j	Gurobi	-	-	-	453.958	4	-	253.958	0.00661	-	-
	PG	0	47	12,942	275.76	19	-0.393	57.93	0.00613	$0 \cdot 10^0$	$0 \cdot 10^0$
	PANOC+	0	47	2,075	278.809	20	-0.386	55.363	0.00604	$0 \cdot 10^0$	$0 \cdot 10^0$

Table 3: Sparse portfolio optimization results with $\ell^{\frac{1}{2}}$ -regularization

References

- [1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. “On augmented Lagrangian methods with general lower-level constraints”. In: *SIAM Journal on Optimization* 18.4 (2007), pp. 1286–1309. DOI: 10.1137/060654797.
- [2] R. Andreani, G. Haeser, R. Prado, and L. Secchin. *Primal-dual global convergence of an augmented Lagrangian method under the error bound condition*. July 22, 2025. URL: <https://optimization-online.org/?p=31199>.
- [3] R. Andreani, G. Haeser, M. L. Schuverdt, and L. D. Secchin. *A relaxed quasnormality condition and the boundedness of dual augmented Lagrangian sequences*. Oct. 8, 2024. URL: <https://optimization-online.org/?p=25207>.
- [4] R. Andreani, J. M. Martínez, and B. F. Svaiter. “A new sequential optimality condition for constrained optimization and algorithmic consequences”. In: *SIAM Journal on Optimization* 20.6 (2010), pp. 3533–3554. DOI: 10.1137/090777189.
- [5] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. “Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality”. In: *Mathematics of Operations Research* 35.2 (2010), pp. 438–457. DOI: 10.1287/moor.1100.0449.
- [6] W. Bian and X. Chen. “Linearly constrained non-Lipschitz optimization for image restoration”. In: *SIAM Journal on Imaging Sciences* 8.4 (2015), pp. 2294–2322. DOI: 10.1137/140985639.
- [7] E. G. Birgin and J. M. Martínez. *Practical augmented Lagrangian methods for constrained optimization*. Vol. 10. Fundamentals of Algorithms. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2014. DOI: 10.1137/1.9781611973365.
- [8] J.-F. Cai, E. J. Candès, and Z. Shen. “A singular value thresholding algorithm for matrix completion”. In: *SIAM Journal on Optimization* 20.4 (2010), pp. 1956–1982. DOI: 10.1137/080738970.
- [9] X. Chen, L. Guo, Z. Lu, and J. J. Ye. “An augmented Lagrangian method for non-Lipschitz nonconvex programming”. In: *SIAM Journal on Numerical Analysis* 55.1 (2017), pp. 168–193. DOI: 10.1137/15M1052834.
- [10] G. Cornuéjols, J. Peña, and R. Tütüncü. *Optimization Methods in Finance*. 2nd edition. Cambridge: Cambridge University Press, 2018. DOI: 10.1017/9781107297340.
- [11] Y. Cui and J.-S. Pang. *Modern Nonconvex Nondifferentiable Optimization*. Vol. 29. MOS-SIAM Series on Optimization. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2021. DOI: 10.1137/1.9781611976748.
- [12] S. vom Dahl, A. D. Marchi, and C. Kanzow. *Proximal limited-memory quasi-Newton methods for nonsmooth nonconvex optimization*. 2026. arXiv: 2605.11627 [math.OA].
- [13] A. De Marchi. “Constrained and sparse switching times optimization via augmented Lagrangian proximal methods”. In: *2020 American Control Conference, ACC 2020, Denver, CO, USA, July 1-3, 2020*. IEEE, 2020, pp. 3633–3638. DOI: 10.23919/ACC45564.2020.9147892.
- [14] A. De Marchi. “Implicit augmented Lagrangian and generalized optimization”. In: *Journal of Applied and Numerical Optimization* 6.2 (2024), pp. 291–320. DOI: 10.23952/jano.6.2024.2.08.
- [15] A. De Marchi, T. Hoheisel, and P. Mehlitz. *Augmented Lagrangian methods for fully convex composite optimization*. 2026. arXiv: 2511.07117 [math.OA].

- [16] A. De Marchi, X. Jia, C. Kanzow, and P. Mehlitz. “Constrained composite optimization and augmented Lagrangian methods”. In: *Mathematical Programming* 201.1-2 (A) (2023), pp. 863–896. DOI: 10.1007/s10107-022-01922-4.
- [17] A. De Marchi and A. Themelis. “Proximal gradient algorithms under local Lipschitz gradient continuity. A convergence and robustness analysis of PANOC”. In: *Journal of Optimization Theory and Applications* 194.3 (2022), pp. 771–794. DOI: 10.1007/s10957-022-02048-5.
- [18] D. Drusvyatskiy and A. S. Lewis. “Error bounds, quadratic growth, and linear convergence of proximal methods”. In: *Mathematics of Operations Research* 43.3 (2018), pp. 919–948. DOI: 10.1287/moor.2017.0889.
- [19] J. Fan and R. Li. “Variable selection via nonconcave penalized likelihood and its oracle properties.” In: *Journal of the American Statistical Association* 96.456 (2001), pp. 1348–1360. DOI: 10.1198/016214501753382273.
- [20] M. Feng, J. E. Mitchell, J.-S. Pang, X. Shen, and A. Wächter. “Complementarity formulations of l_0 -norm optimization problems”. In: *Pacific Journal of Optimization* 14.2 (2018), pp. 273–305.
- [21] A. Frangioni and C. Gentile. “SDP diagonalizations and perspective cuts for a class of nonseparable MIQP”. In: *Operations Research Letters* 35.2 (2007), pp. 181–185. DOI: 10.1016/j.orl.2006.03.008.
- [22] M. R. Hestenes. “Multiplier and gradient methods”. In: *Journal of Optimization Theory and Applications* 4 (1969), pp. 303–320. DOI: 10.1007/BF00927673.
- [23] A. J. Hoffman. “On approximate solutions of systems of linear inequalities”. In: *Journal of research of the National Bureau of Standards* 49 (1952), p. 263. URL: <https://api.semanticscholar.org/CorpusID:17483872>.
- [24] C. Kanzow and P. Mehlitz. “Convergence properties of monotone and nonmonotone proximal gradient methods revisited”. In: *Journal of Optimization Theory and Applications* 195.2 (2022), pp. 624–646. DOI: 10.1007/s10957-022-02101-3.
- [25] C. Kanzow and D. Steck. “An example comparing the standard and safeguarded augmented Lagrangian methods”. In: *Operations Research Letters* 45.6 (2017), pp. 598–603. DOI: 10.1016/j.orl.2017.09.005.
- [26] H. Karimi, J. Nutini, and M. Schmidt. “Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition”. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by P. Frasconi, N. Landwehr, G. Manco, and J. Vreeken. Cham: Springer International Publishing, 2016, pp. 795–811. DOI: <https://doi.org/10.1007/978-3-319-46128-1>.
- [27] Y. Liu and R. Lin. “A bisection method for computing the proximal operator of the ℓ_p -norm for any $0 < p < 1$ with application to Schatten p -norms”. In: *Journal of Computational and Applied Mathematics* 447 (2024). Id/No 115897, p. 9. DOI: 10.1016/j.cam.2024.115897.
- [28] S. Łojasiewicz. “A topological property of real analytic subsets”. In: *Les Équations aux Dérivées Partielles* (1963), pp. 87–89.
- [29] Z. Lu and Y. Zhang. “An augmented Lagrangian approach for sparse principal component analysis”. In: *Mathematical Programming* 135.1-2 (A) (2012), pp. 149–193. DOI: 10.1007/s10107-011-0452-4.
- [30] B. S. Mordukhovich. *Variational Analysis and Applications*. Springer Monographs in Mathematics. Cham: Springer, 2018. DOI: 10.1007/978-3-319-92775-6.

- [31] I. Necoara, Y. Nesterov, and F. Glineur. “Linear convergence of first order methods for non-strongly convex optimization”. In: *Mathematical Programming* 175.1-2 (A) (2019), pp. 69–107. DOI: 10.1007/s10107-018-1232-1.
- [32] J.-S. Pang. “Error bounds in mathematical programming”. In: *Mathematical Programming* 79.1-3 (B) (1997), pp. 299–332. DOI: 10.1007/BF02614322.
- [33] F. Pedregosa, G. Negiar, and G. Dresdner. “copt: composite optimization in Python”. In: (2020). DOI: 10.5281/zenodo.1283339. URL: <http://openopt.github.io/copt/>.
- [34] B. T. Polyak. “Gradient methods for the minimisation of functionals”. In: *Zhurnal Vychislitel’noi Matematiki i Matematicheskoi Fiziki* 3 (1963), pp. 643–653.
- [35] M. J. D. Powell. *A method for nonlinear constraints in minimization problems*. Optimization, Sympos. Inst. Math. Appl. Univ. Keele, England. 1969.
- [36] R. T. Rockafellar. “The multiplier method of Hestenes and Powell applied to convex programming”. In: *Journal of Optimization Theory and applications* 12.6 (1973), pp. 555–562.
- [37] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Vol. 317. Grundlehren der Mathematischen Wissenschaften. Berlin: Springer, 1998. DOI: 10.1007/978-3-642-02431-3.
- [38] L. Shen, B. W. Suter, and E. E. Tripp. “Structured sparsity promoting functions”. In: *Journal of Optimization Theory and Applications* 183.2 (2019), pp. 386–421. DOI: 10.1007/s10957-019-01565-0.
- [39] L. Song, A. Smola, A. Gretton, and K. M. Borgwardt. “A dependence maximization view of clustering”. In: *Proceedings of the 24th International Conference on Machine Learning*. ICML ’07. Corvallis, Oregon, USA: Association for Computing Machinery, 2007, pp. 815–822. DOI: 10.1145/1273496.1273599. URL: <https://doi.org/10.1145/1273496.1273599>.
- [40] A. Themelis, L. Stella, and P. Patrinos. “Forward-backward envelope for the sum of two nonconvex functions: further properties and nonmonotone linesearch algorithms”. In: *SIAM Journal on Optimization* 28.3 (2018), pp. 2274–2303. DOI: 10.1137/16M1080240.
- [41] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. “Sparse reconstruction by separable approximation”. In: *IEEE Transactions on Signal Processing* 57.7 (2009), pp. 2479–2493. DOI: 10.1109/TSP.2009.2016892.
- [42] C.-H. Zhang. “Nearly unbiased variable selection under minimax concave penalty”. In: *The Annals of Statistics* 38.2 (2010), pp. 894–942.