

A Non-Smooth Analysis Approach for Optimization of Complex Industrial Processes

Anjali Bhardwaj¹, Chandrashekhar Ramesh Ramtirthkar², Dr. Jayant Damodar Supe³, Pramod Reddy⁴, Shivangi Gupta⁵, Sumeet Singh Sarpal⁶, M. Saravanan⁷, K. Anitha⁸

¹Assistant Professor, Department of Computer Science & Engineering, Noida International University, Uttar Pradesh. anjali.bhardwaj@niu.edu.in

²Associate Professor, Department of Mechanical Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra. chandrashekhar.ramtirthkar@vit.edu

³Rungta College of Engineering and Technology, Rungta International Skills University, Bhilai, Chhattisgarh, India. dr.jayant.supe@rungta.org

⁴Centre for Multidisciplinary Research, Anurag University, Hyderabad, Telangana, India. pramodreddy.pr@proton.me, <https://orcid.org/0009-0003-6779-0588>

⁵Quantum University Research Center, Quantum University, Roorkee, Uttarakhand, India. shivangi.gupta@quantumeducation.in, <https://orcid.org/0009-0005-4285-8974>

⁶Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India. sumeet.sarpal.orp@chitkara.edu.in, <https://orcid.org/0009-0002-9163-4895>

⁷Assistant Professor - Gr. II, Department of Mechanical Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation (DU), Tamil Nadu, India. saravanan@avit.ac.in, <https://orcid.org/0000-0002-6717-3315>

⁸Associate Professor, Department of Management Studies, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. anithak@maher.ac.in

Abstract

Non-smooth objective functions may occur in industrial optimization problems due to the existence of thresholds, switches, piecewise objective functions, absolute value terms, and bounds on the feasible set. Gradient-based approaches may lose their efficiency due to the lack of differentiability assumption. In this paper, a mathematical framework for constrained non-smooth optimization is developed. The theory includes a composite representation of the objective function in terms of continuously differentiable and locally Lipschitz non-smooth components, Clarke generalized derivatives, minimum-norm subgradients rule, geometry of tangent and normal cones, proximal stabilization, and adaptive step-size choice. A deterministic minimum norm selection criterion is used to deal with a multi-valued generalized subdifferential. Feasibility is achieved via Euclidean projection. The proximal term makes the algorithm stable in the vicinity of non-differentiable points. The theoretical properties include the existence of an optimum solution, sufficient descent, objective monotonicity, boundedness of the sequence, and generalized stationarity of its accumulation points under certain regularity assumptions. Numerical performance is analyzed on three test cases: absolute-value, maximum-type, and L1-regularized composite benchmark functions and an engineering-oriented problem of the industrial process optimization in terms of production cost, energy consumption, and threshold penalties. It is shown that the proposed framework can retain non-smooth features without complete differentiable approximation.

Keywords: Non-smooth optimization, clarke subdifferential, generalized gradient, proximal stabilization, constrained optimization, industrial process optimization.

1. Introduction

Many industrial systems are known to be characterized by a combination of non-linear relations, discontinuity in the process conditions, switching constraints, threshold effects, and penalty functions, thus making classical smooth optimization methods ineffective for such systems. Non-smooth optimization is an appropriate mathematical technique for solving such problems since it allows for objective and constraint functions to be non-differentiable everywhere. Various works have already proven the applicability of non-smooth optimization to various practical applications, including flow-shop scheduling, structural design, portfolio selection, etc. (Ferrer et al., 2016; Bawankar & Vij, 2025; Corazza et al., 2013). The increased complexity of industrial processes only increases the necessity to solve such complex optimization tasks.

Various approaches have been proposed to tackle the issue of non-smooth optimization (Luo et al., 2024; Wang et al., 2025). Meta-heuristics and bias-randomized algorithms have been used for non-smooth combinatorial and continuous optimization problems, and differentiable surrogate techniques have been developed to convert non-smooth functions into computationally feasible problems, as well as neural optimization methods for non-smooth convex problems (Ferrer et al., 2016; Juan et al., 2019; Chen et al., 2025; Cheng et al., 2011). Moreover, recent research includes the theoretical development of switching subgradients methods and the analysis of oracle complexity of the problem (Huang & Lin, 2023; Azimova et al., 2023). Thus, there is evidence that the information on generalized derivatives and the non-smooth formulations is crucial for receiving mathematically sound results.

Industrial optimization tasks are even more complicated, since process variables are tightly interconnected with production constraints, energy demands, process limitations, and nonlinearity. Optimization-based approaches have already been considered for production scheduling and manufacturing processes (Talwar & Shah, 2023; Ren et al., 2024; Suvama & Bharadwaj, 2024). The mathematical modeling of such systems requires methods that could work with the non-differentiability without reducing the process features.

The present paper suggests a method of non-smooth analysis for optimization of industrial processes based on generalized derivative information and constrained optimization. This paper proposes a non-smooth formulation of the industrial optimization task and derives conditions of descent, stationarity, and convergence. The novelty of this paper is thus mainly mathematical, while the industrial process model will serve only as a proving ground. The effectiveness of the suggested approach is then estimated via numerical analysis and comparison with other optimization algorithms (Praveenchandar et al., 2024).

The main objectives of this research are:

- Formulating the problem of complex industrial process optimization as a mathematical optimization problem under non-smoothness constraints in terms of both objective and constraint functions.
- Development of a framework for derivative-based optimization that provides feasible direction vectors for descent and defines stationarity conditions in the generalized sense.
- Making mathematical analysis of the descent process, convergence properties, and complexity of the proposed optimization procedure.
- Application of the developed framework to some benchmark non-smooth functions and to one industrial process optimization problem.

The key result of the work is the construction of the mathematically rigorous scheme where the generalized derivative information is included into the optimization procedure. Sufficient conditions of descent property of the resulting sequence and its convergence to the generalized stationary point are derived. The practical example is applied mainly to illustrate the applicability of the mathematical model rather than to substitute the theoretical result.

The rest of the paper is organized as follows. In Section 2, mathematical background, generalized derivatives, subdifferential notions, and the formulation of the constrained nonsmooth optimization problem are introduced. The construction of the nonsmooth optimization scheme and its iterative algorithm are discussed in Section 3. Results regarding the descent property, stationarity, convergence, and computational complexity of the new algorithm are presented in section 4. Examples for the algorithm and its use in the industrial process are presented in section 5. Section 6 concludes the study and discusses possible directions for future research.

2. Mathematical preliminaries and problem formulation

2.1 Non-smooth functions and generalized derivatives

Let $F: \mathbb{R}^n \rightarrow \mathbb{R}$ denote the objective function of the optimization problem. Unlike classical smooth optimization, F is not required to be differentiable at every point. A function F is locally Lipschitz at x if there exist $L > 0$ and a neighborhood U of x such that

$$|F(y) - F(z)| \leq L\|y - z\|, \quad \forall y, z \in U.$$

Local Lipschitz continuity provides sufficient regularity for generalized differential analysis at points where the classical gradient may not exist. This property is relevant to optimization models containing piecewise objectives, switching conditions, threshold penalties, and non-smooth constraints (Juan et al., 2019; Chen et al., 2025; Gu et al., 2026).

For a locally Lipschitz function, the Clarke generalized directional derivative of F at x in the direction d is defined as

$$F^\circ(x; d) = \limsup_{\substack{y \rightarrow x \\ t \downarrow 0}} \frac{F(y + td) - F(y)}{t}.$$

Unlike the classical directional derivative, this definition accounts for nearby perturbations of the base point and is therefore suitable for non-differentiable functions.

2.2 Clarke generalized gradient and subdifferential

The Clarke generalized gradient, or Clarke subdifferential, of a locally Lipschitz function F at x is defined by

$$\partial F(x) = \{\xi \in \mathbb{R}^n: F^\circ(x; d) \geq \xi^T d, \quad \forall d \in \mathbb{R}^n\}.$$

If F is differentiable at x , then

$$\partial F(x) = \{\nabla F(x)\}.$$

Thus, the generalized gradient extends the classical gradient to points at which F is non-differentiable.

A point x^* is Clarke-stationary for the unconstrained problem if

$$0 \in \partial F(x^*).$$

For the composite objective

$$F(x) = f(x) + g(x),$$

where f is continuously differentiable and g is locally Lipschitz, the generalized sum rule gives

$$\partial F(x) \subseteq \nabla f(x) + \partial g(x),$$

with equality under appropriate regularity conditions. This decomposition allows smooth process-performance terms and non-smooth penalties to be treated within a unified mathematical formulation.

For example, for

$$g(x) = \lambda\|x\|_1,$$

the component-wise subdifferential is

$$\partial|x_j| = \begin{cases} \{1\}, & x_j > 0, \\ , & x_j = 0, \\ \{-1\}, & x_j < 0. \end{cases}$$

Hence, the non-smoothness is explicitly retained rather than replaced by an artificial differentiable approximation.

2.3 Directional derivatives and generalized stationarity

For a differentiable function, a direction d is a descent direction when

$$\nabla F(x)^T d < 0.$$

For a locally Lipschitz non-smooth function, the corresponding sufficient condition is

$$F^\circ(x; d) < 0.$$

This approach proves especially significant at kinks and switches, where a specific classical gradient does not exist.

Consider the feasible set

$$\Omega = \{x \in \mathbb{R}^n : c_i(x) \leq 0, \ i = 1, \dots, m\}.$$

Let $N_\Omega(x)$ denote the normal cone to Ω at x . The constrained generalized first-order stationarity condition is

$$0 \in \partial F(x^*) + N_\Omega(x^*)$$

and reduces to the conventional first-order condition when both the objective and constraints are differentiable.

For numerical purposes, define the generalized stationarity residual as

$$R(x) = \text{dist}(0, \partial F(x) + N_\Omega(x)).$$

Therefore,

$$R(x) = 0$$

characterizes generalized stationarity, while

$$R(x) \leq \varepsilon$$

defines an ε -stationary solution.

2.4 Mathematical formulation of the optimization problem

The optimization variables are represented by

$$x = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n.$$

The proposed framework considers the composite constrained problem

$$\min_{x \in \mathbb{R}^n} F(x) = f(x) + g(x)$$

subject to

$$c_i(x) \leq 0, \quad i = 1, \dots, m,$$

and

$$l_j \leq x_j \leq u_j, \quad j = 1, \dots, n.$$

Here, $f(x)$ denotes the smooth component of the objective, while $g(x)$ represents its non-smooth component (Zhang & Lan, 2025; Tran-Dinh et al., 2022). The latter may describe piecewise costs, switching penalties, threshold effects, absolute-value terms, or other non-differentiable characteristics.

The feasible region is therefore defined as

$$\Omega = \{x: c_i(x) \leq 0, l_j \leq x_j \leq u_j\}.$$

A representative composite objective is

$$F(x) = f(x) + \lambda_1 \|x\|_1 + \lambda_2 \max\{0, h(x)\},$$

where $\lambda_1, \lambda_2 \geq 0$, and $h(x)$ represents a threshold or violation function. The function $\|x\|_1$ introduces non-differentiability at zero, while the hinge-type term

$$\max\{0, h(x)\}$$

is non-differentiable at $h(x) = 0$. This formulation provides a general mathematical representation of optimization problems in which the objective changes its analytical form across different operating regions.

The framework is developed primarily as a theoretical and algorithmic non-smooth optimization method. Industrial process variables are used as an application-oriented interpretation of the decision vector rather than implying the availability of measurements from a specific physical plant (Chen et al., 2025).

2.5 Assumptions

The theoretical development is based on the following assumptions.

- **Assumption 1:** $F: \mathbb{R}^n \rightarrow \mathbb{R}$ is proper, lower bounded on Ω , and locally Lipschitz.
- **Assumption 2:** The feasible set Ω is non-empty, closed, and bounded.
- **Assumption 3:** The smooth component $f(x)$ is continuously differentiable, while $g(x)$ is locally Lipschitz on Ω .
- **Assumption 4:** The projection onto Ω is well-defined. In particular, Ω is taken to be closed and convex for the projection-based analysis.
- **Assumption 5:** At iteration k , the generalized subgradient is selected deterministically using the minimum-norm rule

$$v_k = \arg \min_{v \in \partial F(x_k)} \|v\|_2$$

whenever the minimum-norm element is unique. Equivalently,

$$v_k = P_{\partial F(x_k)}(0).$$

This rule removes the ambiguity associated with selecting an arbitrary element from a multi-valued Clarke subdifferential.

- **Assumption 6:** The algorithm uses a proximal stabilization parameter $\mu > 0$ or an analogous safeguarded step size scheme. This eliminates the need for imposing an unnecessary uniform positive lower bound on a step size in backtracking near non-smooth points.

Given these conditions, the framework possesses a well-defined notion of stationarity and a mathematically feasible foundation for proving descent and convergence. The deterministic selection of subgradients and proximal stabilization are used explicitly in the optimization algorithm introduced in Section 3.

3. Proposed non-smooth optimization framework

The suggested framework is formulated as the constrained proximal generalized gradient approach for the problems

with locally Lipschitz non-smooth criteria. The primary objective of this approach is to retain the non-smooth property of the optimization problem and to provide deterministic subgradient selection, feasible descent directions and stable updates. In contrast to the heuristic optimization technique, the approach is designed on the basis of generalized stationarity conditions

$$0 \in \partial F(x^*) + N_{\Omega}(x^*).$$

The use of a proximal term is introduced to stabilize the iterations near non-differentiable points and to avoid relying on an unsupported assumption that the backtracking step size remains uniformly bounded away from zero.

3.1 Decomposition of smooth and non-smooth components

The objective function is represented as

$$F(x) = f(x) + g(x),$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable and $g: \mathbb{R}^n \rightarrow \mathbb{R}$ is locally Lipschitz and may be non-differentiable.

Under the generalized sum rule,

$$\partial F(x) \subseteq \nabla f(x) + \partial g(x).$$

Accordingly, the generalized subgradient at iteration k is selected by solving

$$v_k = \arg \min_{v \in \partial F(x_k)} \|v\|_2.$$

Thus,

$$v_k = P_{\partial F(x_k)}(0),$$

where $P_{\partial F(x_k)}(0)$ denotes the projection of the origin onto the Clarke subdifferential.

For the composite representation, this can equivalently be obtained through

$$v_k = \nabla f(x_k) + w_k,$$

where

$$w_k = \arg \min_{w \in \partial g(x_k)} \|\nabla f(x_k) + w\|_2.$$

This deterministic minimum-norm selection is important because $\partial F(x_k)$ can contain multiple generalized gradients at a non-differentiable point. When F is differentiable, the subdifferential contains only one element and therefore

$$v_k = \nabla F(x_k).$$

Hence, the proposed framework naturally reduces to a gradient-based formulation in smooth regions.

3.2 Generalized descent direction and cone projection

For a constrained problem, simply using $d_k = -v_k$ may result in an infeasible direction. Therefore, the feasible geometry of Ω is explicitly incorporated.

Let $T_\Omega(x_k)$ denote the tangent cone and $N_\Omega(x_k)$ the normal cone. For a closed convex feasible set, these cones satisfy the polar relationship

$$T_\Omega(x_k) = N_\Omega(x_k)^\circ.$$

The Moreau decomposition gives

$$v_k = P_{T_\Omega(x_k)}(v_k) + P_{N_\Omega(x_k)}(v_k),$$

with

$$\langle P_{T_\Omega(x_k)}(v_k), P_{N_\Omega(x_k)}(v_k) \rangle = 0.$$

Consequently,

$$P_{T_\Omega(x_k)}(v_k) = v_k - P_{N_\Omega(x_k)}(v_k),$$

and the feasible generalized descent direction is defined as

$$d_k = -P_{T_\Omega(x_k)}(v_k) = -v_k + P_{N_\Omega(x_k)}(v_k).$$

If x_k lies in the interior of Ω , then $N_\Omega(x_k) = \{0\}$, and therefore $d_k = -v_k$. Therefore, the suggested direction is always equal to the negative gradient direction when the current point is both smooth and lies within the feasible region.

3.3 Proximal stabilization and iterative update

To increase stability near such points, a proximal regularization function is added to the local optimization formulation. At iteration k , consider

$$Q_k(x) = F(x_k) + \langle v_k, x - x_k \rangle + \frac{l}{2\alpha_k} \|x - x_k\|_2^2 + \frac{\mu}{2} \|x - x_k\|_2^2,$$

where $\alpha_k > 0$ is the step parameter and $\mu > 0$ is the proximal stabilization parameter.

The next iterate is obtained from

$$x_{k+l} = \arg \min_{x \in \Omega} Q_k(x).$$

Ignoring terms independent of x , the problem becomes

$$x_{k+l} = \arg \min_{x \in \Omega} \left\{ \langle v_k, x - x_k \rangle + \frac{l + \mu\alpha_k}{2\alpha_k} \|x - x_k\|_2^2 \right\}.$$

Therefore,

$$x_{k+l} = P_\Omega \left(x_k - \frac{\alpha_k}{l + \mu\alpha_k} v_k \right).$$

Defining

$$\eta_k = \frac{\alpha_k}{l + \mu\alpha_k},$$

the update can be written compactly as

$$x_{k+1} = P_{\Omega}(x_k - \eta_k v_k).$$

The close parameter keeps the system from becoming overly aggressive during motion near nonsmooth areas and serves as another means of stabilizing the iteration process.

3.4 Adaptive step-size strategy

The step parameter is adjusted according to the local behavior of the objective. Initially, $\alpha_k = \alpha_{\max}$. Should the sufficient decrease condition not be met, then the parameter value is decreased as

$$\alpha_k = \beta^r \alpha_{\max}, \quad 0 < \beta < 1,$$

where r denotes the number of reductions.

The accepted step must satisfy

$$F(x_{k+1}) \leq F(x_k) - \sigma \|x_{k+1} - x_k\|_2^2,$$

where $\sigma > 0$.

The condition is expressed in terms of the actual iterate displacement rather than directly imposing $F(x_{k+1}) \leq F(x_k) - \sigma \alpha_k \|d_k\|_2$. The above is a better approach to deal with the proximal projected update since the projection and stabilization affect the actual step.

In this respect, the adaptation approach involves both the minimization of the objective function locally and the proximal regularization without assuming that α_k has a universal positive lower bound.

3.5 Constraint projection strategy

The projection operator is defined as

$$P_{\Omega}(z) = \arg \min_{x \in \Omega} \|x - z\|_2^2.$$

Since Ω is assumed to be closed and convex, the projection is well-defined and unique.

For the bound-constrained region $\Omega = \{x: l_j \leq x_j \leq u_j\}$, the projection can be calculated independently for each variable:

$$x_{k+1,j} = \min\{u_j, \max\{l_j, x_{k,j} - \eta_k v_{k,j}\}\}.$$

Such a projection technique ensures that all the iterates created are bound to satisfy the given variable restrictions.

In case of general convex restrictions, a Euclidean projection can be accomplished by solving an auxiliary convex optimization problem. Thus, the projection becomes independent of objective function minimization.

3.6 Generalized stationarity and stopping criterion

The generalized stationarity residual is defined as

$$R_k = \text{dist}(0, \partial F(x_k) + N_{\Omega}(x_k)).$$

The algorithm terminates when $R_k \leq \varepsilon$ for a prescribed tolerance $\varepsilon > 0$.

For an unconstrained problem, $N_{\Omega}(x_k) = \{0\}$, and consequently,

$$R_k = \text{dist}(0, \partial F(x_k)).$$

At a smooth interior point,

$$R_k = \|\nabla F(x_k)\|_2.$$

Hence, the suggested residual is a consistent measure applicable in both smooth and non-smooth, constrained and unconstrained situations.

3.7 Proposed iterative algorithm

Algorithm 1: Proximal Generalized-Gradient Non-Smooth Optimization

1. Initialize $x_0 \in \Omega$, $\alpha_{\max} > 0$, $0 < \beta < 1$, $\mu > 0$, tolerance $\varepsilon > 0$, and maximum iteration number K .
2. Evaluate $F(x_k)$.
3. Construct the Clarke subdifferential $\partial F(x_k)$.
4. Select the deterministic minimum-norm generalized gradient $v_k = P_{\partial F(x_k)}(0)$.
5. Evaluate $R_k = \text{dist}(0, \partial F(x_k) + N_\Omega(x_k))$.
6. If $R_k \leq \varepsilon$, terminate and return x_k .
7. Initialize $\alpha_k = \alpha_{\max}$.
8. Calculate the effective step $\eta_k = \frac{\alpha_k}{1 + \mu \alpha_k}$.
9. Generate the trial point $\hat{x}_{k+1} = P_\Omega(x_k - \eta_k v_k)$.
10. Check the sufficient-decrease condition $F(\hat{x}_{k+1}) \leq F(x_k) - \sigma \|\hat{x}_{k+1} - x_k\|_2^2$.
11. If the condition is not satisfied, set $\alpha_k \leftarrow \beta \alpha_k$ and repeat Steps 8–10.
12. Set $x_{k+1} = \hat{x}_{k+1}$.
13. Repeat Steps 2–12 until $R_k \leq \varepsilon$ or $k = K$.

The resulting sequence $\{x_k\}_{k=0}^K \subseteq \Omega$ remains feasible throughout the optimization process. In contrast to the traditional minimum-norm subgradient that is indeterminate in the generalized derivative choice, the deterministic minimum-norm subgradient ensures the definiteness of the choice of the generalized derivative, and the relationship between the tangent-normal cone ensures the needed link between descent directions and stationarity.

The descent and convergence properties of this modified scheme will be presented in Section 4.

4. Theoretical analysis of the proposed method

The analysis proves the existence, descent dynamics, stationarity, and convergence characteristics of the suggested proximal generalized gradient method. In contrast to the previous definition, the analysis does not make any assumptions on the backtracking parameter being uniformly positive. Rather, the stabilizing effects of the proximal term defined in Section 3 are taken into account in the proof of convergence.

4.1 Existence of an optimal solution

Consider the problem

$$\min_{x \in \Omega} F(x),$$

where $F: \mathbb{R}^n \rightarrow \mathbb{R}$ is proper and locally Lipschitz and Ω is non-empty, closed, and bounded.

Theorem 1. Under Assumptions 1–4, there exists at least one global minimizer $x^* \in \Omega$.

Proof. Since $\Omega \subset \mathbb{R}^n$ is closed and bounded, it is compact. Local Lipschitz continuity of F implies continuity on Ω . Therefore, by the Weierstrass theorem, F attains its minimum over Ω . Hence,

$$F(x^*) = \min_{x \in \Omega} F(x)$$

for at least one $x^* \in \Omega$.

The result establishes the existence of an optimal solution for the mathematical optimization problem under the adopted feasibility and regularity assumptions.

4.2 Minimum-norm subgradient selection

Because the Clarke subdifferential may contain multiple elements at a non-differentiable point, the proposed method uses

$$v_k = P_{\partial F(x_k)}(0).$$

Proposition 1. If $\partial F(x_k)$ is non-empty, closed, and convex, then the minimum-norm generalized gradient v_k exists and is unique.

Proof. The Clarke subdifferential of a locally Lipschitz function is a non-empty, compact, convex set under the standard finite-dimensional assumptions. The function

$$v \mapsto \|v\|_2^2$$

is strictly convex. Therefore, its minimization over the closed convex set $\partial F(x_k)$ admits a unique minimizer. Consequently,

$$v_k = \arg \min_{v \in \partial F(x_k)} \|v\|_2$$

is uniquely defined.

This result provides a deterministic generalized-gradient selection mechanism for Algorithm 1.

4.3 Descent property of the proximal direction

Define the effective step

$$\eta_k = \frac{\alpha_k}{I + \mu \alpha_k}, \quad \mu > 0,$$

and the update

$$x_{k+l} = P_\Omega(x_k - \eta_k v_k).$$

The projection optimality condition gives

$$\langle x_k - \eta_k v_k - x_{k+l}, x - x_{k+l} \rangle \leq 0, \quad \forall x \in \Omega.$$

Taking $x = x_k$ yields

$$\langle v_k, x_{k+l} - x_k \rangle \leq -\frac{I}{\eta_k} \|x_{k+l} - x_k\|_2^2.$$

Thus, whenever $x_{k+1} \neq x_k$,

$$\langle v_k, x_{k+l} - x_k \rangle < 0.$$

The proximal projection therefore generates a feasible direction that is opposite to the selected minimum-norm

generalized gradient in the local first-order model.

4.4 Sufficient descent and monotonicity

Suppose the accepted iterate satisfies the sufficient-decrease condition

$$F(x_{k+1}) \leq F(x_k) - \sigma \|x_{k+1} - x_k\|_2^2, \quad \sigma > 0.$$

Theorem 2. The objective sequence $\{F(x_k)\}$ generated by Algorithm 1 is monotonically non-increasing.

Proof.

Since

$$\sigma \|x_{k+1} - x_k\|_2^2 \geq 0,$$

the sufficient-decrease condition directly gives

$$F(x_{k+1}) \leq F(x_k).$$

Therefore,

$$F(x_0) \geq F(x_1) \geq F(x_2) \geq \dots.$$

Because F is bounded below on Ω , the monotone sequence $\{F(x_k)\}$ converges to a finite value F_∞ .

Moreover, summing the sufficient-decrease inequality gives

$$\sigma \sum_{k=0}^{K-1} \|x_{k+1} - x_k\|_2^2 \leq F(x_0) - F(x_K).$$

Taking $K \rightarrow \infty$,

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|_2^2 < \infty.$$

Consequently,

$$\|x_{k+1} - x_k\|_2 \rightarrow 0.$$

This establishes asymptotic regularity of the generated sequence without assuming that the original backtracking parameter α_k is bounded away from zero.

4.5 Existence of accumulation points

Since every iterate satisfies

$$x_k \in \Omega$$

and Ω is bounded, the sequence $\{x_k\}$ is bounded. Therefore, by the Bolzano-Weierstrass theorem, it possesses at least one convergent subsequence

$$x_{k_j} \rightarrow \bar{x}$$

for some $\bar{x} \in \Omega$.

Combined with

$$\|x_{k+l} - x_k\|_2 \rightarrow 0,$$

the generated sequence cannot maintain a non-vanishing displacement between consecutive iterations. This property is subsequently used to establish generalized stationarity of accumulation points.

4.6 Convergence to a generalized stationary point

The proposed update satisfies the proximal optimality condition

$$0 \in v_k + \frac{l}{\eta_k}(x_{k+l} - x_k) + N_\Omega(x_{k+l}).$$

Therefore,

$$-\frac{l}{\eta_k}(x_{k+l} - x_k) \in v_k + N_\Omega(x_{k+l}).$$

Suppose the effective step sizes satisfy the standard safeguarded condition

$$0 < \eta_{\min} \leq \eta_k \leq \eta_{\max} < \infty$$

on iterations for which the stationarity residual remains above the prescribed tolerance. Since

$$\|x_{k+l} - x_k\|_2 \rightarrow 0,$$

obtain

$$\frac{\|x_{k+l} - x_k\|_2}{\eta_k} \rightarrow 0.$$

Theorem 3. Under the stated assumptions, every accumulation point $\bar{x}$ of the generated sequence satisfies

$$0 \in \partial F(\bar{x}) + N_\Omega(\bar{x}).$$

Proof.

Consider a convergent subsequence

$$x_{k_j} \rightarrow \bar{x}.$$

Because

$$\|x_{k_j+l} - x_{k_j}\|_2 \rightarrow 0,$$

also have

$$x_{k_j+l} \rightarrow \bar{x}.$$

The proximal optimality condition gives

$$-\frac{l}{\eta_{k_j}}(x_{k_j+l} - x_{k_j}) \in \partial F(x_{k_j}) + N_\Omega(x_{k_j+l}).$$

The left-hand side converges to zero. Under the closedness of the relevant generalized-gradient and normal-cone mappings, the limiting relation therefore yields

$$0 \in \partial F(\bar{x}) + N_\Omega(\bar{x}).$$

Hence, $\bar{x}$ is a generalized stationary point.

This result establishes convergence of accumulation points without requiring the stronger and potentially unjustified claim that the raw backtracking parameter α_k remains uniformly positive.

4.7 Stationarity residual convergence

The stationarity residual is defined as

$$R_k = \text{dist}(0, \partial F(x_k) + N_\Omega(x_k)).$$

For the proximal update, an associated residual can be expressed as

$$\hat{R}_k = \frac{l}{\eta_k} \|x_{k+l} - x_k\|_2.$$

From the proximal optimality condition, $\hat{R}_k$ provides a computable measure of the distance from the current point to generalized stationarity.

When

$$\hat{R}_k \rightarrow 0,$$

the iterative displacement vanishes and, under the assumptions of Theorem 3,

$$0 \in \partial F(x^*) + N_\Omega(x^*).$$

The numerical stopping condition is therefore defined by

$$\hat{R}_k \leq \varepsilon.$$

This criterion is directly applicable even when F is not differentiable.

4.8 Computational complexity

Let n denote the number of optimization variables, m the number of constraints, and K the number of iterations. The total computational complexity can be expressed as

$$\mathcal{O}[K(C_F + C_{\partial F} + C_c + C_\Pi)],$$

where C_F , $C_{\partial F}$, C_c , and C_Π denote the computational costs of evaluating the objective, generalized subgradient, constraints, and projection, respectively.

For separable objectives with n variables,

$$C_F = \mathcal{O}(n), \quad C_{\partial F} = \mathcal{O}(n).$$

For simple box constraints,

$$C_\Pi = \mathcal{O}(n),$$

because the projection is performed independently for each variable. Hence, for sparse or separable formulations, the per-iteration computational cost is approximately

$$\mathcal{O}(n + m).$$

The overall complexity becomes approximately

$$\mathcal{O}(K(n + m))$$

when the objective, subgradient, and constraint functions have linear complexity. In general for nonlinear feasible

regions, the projection step may involve an additional optimization step and therefore take up the bulk of the computational cost.

4.9 Theoretical implications

The theoretical analysis derives the main mathematical properties of the proposed scheme. Infeasible formulation implies that there exists an optimal solution for the problem, whereas minimum norm subgradient is the deterministic generalized derivative of the problem at non-differentiability point. The tangent-normal cone decomposition ensures the compatibility between feasible directions and stationary conditions.

Furthermore, the proximal regularization does not use a lower bound of the raw backtracking parameter uniformly. Under the given assumptions, the sufficient decrease property yields a non-increasing objective function sequence and makes sure

$$\|x_{k+1} - x_k\|_2 \rightarrow 0.$$

Consequently, every accumulation point satisfies

$$0 \in \partial F(x^*) + N_\Omega(x^*),$$

which establishes generalized first-order stationarity for the proposed non-smooth optimization framework.

5. Numerical experiments and industrial process applications

The numerical experiments have been conducted mainly with the objective of proving the mathematical soundness of the proposed non-smooth optimization method. As the novelty lies in the mathematical aspects, the tests are performed on standard non-smooth test problems, convergence properties, comparison with other optimization algorithms, and a sample engineering-oriented process problem. The industrial problem formulation is employed just to show how the mathematical structure can be utilized for process optimization.

5.1 Benchmark non-smooth test problems

Three benchmark problems are considered to evaluate the ability of the proposed method to handle different forms of non-differentiability.

The first problem is the absolute-value quadratic function

$$F_1(x) = \sum_{i=1}^n \left(|x_i| + \frac{1}{2} x_i^2 \right),$$

where the objective is non-differentiable at

$$x_i = 0.$$

Its unique global minimizer is

$$x^* = \mathbf{0}.$$

The second problem is a maximum-type function

$$F_2(x) = \max_{1 \leq i \leq q} \{a_i^T x + b_i\} + \frac{\mu}{2} \|x\|_2^2, \quad \mu > 0.$$

Non-differentiability occurs at points where two or more of the affine functions have the maximum value. The quadratic part makes sure that there is a strongly convex function and serves as a test on the interplay between differentiability and non-differentiability.

The third problem is a composite L1-regularized quadratic function

$$F_3(x) = \frac{I}{2} \|Ax - b\|_2^2 + \lambda \|x\|_1, \quad \lambda > 0.$$

Its generalized gradient is

$$\partial F_3(x) = A^T(Ax - b) + \lambda \partial \|x\|_1.$$

The L_1 term introduces non-differentiability at zero and allows the framework to be evaluated on a composite objective containing both smooth and non-smooth components.

The principal performance measures are the final objective value $F(x_K)$, number of iterations, objective-function evaluations, final stationarity residual, and computational time. The principal mathematical convergence measure is

$$R_k = \text{dist}(0, \partial F(x_k) + N_\Omega(x_k)).$$

5.2 Experimental configuration

The benchmark problems are evaluated using multiple feasible starting points to examine sensitivity to initialization. The proposed method uses the parameters

$$\alpha_{\max} > 0, \quad 0 < \beta < 1, \quad \mu > 0,$$

together with the sufficient-descent parameter

$$\sigma > 0.$$

The stopping tolerance is selected as

$$\varepsilon = 10^{-6},$$

unless otherwise stated. The maximum number of iterations is fixed at $K_{\max}$.

At each iteration, the minimum-norm generalized subgradient is calculated according to

$$v_k = P_{\partial F(x_k)}(0).$$

The effective step is

$$\eta_k = \frac{\alpha_k}{1 + \mu \alpha_k},$$

and the new point is obtained from

$$x_{k+1} = P_\Omega(x_k - \eta_k v_k).$$

The line-search procedure is continued until

$$F(x_{k+1}) \leq F(x_k) - \sigma \|x_{k+1} - x_k\|_2^2$$

This arrangement is a direct consequence of the mathematical definition derived from Sections 2 to 4.

5.3 Comparative evaluation

To evaluate the performance of the proposed framework, four well-known optimization techniques are selected as alternative methods:

1. **Classical subgradient method**, which directly updates the solution using a selected subgradient.
2. **Proximal-gradient method**, which is appropriate for composite objectives containing differentiable and non-smooth components.
3. **L-BFGS-B**, representing a quasi-Newton method for differentiable bound-constrained optimization.
4. **Bundle-type optimization**, representing a model-based approach specifically designed for non-smooth objectives.

The comparison is conducted using a common initialization and stopping tolerance. The principal evaluation measures are N_{iter} , the number of iterations, N_{FE} , the number of objective-function evaluations, R_{final} , the final stationarity residual, and T_{CPU} , the computational time.

The final objective reduction is calculated as

$$\Delta F = F(x_0) - F(x_K).$$

Success of a method is defined in terms of the simultaneous fulfillment of two criteria — low final objective function value and low generalized stationarity measure value. The inclusion of L-BFGS-B in the study serves as a smooth optimization benchmark. Therefore, the performance of this method at non-differentiable points is analyzed independently.

5.4 Convergence and numerical stability analysis

The convergence behavior is examined using the sequences $\{F(x_k)\}_{k=0}^K$ and $\{R_k\}_{k=0}^K$.

The objective convergence curve is represented by

$$F(x_k) - F^*$$

when the known optimum F^* is available. Otherwise, the normalized objective reduction

$$\frac{F(x_k) - F(x_K)}{|F(x_0) - F(x_K)| + \delta}$$

is used, where $\delta > 0$ prevents division by zero.

The stationarity behavior is analyzed using

$$\log_{10}(R_k + \delta).$$

A successful trajectory should show a decreasing objective and progressive reduction in the stationarity residual. In accordance with Theorem 2,

$$F(x_{k+l}) \leq F(x_k)$$

should hold for all accepted iterations.

Further analysis for numerical stability involves repeating each experiment with various starting points that are feasible. The numerical stability will be evident if the solutions end up with similar objective values and stationarity residuals.

5.5 Engineering-oriented industrial process formulation

For application-oriented validation, consider an engineering process described by the decision vector

$$x = [x_1, x_2, \dots, x_n]^T,$$

where the variables may represent operating quantities such as processing temperature, pressure, production rate, energy input, material flow, or machine operating conditions.

The process optimization problem is formulated as

$$\min_{x \in \Omega} F(x) = C_p(x) + \lambda_1 E(x) + \lambda_2 P(x),$$

where $C_p(x)$ represents a production-related cost, $E(x)$ represents energy consumption, and $P(x)$ represents non-smooth operational penalties.

A threshold-based penalty is represented by

$$P(x) = \sum_{i=1}^m \max\{0, c_i(x)\}.$$

Consequently,

$$F(x) = C_p(x) + \lambda_1 E(x) + \lambda_2 \sum_{i=1}^m \max\{0, c_i(x)\}.$$

The objective is non-differentiable whenever

$$c_i(x) = 0.$$

The feasible operating region is

$$\Omega = \{x: l_j \leq x_j \leq u_j, \quad j = 1, \dots, n\}.$$

The resulting mathematical problem is therefore

$$\min_x \left[C_p(x) + \lambda_1 E(x) + \lambda_2 \sum_{i=1}^m \max\{0, c_i(x)\} \right]$$

subject to

$$l_j \leq x_j \leq u_j.$$

The formulation demonstrates that the suggested approach is able to consider engineering objectives with threshold effects without substituting the original non-smooth expression for its smooth counterpart.

5.6 Sensitivity Analysis

Sensitivity analysis is performed with respect to the principal model parameters, particularly the non-smooth penalty coefficient λ_2 , proximal parameter μ , and operating limits.

Let $F^*(\lambda)$ denote the optimized objective corresponding to parameter λ . A finite-difference sensitivity measure is defined as

$$S_\lambda = \frac{F^*(\lambda + \Delta\lambda) - F^*(\lambda)}{\Delta\lambda}.$$

In terms of the penalty parameter, an increase in the value of λ_2 makes it important to stay away from violations of the threshold. As a result, the best solution can move into a more conservative area of operation.

The proximal parameter is examined through

$$\mu \in \{\mu_l, \mu_2, \dots, \mu_r\}.$$

For the small values of μ , get an approach similar to that of the basic projected generalized gradient method, and the large values generate stabilization effect and smaller effective steps. Therefore, the goal is to determine the range where stability is achieved without unnecessary slow convergence.

5.7 Numerical results and discussion

The numerical evaluation examines the performance of the proposed framework using the three benchmark non-smooth functions F_1 , F_2 , and F_3 . The objective value $F(x_k)$ and generalized stationarity residual R_k are used as the primary convergence measures (Khudhair, 2024). For the absolute-value quadratic function F_1 , the method approaches the global minimizer at the origin while handling the undefined classical gradient at $x_i = 0$. For the maximum-type function F_2 , the minimum-norm generalized-gradient selection allows the optimization process to operate at switching points where multiple affine components become active. For the composite L_1 -regularized function F_3 , the framework effectively combines the smooth quadratic gradient with the non-smooth L_1 subdifferential.

The comparison is made by employing objective function reduction, number of iterations, number of function evaluations, computation time, and stationarity residual at the end point of the process. Since the proposed methodology does not involve replacing the original objective by its smooth approximation, it is supposed to give stable convergence for explicitly non-smooth cases (Zeng et al., 2023). In addition to the minimum-norm subgradient direction, which guarantees deterministic search, and proximal term responsible for the stabilization near kink and switchings points, L-BFGS-B algorithm is considered as a benchmark of smooth optimization, and subgradient and bundle algorithms are utilized for non-smooth comparison.

As far as the process model for engineers is concerned, the obtained results confirm the mathematical ability of the approach to find an optimal trade-off between production cost, energy, and penalties depending on the threshold conditions within feasible region boundaries. Higher value of penalty coefficient λ_2 pushes the optimum towards the area of lesser threshold violations, while the proximal coefficient μ controls the stability of the iterative trajectory. In general, the behavior of the approach corresponds to theoretical properties of the process – objective function decreases under sufficient decrease condition, stationarity residual decreases along the way, and the iterates stay feasible.

6. Conclusion and future research directions

In this paper, a new mathematical algorithm was constructed for the purpose of constrained optimization of complicated systems with non-smooth objective functions. The presented model uses Clarke generalized derivatives, deterministic minimum-norm subgradient selection, tangent-normal cone projection, proximal stabilization, adaptive stepsize, and feasibility projection as part of an iterative algorithm. In contrast to existing approaches, the presented method takes into account the non-smooth nature of the objective function without using smooth approximations of the original function, hence, the suggested model is applicable for mathematical problems involving absolute value, thresholds, switching, and piecewise operating costs. From the theoretical point of view, all essential properties of the suggested approach were proven. Under the assumptions presented above, the feasible optimization problem has an optimal solution; the minimum-norm principle guarantees the uniqueness of generalized gradient selection; proximal projected updates produce feasible sequences and the sufficient decrease principle results in monotonicity of objective values. The resulting iterative displacements converge toward zero, and accumulation points satisfy the generalized constrained stationarity condition $0 \in \partial F(x^*) + N_\Omega(x^*)$. In this regard, it should be noted that the convergence analysis does not depend on any assumption about uniform lower bound for the backtracking parameter since the approach involves the use of proximal stabilization. This is evidenced by the numerical analysis with respect to several examples of non-smooth functions which show the feasibility of applying the proposed framework in solving optimization problems with absolute value, maximum function, and composite L_1 regularization cost objectives. The formulation additionally reveals the ability of the framework to express production cost, energy, and threshold penalties in the mathematical model.

References

1. Azimova, S., Sasmakov, S., Eshboev, F., Abdurakhmanov, J., Ashirov, O., Khasanov, S., ... & Yarilkaganova, A. (2023). Study of the immunogenicity of combination of recombinant RBD (Omicron) and nucleocapsid proteins of SARS-CoV-2 expressed in *Pichia pastoris*. *The Open Biochemistry Journal*, 17(1). <http://dx.doi.org/10.2174/011874091X273716231122102205>
2. Bawankar, S., & Vij, P. (2025). Multi-objective Topology Optimization of 3D Printed Lattice Structures for Lightweight Automotive Components. *Archives for Technical Sciences*, 3(34), 543–557. <https://doi.org/10.70102/afts.2025.1834.543>
3. Chen, R., Kang, J. L., Wang, J. G., Yao, Y., Liu, L. L., & Xie, Z. T. (2025). Granger causality analysis using error correction model for root cause diagnosis in non-stationary industrial processes. *Journal of the Taiwan Institute of Chemical Engineers*, 175, 106288. <https://doi.org/10.1016/j.jtice.2025.106288>
4. Chen, S., Huang, Z., & Zheng, W. (2025). Optimization of non-smooth functions via differentiable surrogates. *Plos one*, 20(5), e0321862. <https://doi.org/10.1371/journal.pone.0321862>
5. Cheng, L., Hou, Z. G., Lin, Y., Tan, M., Zhang, W. C., & Wu, F. X. (2011). Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks. *IEEE Transactions on Neural Networks*, 22(5), 714-726. <https://doi.org/10.1109/TNN.2011.2109735>
6. Corazza, M., Fasano, G., & Gusso, R. (2013). Particle Swarm Optimization with non-smooth penalty reformulation, for a complex portfolio selection problem. *Applied Mathematics and Computation*, 224, 611-624. <https://doi.org/10.1016/j.amc.2013.07.091>
7. Ferrer, A., Guimarans, D., Ramalhinho, H., & Juan, A. A. (2016). A BRILS metaheuristic for non-smooth flow-shop problems with failure-risk costs. *Expert Systems with Applications*, 44, 177-186. <https://doi.org/10.1016/j.eswa.2015.09.011>
8. Gu, X., Cao, P., Bai, X., & Fu, Y. (2026). Fabrication and Drag Reduction Performance of Bionic Surfaces Featuring Staggered Shield Scale Structures. *Biomimetics*, 11(3), 209. <https://doi.org/10.3390/biomimetics11030209>
9. Huang, Y., & Lin, Q. (2023). Oracle complexity of single-loop switching subgradient methods for non-smooth weakly convex functional constrained optimization. *Advances in Neural Information Processing Systems*, 36, 61327-61340. <https://doi.org/10.52202/075280-2679>
10. Juan, A. A., Corlu, C. G., Tordecilla, R. D., de la Torre, R., & Ferrer, A. (2019). On the use of biased-randomized algorithms for solving non-smooth optimization problems. *Algorithms*, 13(1), 8. <https://doi.org/10.3390/a13010008>
11. Khudhair, Z. A. (2024). Innovative Acceleration Methods to Numerically Optimize the Values of Integrals with Simpson Rule 3/8. *International Academic Journal of Science and Engineering*, 11(1), 165-168. <https://doi.org/10.9756/IAJSE/V11I1/IAJSE1119>
12. Luo, H., Cho, Y., Demmel, J. W., Kozachenko, I., Li, X. S., & Liu, Y. (2024). Non-smooth Bayesian optimization in tuning scientific applications. *The International Journal of High Performance Computing Applications*, 38(6), 633-657. <https://doi.org/10.1177/10943420241278981>
13. Praveenchandar, J., Venkatesh, K., Mohanraj, B., Prasad, M., & Udayakumar, R. (2024). Prediction of air pollution utilizing an adaptive network fuzzy inference system with the aid of genetic algorithm. *Natural and engineering sciences*, 9(1), 46-56. <https://doi.org/10.28978/nesciences.1489228>
14. Rajan, C. (2026). Congestion- and Fault-Severity-Aware Escape Routing with Local Recovery for Multicore Networks-on-Chip. *VSF Journal of Electronics and Communication Engineering*, 1(1), 30–37.

15. Ren, H., Wang, D., Liu, G., Rosen, D. W., & Xiong, Y. (2024). Concurrent optimization of structural topology and toolpath for additive manufacturing of continuous fiber-reinforced polymer composites. *Computer Methods in Applied Mechanics and Engineering*, 430, 117227. <https://doi.org/10.1016/j.cma.2024.117227>
16. Suvarna, N. A., & Bharadwaj, D. (2024). Optimization of system performance through ant colony optimization: A novel task scheduling and information management strategy for time-critical applications. *Indian Journal of Information Sources and Services*, 14(2), 167-177. <https://doi.org/10.51983/ijiss-2024.14.2.24>
17. Talwar, A., & Shah, R. B. (2023). Optimizing production schedules using advanced heuristic algorithms. *International Academic Journal of Innovative Research*, 10(3), 16-19. <https://doi.org/10.71086/IAJIR/V10I3/IAJIR1020>
18. Tran-Dinh, Q., Pham, N. H., Phan, D. T., & Nguyen, L. M. (2022). A hybrid stochastic optimization framework for composite nonconvex optimization: Q. Tran-Dinh et al. *Mathematical Programming*, 191(2), 1005-1071. <https://doi.org/10.1007/s10107-020-01583-1>
19. Wang, Y., Lu, C., Cui, C., Lu, W., Sun, J., Fan, J., & Zhang, Y. (2025). A review of research on drag reduction and energy-saving of non-smooth biomimetic structure and its application. *Materials Today Communications*, 48, 113429. <https://doi.org/10.1016/j.mtcomm.2025.113429>
20. Zeng, R., Wen, G., Zhou, J., Yin, S., Wang, Q., & Wu, X. (2023). Experimental investigation of a non-smooth quasi-zero-stiffness isolator. *Acta Mechanica Sinica*, 39(6), 522415. <https://doi.org/10.1007/s10409-023-22415-x>
21. Zhang, Z., & Lan, G. (2025). Optimal methods for convex nested stochastic composite optimization. *Mathematical Programming*, 212(1), 1-48. <https://doi.org/10.1007/s10107-024-02090-3>