Global convergence of a coderivative-based regularized Newton method with damping for nonsmooth optimization

Wei Ouyang* Zhenghong Tan† Jiangxing Zhu‡

Abstract

In this paper, we propose and analyze a globally convergent regularized Newton method with positive definite regularization for solving nonsmooth optimization problems. Our approach leverages the coderivative-generated second-order subdifferential (generalized Hessian) and replaces the identity matrix in traditional algorithms with a general positive-definite symmetric matrix to regularize the generalized Hessian. By appropriately selecting the regularization matrix, we enhance the practical performance of the algorithm. Under suitable assumptions, we establish the well-posedness of the proposed algorithm. Using tools from variational analysis and generalized differentiation, we derive explicit convergence rates under the Hölder strong metric subregularity condition. Specifically, we quantify the precise relationship between the algorithm's convergence rate and the order of Hölder strong metric subregularity. For a class of nonsmooth functions, namely prox-regular functions, corresponding algorithms have also been developed via their Moreau envelopes. As an application, we apply the proposed method to convex composite optimization problems within the forward–backward envelope framework. Numerical experiments on Lasso problems demonstrate that our algorithm outperforms some existing methods.

Key Words. Regularized Newton method; Second-order subdifferential; Hölder strong metric subregularity; Prox-regular functions; Convex composite optimization

AMS subject classification: 90C31; 49J52; 49J53.

1 Introduction

It is well recognized that the classical Newton's method provides a highly efficient algorithm for solving C^2-smooth unconstrained optimization problems of the form

$$\text{minimize } f(x) \text{ subject to } x \in \mathbb{R}^n \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ denotes the objective function. The classical Newton's method constructs the iterative procedure

$$x^{k+1} := x^k + d^k \quad \text{for all } k \in \mathbb{N} := \{1, 2, \dots\} \tag{2}$$

*School of Mathematics, Yunnan Normal University, Kunming 650500, People's Republic of China; Yunnan Key Laboratory of Modern Analytical Mathematics and Applications, Kunming 650500, China. Email: weiouyangxe@hotmail.com.

†School of Mathematics, Yunnan Normal University, Kunming 650500, People's Republic of China. Email: tanzhenghong@ynnu.edu.cn.

‡Department of Mathematics, Yunnan University, Kunming 650500, People's Republic of China. Email: jxzhu@ynu.edu.cn.

with $x^0 \in \mathbb{R}^n$ being a given starting point. The search directions d^k are determined by solving the linear system

$$-\nabla f(x^k) = \nabla^2 f(x^k)\, d^k, \quad k = 0, 1, \ldots \tag{3}$$

where $\nabla^2 f(x^k)$ represents the Hessian matrix of f at x^k. Newton iterations exhibit local quadratic convergence when the Hessian matrix $\nabla^2 f(\bar{x})$ is positive definite at the reference solution $\bar{x}$ and the starting point x^0 is chosen sufficiently close to $\bar{x}$; see, for example, [5, 10].

To ensure the global convergence of Newton's method, the Newton direction is typically combined with various line search strategies. These strategies are implemented via iterative procedures of the form

$$x^{k+1} := x^k + \tau_k\, d^k \quad \text{for all } k \in \mathbb{N} \tag{4}$$

with a step size $\tau_k \geq 0$. If $\nabla^2 f(x^k)$ is positive-definite, then algorithm (4) equipped with a backtracking line search is referred to as the *damped Newton method* [5], distinguishing it from the pure Newton method that employs a fixed step size. When $\nabla^2 f(x^k)$ is merely positive-semidefinite, $\nabla^2 f(x^k)$ in (3) is often replaced by a regularized Hessian $\nabla^2 f(x^k) + \rho_k I$, where the sequence $\{\rho_k\}$ is commonly set as $\rho_k := c\,\|\nabla f(x^k)\|$ for some constant $c > 0$. The corresponding algorithm is termed the *regularized Newton method*. For numerous intriguing results in this regard, we refer the reader to [8, 17].

For solving second-order nonsmooth optimization problems—i.e., optimization problem (1) where the objective function belongs to the $C^{1,1}$ class (continuously differentiable with Lipschitz continuous gradients) but is not twice continuously differentiable—the semismooth Newton method stands out as one of the most popular Newton-type algorithms. The core idea of this method is to replace the Hessian matrix in the classical Newton method with Clarke's generalized Jacobian of Lipschitzian mappings. Under the assumption that the generalized Jacobian is nonsingular, both local and global convergence results for the method have been established, as demonstrated in [9, 10, 12, 25]. Another approach involves replacing the classical Hessian matrix $\nabla^2 f$ of C^2-smooth functions with the generalized Hessian matrix (or second-order subdifferential) $\partial^2 f$. Proposed by Mordukhovich [19] as the coderivative of the subgradient mapping, this generalized Hessian matrix exhibits wide applications in variational analysis and optimization theory; see Section 2 for details.

Gfrerer and Outrata [11] recently proposed a pure Newton-type algorithm for solving generalized equations by virtue of the coderivative. For $C^{1,1}$ objective functions, Mordukhovich and Sarabi [20] developed a generalized Newton method based on the generalized Hessian matrix to find tilt-stable minimizers of the optimization problem (1). They further extended these findings to a class of general prox-regular functions via the Moreau envelope. The local superlinear convergence of the aforementioned Newtonian algorithms, established in [11, 20], hinges on the semismooth* assumption imposed on the associated mapping. First introduced in [11], the semismooth* property constitutes a relaxed variant of the classical semismoothness. Khanh et al. [15] constructed a damped Newton method based on the generalized Hessian matrix and established its global convergence under the positive definiteness assumption of the generalized Hessian matrix $\partial^2 f$. The positive definiteness of $\partial^2 f$ implies that the gradient mapping ∂f is strongly metrically regular [23, Theorem 5.13]. However, strong metric regularity is a relatively restrictive assumption, imposing certain limitations in applications. Subsequently, Khanh et al. [13] proposed the generalized regularized Newton method (GRNM), which is well-defined even when the generalized Hessians are merely positive-semidefinite. Under the additional assumption of metric regularity, the convergence rate achieved is at least linear. Furthermore, combined with the semismooth* property of the gradient mapping ∇f, the GRNM is guaranteed to have a Q-superlinear convergence rate. Recently, Shi and Chao [33] further generalized the GRNM. For the nonsmooth case, Khanh et al. [14] constructed a generalized Newton method to solve subgradient inclusions.

Under the positive definiteness assumption, or under the assumptions of positive semi-definiteness and metric regularity, the algorithms studied in [13, 14, 15, 20] typically converge to the tilt-stable minimizer of (1), a special type of minimizer (see [30]). Recently, Aragón et al. [1] and Shi and Chao [33] investigated the convergence of Newton algorithms based on the generalized Hessian matrix under a weaker metric subregularity assumption.

In this paper, to solve $C^{1,1}$ optimization problems, we propose a generalized positive definite regularized Newton method based on the generalized Hessian matrix by replacing the identity matrix in the regularized Newton method with a positive definite matrix. This algorithmic framework includes the generalized regularized Newton method in references [13, 15, 33]. By appropriately selecting the positive definite matrix in the regularization process, our algorithm can have broader applicability in practice. The proposed positive definite regularized Newton method (PDRNM) generally does not require the generalized Hessian matrix $\partial^2 f$ to be positive definite or positive semi-definite; we construct this algorithm under the more general ξ-lower-definite assumption and establish its well-posedness and global convergence to a stationary point of f. To obtain the convergence rate results of PDRNM, only the Hölder strong metric subregularity of the gradient mapping is additionally imposed, without the need for the semismooth* property, semismoothly differentiable property, and other properties that are additionally required in references [1, 15, 13, 33]. It is worth noting that the metric subregularity property is weaker than the metric regularity property imposed in reference [13] for solving tilt-stable minimizers. We also present the exact quantitative relationship between the convergence rate of the studied algorithm and the order of Hölder strong metric subregularity. Another difference from the generalized regularized Newton method is that ρ_k in the regularized Hessian of the algorithm iteration process can be prespecified, rather than being generated passively by $\|\nabla f(x^k)\|$, which can reduce the amount of computation. Of course, ρ_k needs to satisfy a corresponding relationship with ξ in our basic assumption of ξ-lower-definite; see Section 3 for details. The advantage of this is that we can adjust the value of the parameter ρ_k as needed.

The rest of this paper is organized as follows. In Section 2, we present some commonly used notations as well as basic concepts and properties in variational analysis. In Section 3, we describe the generalized positive definite regularized Newton method based on the generalized Hessian matrix and establish its feasibility and convergence conclusions. In Section 4, for a class of general prox-regular functions, we present the application of our algorithm in nonsmooth optimization using the Moreau envelope. In Section 5, we demonstrate the application of our algorithm in a class of convex composite optimization via the forward-backward envelope. In Section 6, we provide the application of the proposed algorithm in the LASSO problem and numerical experimental results. In Section 7, we summarize the main achievements of this paper and discuss future research plans.

2 Notations and Preliminary Results

Throughout this paper, we work within finite-dimensional Euclidean spaces and adopt the standard notation and terminology from variational analysis and generalized differentiation; see, e.g., [22, 23, 31] for most of the results presented in this section. Recall that

$\mathbb{B}_r(x)$ denotes the closed ball centered at $x \in \mathbb{R}^n$ with radius $r > 0$, and $\mathbb{N} := \{1, 2, \ldots\}$. For a symmetric matrix $Q \in \mathbb{R}^{n\times n}$, let $\lambda_{\min(Q)}$ and $\lambda_{\max(Q)}$ denote the smallest and largest eigenvalues of Q, respectively. It is straightforward that for any $x \in \mathbb{R}^n$,

$$\lambda_{\min(Q)}\|x\|^2 \leq \langle Qx, x\rangle \leq \lambda_{\max(Q)}\|x\|^2 \quad \text{and} \quad \lambda_{\min(Q)}\|x\| \leq \|Qx\| \leq \lambda_{\max(Q)}\|x\|.$$

For a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, its graph is defined as $\operatorname{gph} F := \{(v, w) \in \mathbb{R}^n \times \mathbb{R}^m \mid w \in F(v)\}$. The (Painlevé–Kuratowski) outer limit of F at $x \in \mathbb{R}^n$ is given by:

$$\operatorname*{Lim\,sup}_{u\to x} F(u) := \{y \in \mathbb{R}^m \mid \exists\, u_k \to x,\, y_k \to y \text{ with } y_k \in F(u_k) \text{ for all } k \in \mathbb{N}\}.$$

For a nonempty set $S \subseteq \mathbb{R}^n$, the (Fréchet) regular normal cone at $x \in S$ is

$$\widehat{N}(S; x) := \left\{ v \in \mathbb{R}^n \mid \limsup_{u \xrightarrow{S} x} \left\langle v, \frac{u - x}{\|u - x\|} \right\rangle \leq 0 \right\},$$

where $u \xrightarrow{S} x$ indicates that $u \to x$ with $u \in S$. The (Mordukhovich) basic/limiting normal cone at $x \in S$ is

$$N(S; x) := \limsup_{u \xrightarrow{S} x} \widehat{N}(S; u).$$

By convention, $\widehat{N}(S; x) = N(S; x) := \emptyset$ if $x \notin S$. The indicator function of S, denoted $\delta_S(x)$, equals 0 for $x \in S$ and ∞ otherwise.

For a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, the regular and basic coderivatives at $(x, y) \in \operatorname{gph} F$ are defined for all $d \in \mathbb{R}^m$ using the corresponding normal cones to $\operatorname{gph} F$

$$\widehat{D}^* F(x, y)(d) := \left\{ w \in \mathbb{R}^n \mid (w, -d) \in \widehat{N}\big(\operatorname{gph} F; (x, y)\big) \right\}$$

and

$$D^* F(x, y)(d) := \left\{ w \in \mathbb{R}^n \mid (w, -d) \in N\big(\operatorname{gph} F; (x, y)\big) \right\}.$$

If F is single-valued at x, the argument y is omitted.

We now proceed to recall the definition of Hölder metric subregularity, a notion that assumes significant importance in the convergence analysis of numerical algorithms.

Definition 2.1. *Let $p \in (0, \infty)$. Consider a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a pair $(\bar{x}, \bar{y}) \in \operatorname{gph} F$. We say that F is* Hölder metrically subregular *at $(\bar{x}, \bar{y})$ of order p if there exist constants $\kappa, \delta \in (0, \infty)$ such that the inequality*

$$d(x, F^{-1}(\bar{y})) \leq \kappa\, d(\bar{y}, F(x))^p$$

holds for all $x \in \mathbb{B}_\delta(\bar{x})$, where $F^{-1}(\bar{y}) = \{x \in \mathbb{R}^n \mid \bar{y} \in F(x)\}$.

Moreover, if the set $F^{-1}(\bar{y}) \cap \mathbb{B}_\delta(\bar{x})$ contains only the point $\bar{x}$ (i.e., $F^{-1}(\bar{y}) \cap \mathbb{B}_\delta(\bar{x}) = \{\bar{x}\}$), then F is said to be Hölder strongly metrically subregular *at $(\bar{x}, \bar{y})$ of order p.*

When $p = 1$, the aforementioned notion reduces to the well-known concept of (strong) metric subregularity. The order of Hölder metric subregularity can be any positive number. For instance, consider the mapping $F(x) := |x|^{\frac{1}{p}}$ for all $x \in \mathbb{R}$: this mapping is Hölder strongly metrically subregular at $(0, 0)$ of order p, where $p \in (0, \infty)$. A characterization of the strong metric subregularity property is known as the Levy–Rockafellar criterion; readers can be referred to the monograph by Dontchev and Rockafellar [9, Theorem 4E.1] for more details. For properties of the Hölder metric subregularity case and its applications, see [7, 28, 36, 37].

Next, we present an extension of the positive-definiteness notion for multifunctions, in which the associated constant need not be positive (see [1]).

Definition 2.2. *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $\xi \in \mathbb{R}$. We say that F is ξ-lower-definite if*

$$\langle y, x\rangle \geq \xi\|x\|^2 \quad \text{for all } (x, y) \in \text{gph}\, F.$$

Remark 2.3. *The following observations can be readily verified:*

(i) For any symmetric matrix Q, the linear function $f(x) = Qx$ is $\lambda_{\min(Q)}$-lower-definite.

(ii) If a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex, then by [6, Theorem 3.2], $\partial^2 f(x, v)$ is 0-lower-definite for all $(x, v) \in \text{gph}\,\partial f$.

(iii) If $F_1, F_2 : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are ξ_1- and ξ_2-lower-definite respectively, their sum $F_1 + F_2$ is $(\xi_1 + \xi_2)$-lower-definite.

For a lower semicontinuous (l.s.c.) function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, its domain and epigraph are defined respectively as

$$\text{dom}\, f := \{x \in \mathbb{R}^n \mid f(x) < \infty\} \quad \text{and} \quad \text{epi}\, f := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\}.$$

The function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be proper if $\text{dom}\, f \neq \varnothing$.

The regular and basic subdifferentials of f at $x \in \text{dom}\, f$ are defined via the corresponding normal cones to its epigraph

$$\hat{\partial} f(x) := \Big\{v \in \mathbb{R}^n \mid (v, -1) \in \hat{N}\big(\text{epi}\, f; (x, f(x))\big)\Big\}$$

and

$$\partial f(x) := \big\{v \in \mathbb{R}^n \mid (v, -1) \in N\big(\text{epi}\, f; (x, f(x))\big)\big\}.$$

For a single-valued mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ which is locally Lipschitzian around x, the basic coderivative admits the following representation via the basic subdifferential of its scalarization

$$D^*F(x)(d) = \partial\langle d, F\rangle(x), \quad \text{where } \langle d, F\rangle(x) := \langle d, F(x)\rangle. \tag{5}$$

A lower semicontinuous (l.s.c.) function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *prox-regular at a point* $\bar{x} \in \text{dom}\, f$ for a *subgradient* $\bar{v} \in \partial f(\bar{x})$ with *modulus* $r > 0$ if there exists $\varepsilon > 0$ such that, for all $x, u \in \mathbb{B}_\varepsilon(\bar{x})$ satisfying $|f(u) - f(\bar{x})| < \varepsilon$, the inequality

$$f(x) \geq f(u) + \langle v, x - u\rangle - \frac{r}{2}\|x - u\|^2 \tag{6}$$

holds whenever $v \in \partial f(u) \cap \mathbb{B}_\varepsilon(\bar{v})$. If this condition is satisfied for all $\bar{v} \in \partial f(\bar{x})$, then f is said to be *prox-regular* at $\bar{x} \in \text{dom}\, f$.

A function f is said to be *subdifferentially continuous* at $\bar{x} \in \text{dom}\, f$ *for* $\bar{v} \in \partial f(\bar{x})$ if, for any sequence $(x_k, v_k) \to (\bar{x}, \bar{v})$ with $v_k \in \partial f(x_k)$, we have $f(x_k) \to f(\bar{x})$ as $k \to \infty$. If this holds for all $\bar{v} \in \partial f(\bar{x})$, then f is said to be *subdifferentially continuous* at $\bar{x}$.

In what follows, we say that a function f is *continuously prox-regular at $\bar{x}$ for $\bar{v}$* (and simply *at $\bar{x}$*) if it is both prox-regular and subdifferentially continuous at that point. Given a set $\Omega \subseteq \mathbb{R}^n$, f is said to be *continuously prox-regular on* Ω *with modulus* $r > 0$ if it is continuously prox-regular at every point of Ω with the same modulus r.

Note that f is trivially subdifferentially continuous at any point $\bar{x} \in \operatorname{dom} f$ where f is continuous relative to its domain. As discussed by Rockafellar and Wets [31], the class of continuously prox-regular functions is quite broad, encompassing not only C^2-smooth functions but also $C^{1,1}$-class functions, convex l.s.c. functions, lower C^2 functions, strongly amenable functions, etc..

In addition to the (first-order) basic subdifferential, we consider the second-order subdifferential (or generalized Hessian) of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ at $x \in \operatorname{dom} f$ relative to $v \in \partial f(x)$, defined by

$$\partial^2 f(x, v)(d) = (D^*\partial f)(x, v)(d) \quad \text{for all } d \in \mathbb{R}^n.$$

When $\partial f(x)$ is a singleton, this is simplified to $\partial^2 f(x)(d)$. For a function f that is twice continuously differentiable (C^2-smooth) in a neighborhood of x, we have $\partial^2 f(x)(d) = \{\nabla^2 f(x)d\}$.

Next, we recall, for completeness, the notions of convergence rates used in our analysis of the algorithms. Let $\{x^k\} \subset \mathbb{R}^n$ be a sequence of vectors satisfying $x^k \to \bar{x}$ as $k \to \infty$, with $\bar{x} \neq x^k$ for all $k \in \mathbb{N}$ (i.e., no term in the sequence is equal to the limit $\bar{x}$ prior to convergence).

The sequence $\{x^k\}$ is said to converge R-linearly to $\bar{x}$ if

$$0 < \limsup_{k\to\infty} \left(\|x^k - \bar{x}\|\right)^{1/k} < 1,$$

which is equivalent to the existence of constants $\mu \in (0, 1)$, $c > 0$, and an integer $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$,

$$\|x^k - \bar{x}\| \leq c\,\mu^k.$$

The sequence $\{x^k\}$ is said to converge Q-linearly to $\bar{x}$ if

$$\limsup_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} < 1,$$

which is equivalent to the existence of constants $\mu \in (0, 1)$ and an integer $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$,

$$\|x^{k+1} - \bar{x}\| \leq \mu\,\|x^k - \bar{x}\|.$$

To analyze the convergence rate of the algorithm, we require the following convergence rate results for monotone sequences.

Lemma 2.4. *Let $q, \eta \in (0, \infty)$, and consider a nonnegative sequence $\{\alpha_k\}$ that satisfies the recurrence relation*

$$\alpha_{k+1} \leq \alpha_k - \eta\alpha_k^q \quad \textit{for all integers } k \geq 0. \tag{7}$$

The subsequent convergence properties hold:

(i) When $q \in (0, 1)$, there exists an integer $k_0 \in \mathbb{N}$ such that $\alpha_k = 0$ for every $k \geq k_0$. In other words, the sequence terminates at zero after a finite number of iterations.

(ii) When $q = 1$, the sequence $\{\alpha_k\}$ converges to zero with Q-linear rate.

(iii) When $q \in (1, \infty)$, the sequence satisfies the asymptotic bound

$$\alpha_k \leq O\left(k^{\frac{1}{1-q}}\right) \quad \text{for sufficiently large } k \geq 0. \tag{8}$$

Here, the big-O notation denotes that there exists a constant $M > 0$ such that $\alpha_k \leq M \cdot k^{\frac{1}{1-q}}$ for sufficiently large k.

Proof. (i) Let $q \in (0, 1)$. Since $\eta > 0$ and $\alpha_k \geq 0$ for all $k \in \mathbb{N}$, the recurrence relation (7) directly implies that $\alpha_{k+1} \leq \alpha_k$ for all $k \in \mathbb{N}$. By the Monotone Convergence Theorem, this nonnegative, monotone-decreasing sequence $\{\alpha_k\}$ converges to some limit $\alpha \in [0, \infty)$. Taking the limit on both sides of (7) and using $\eta > 0$, we obtain

$$\alpha \leq \alpha - \eta\alpha^q$$

Subtracting α from both sides forces that $0 \leq -\eta\alpha^q$. Since $\eta > 0$ and $\alpha^q \geq 0$, this inequality holds if and only if $\alpha = 0$.

If $\alpha_{k_0} = 0$ for some $k_0 \in \mathbb{N}$, then for all $k \geq k_0$, $\alpha_k \leq \alpha_{k_0} = 0$ (as the sequence is decreasing). Thus $\alpha_k = 0$ for $k \geq k_0$. Otherwise, suppose that $\alpha_k > 0$ for all $k \in \mathbb{N}$. Dividing (7) by α_k^q (valid since $\alpha_k > 0$) gives

$$0 \leq \frac{\alpha_{k+1}}{\alpha_k^q} \leq \alpha_k^{1-q} - \eta$$

Since $q \in (0, 1)$, $1 - q > 0$, so $\alpha_k^{1-q} \to 0$ as $k \to \infty$ (because $\alpha_k \downarrow 0$). Taking the limit as $k \to \infty$, the right-hand side tends to $-\eta$, yielding $0 \leq -\eta$, which is a contradiction to $\eta > 0$. Hence, such a sequence cannot stay positive forever, so $\alpha_{k_0} = 0$ must hold for some k_0.

(ii) Let $q = 1$. From (7), we directly derive

$$\alpha_{k+1} \leq (1 - \eta)\alpha_k \quad \forall\, k \in \mathbb{N},$$

which implies that the sequence $\{\alpha_k\}$ converges Q-linearly to zero.

(iii) Let $q \in (1, \infty)$. Applying [3, Lemma 4.1] with $\delta_k \equiv \eta$, we derive the bound

$$\alpha_k \leq \left(\alpha_0^{1-q} + (q-1)\sum_{i=0}^{k-1} \eta\right)^{\frac{1}{1-q}}$$

The summation $\sum_{i=0}^{k-1} \eta = \eta k$, so by substitution and simplification, we have $\alpha_k \leq O\left(k^{\frac{1}{1-q}}\right)$, which completes the proof. ■

3 Coderivative-based positive definite regularized Newton algorithm for $\mathcal{C}^{1,1}$ functions

To find stationary points of the $C^{1,1}$ problem (1), we propose a coderivative-based positive definite regularized Newton method (PDRNM), labeled as Algorithm 1 below. The goal of this section is to justify the well-posedness and convergence of the novel algorithm PDRNM under appropriate and fairly general assumptions.

Algorithm 1: Coderivative-based positive definite regularized Newton algorithm for $\mathcal{C}^{1,1}$ functions

```
Require: x^0 ∈ R^n and σ, β ∈ (0,1)
1: for k = 0, 1, ... do
2:   if ∇f(x^k) = 0 then
3:     Stop and return x^k
4:   end if
5:   Choose B_k ≻ 0, ρ_k ≥ 0 and d^k ∈ R^n \ {0} such that
     −∇f(x^k) ∈ ∂²f(x^k)(d^k) + ρ_k B_k d^k
6:   Set τ_k = 1
7:   while f(x^k + τ_k d^k) > f(x^k) + στ_k⟨∇f(x^k), d^k⟩ do
8:     Set τ_k := βτ_k
9:   end while
10:  Set x^{k+1} := x^k + τ_k d^k
11: end for
```

Remark 3.1. *(a) The proposed Algorithm 1 generalizes the regularized Newton method (RNM) [17] to second-order nonsmooth setting. Algorithm 1 also extends GRNM in [14] and GRNM-PD in [33]. In particular, when $B_k = I$ and $\rho_k = c\|\nabla\varphi(x^k)\|$ for some constant $c > 0$, Algorithm 1 reduces to the GRNM in [14]. When $\rho_k = c\|\nabla\varphi(x^k)\|^\theta$ for some constant $c > 0$ and $\theta \in [0,2]$, Algorithm 1 reduces to the GRNM-PD in [33].*

(b) Similarly to the GRNM-PD in [33], Algorithm 1 uses the regularization matrix B_k to replace the identity matrix for regularizing the Newton method, and numerical experiments indicate that choosing B_k appropriately can significantly accelerate the algorithm.

The next lemma shows that Algorithm 1 is well-defined by proving the existence of a direction d^k in Step 5 for sufficiently large regularization parameters ρ_k.

Lemma 3.2. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ of class $C^{1,1}$ on $\mathbb{R}^n$, let $\bar{x} \in \mathbb{R}^n$ be such that $\nabla f(\bar{x}) \neq 0$, and the second-order subdifferential $\partial^2 f(\bar{x})$ be ξ-lower-definite for some $\xi \in \mathbb{R}$. And let $B \in \mathbb{R}^{n\times n}$ be a positive definite symmetric matrix.*

Then for any $\zeta > 0$ and any $\rho \geq \frac{\zeta-\xi}{\lambda_{\min(B)}}$, there exists a nonzero direction $d \in \mathbb{R}^n$ such that

$$-\nabla f(\bar{x}) \in \partial^2 f(\bar{x})(d) + \rho B d. \tag{9}$$

Consequently, for each $\sigma \in (0,1)$ and $d \in \mathbb{R}^n \setminus \{0\}$ satisfying (9), we have

$$\langle \nabla f(\bar{x}), d\rangle \leq -\zeta\|d\|^2,$$

and there exists $\delta > 0$ such that

$$f(\bar{x} + \tau d) \leq f(\bar{x}) + \sigma\tau\langle \nabla f(\bar{x}), d\rangle \text{ whenever } \tau \in (0,\delta). \tag{10}$$

Proof. For any $\zeta > 0$ and any $\rho \geq \frac{\zeta-\xi}{\lambda_{\min(B)}}$, consider the function

$$\psi(x) := f(x) + \frac{\rho}{2}\langle B(x-\bar{x}), x-\bar{x}\rangle \quad \forall x \in \mathbb{R}^n.$$

By the second-order subdifferential sum rule in [22, Proposition 1.121], we have

$$\partial^2\psi(\bar{x})(v) = \partial^2 f(\bar{x})(v) + \rho B v \quad \text{for all } v \in \mathbb{R}^n. \tag{11}$$

By the positive definiteness of B, we have $\lambda_{\min(B)} > 0$, and thus $\xi + \rho\lambda_{\min(B)} \geq \zeta$.

Since $\partial^2 f(\bar{x})$ is ξ-lower-definite and ρB is $\rho\lambda_{\min(B)}$-lower-definite, Remark 2.3 (iii) implies that $\partial^2\psi(\bar{x})$ is $(\xi + \rho\lambda_{\min(B)})$-lower-definite, hence is also ζ-lower-definite. Note that $\nabla\psi(\bar{x}) = \nabla f(\bar{x}) \neq 0$ and $\zeta > 0$. By [13, Proposition 1] and (11), there exists a nonzero direction d such that

$$-\nabla f(\bar{x}) = -\nabla\psi(\bar{x}) \in \partial^2\psi(\bar{x})(d) = \partial^2 f(\bar{x})(d) + \rho B d, \tag{12}$$

which verifies (9). Due to the ξ-lower-definiteness of $\partial^2 f(\bar{x})$, we obtain

$$\langle \nabla f(\bar{x}), d\rangle \leq -\xi\|d\|^2 - \rho\langle Bd, d\rangle \leq -(\xi + \rho\lambda_{\min(B)})\|d\|^2 \leq -\zeta\|d\|^2.$$

Since $\zeta > 0$, we also have $\langle \nabla f(\bar{x}), d\rangle < 0$. By [12, Lemmas 2.18 and 2.19], there exists $\delta > 0$ such that (10) holds, completing the proof. ■

Regarding the sequence $\{\tau_k\}$ in Algorithm 1, we have the following estimate.

Lemma 3.3. *For a function $f : \mathbb{R}^n \to \mathbb{R}$ of class $C^{1,1}$ on $\mathbb{R}^n$, let $l > 0$ be the Lipschitz constant of ∇f. If Algorithm 1 generates an infinite sequence $\{x^k\}$, then for each $k \in \mathbb{N}$, the sequence $\{\tau_k\}$ in Algorithm 1 satisfies either $\tau_k = 1$ or $1 > \tau_k \geq 2(1-\sigma)\zeta\beta/l$.*

Proof. For each $k \in \mathbb{N}$, if $\tau_k \neq 1$, then Step 7 of Algorithm 1 ensures the inequality

$$f\left(x^k + \beta^{-1}\tau_k d^k\right) > f(x^k) + \sigma\beta^{-1}\tau_k \left\langle \nabla f(x^k), d^k \right\rangle. \tag{13}$$

Note that ∇f is Lipschitz continuous. By [12, Lemma A.11], we have

$$f\left(x^k + \beta^{-1}\tau_k d^k\right) \leq f(x^k) + \beta^{-1}\tau_k \left\langle \nabla f(x^k), d^k \right\rangle + \frac{l}{2}\beta^{-2}\tau_k^2 \left\|d^k\right\|^2. \tag{14}$$

By Lemma 3.2, $\left\langle \nabla f(x^k), d^k \right\rangle \leq -\zeta \left\|d^k\right\|^2 < 0$. Combining this with (13) and (14), we obtain

$$\begin{aligned} \sigma\beta^{-1}\tau_k \left\langle \nabla f(x^k), d^k \right\rangle &< f\left(x^k + \beta^{-1}\tau_k d^k\right) - f(x^k) \\ &\leq \beta^{-1}\tau_k \left\langle \nabla f(x^k), d^k \right\rangle + \frac{l}{2}\beta^{-2}\tau_k^2 \left\|d^k\right\|^2 \\ &\leq \beta^{-1}\tau_k \left(1 - \frac{l\tau_k}{2\zeta\beta}\right) \left\langle \nabla f(x^k), d^k \right\rangle. \end{aligned} \tag{15}$$

Since $\left\langle \nabla f(x^k), d^k \right\rangle < 0$, it follows from (15) that $\tau_k \geq 2(1-\sigma)\zeta\beta/l$. ■

Now we are ready to establish the aforementioned theorem regarding the performance of Algorithm 1.

Theorem 3.4. *Let $\zeta > 0$, $\xi \in \mathbb{R}$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a $C^{1,1}$-class objective function on $\mathbb{R}^n$ for problem (1) with $\inf f > -\infty$. Choose an initial point $x_0 \in \mathbb{R}^n$ and define the sublevel set*

$$\Omega := \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}.$$

Assume that the eigenvalues of the regularization matrices B_k in Algorithm 1 lie within $[\lambda_{\min}, \lambda_{\max}]$ where $\lambda_{\max} > \lambda_{\min} > 0$, and that the second-order subdifferential $\partial^2 f(x)$ is ξ-lower-definite for all $x \in \Omega$.

Then Algorithm 1 either terminates at a stationary point, or if $\sup_{k\in\mathbb{N}} \rho_k < \infty$ and $\inf_{k\in\mathbb{N}} \rho_k \geq \frac{\zeta-\xi}{\lambda_{\min}}$, it generates sequences $\{x^k\} \subseteq \Omega$, $\{f(x^k)\}$, $\{\nabla f(x^k)\}$, $\{d^k\}$, and $\{\tau_k\}$ satisfying:

(i) The sequence $\{f(x^k)\}$ is monotonically decreasing and convergent.

(ii) $\sum_{k=0}^{\infty} \|\nabla f(x^k)\|^2 < \infty$,

$$\inf_{k\in\mathbb{N}} \tau_k > 0, \quad \sum_{k=0}^{\infty} \|d^k\|^2 < \infty, \quad \textit{and} \quad \sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < \infty. \tag{16}$$

In particular, if $\bar{x}$ is an accumulation point of $\{x^k\}$, then $\bar{x}$ is a stationary point of problem (1) with $f(\bar{x}) = \inf_{k\in\mathbb{N}} f(x^k)$.

(iii) If $\{x^k\}$ has an isolated accumulation point $\bar{x}$, then the entire sequence $\{x^k\}$ converges to $\bar{x}$ as $k \to \infty$, where $\bar{x}$ is a stationary point of (1).

Proof. If Algorithm 1 terminates after a finite number of iterations, it clearly returns a stationary point. Otherwise, it generates an infinite sequence $\{x^k\}$. By Step 7 of Algorithm 1 and Lemma 3.2, we have $\inf f \le f(x^{k+1}) < f(x^k)$ for all $k \in \mathbb{N}$, which proves assertion (i) and implies $\{x^k\} \subseteq \Omega$.

For (ii), we first show that $\inf_{k\in\mathbb{N}} \tau_k > 0$. Assume to the contrary that $\inf_{k\in\mathbb{N}} \tau_k = 0$. Then there exists a subsequence $\{\tau_{k_j}\}$ of $\{\tau_k\}$ such that $\tau_{k_j} \to 0$ as $j \to \infty$. Let $l > 0$ denote the Lipschitz constant of ∇f. Since $\tau_{k_j} \to 0^+$, we may assume that $\tau_{k_j} < 1$ for all $j \in \mathbb{N}$. By Lemma 3.3, this implies that $\tau_{k_j} \ge 2(1-\sigma)\zeta\beta/l > 0$, contradicting $\tau_{k_j} \to 0$. Thus, $\inf_{k\in\mathbb{N}} \tau_k > 0$.

From Step 5 of Algorithm 1 and Lemma 3.2, we have $-\nabla f(x^k) \in \partial^2 f(x^k)(d^k) + \rho_k B_k d^k$, which implies that

$$\langle \nabla f(x^k), d^k\rangle \le -\zeta\|d^k\|^2 \quad \text{for all } k \in \mathbb{N}. \tag{17}$$

Applying the Cauchy-Schwarz inequality to (17) yields

$$\zeta\|d^k\| \le \|\nabla f(x^k)\| \quad \text{for all } k \in \mathbb{N}. \tag{18}$$

From Steps 7–10 of Algorithm 1, we have $-\sigma\tau_k\langle\nabla f(x^k), d^k\rangle \le f(x^k) - f(x^{k+1})$. Combining this with (17) gives

$$\sum_{k=0}^{\infty} \zeta\tau_k\|d^k\|^2 \le -\sum_{k=0}^{\infty} \tau_k\langle\nabla f(x^k), d^k\rangle \le \frac{1}{\sigma}\left(f(x^0) - \inf_{k\in\mathbb{N}} f(x^k)\right) < \infty.$$

Since $\zeta > 0$ and $\inf_{k\in\mathbb{N}} \tau_k > 0$, it follows that $\sum_{k=0}^{\infty} \|d^k\|^2 < \infty$. Moreover, $\|x^{k+1} - x^k\| = \tau_k\|d^k\|$ implies $\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < \infty$.

By the inclusion $-\nabla f(x^k) - \rho_k B_k d^k \in \partial^2 f(x^k)(d^k) := D^*(\nabla f)(x^k)(d^k)$ and the Lipschitz continuity of ∇f, [22, Theorem 1.44] yields

$$\| -\nabla f(x^k) - \rho_k B_k d^k\| \le l\|d^k\|.$$

Using the triangle inequality,

$$\|\nabla f(x^k)\| \le \|\nabla f(x^k) + \rho_k B_k d^k\| + \|\rho_k B_k d^k\| \le (l + \rho_k\lambda_{\max})\|d^k\|. \tag{19}$$

Since $\sup_{k\in\mathbb{N}} \rho_k < \infty$, $\sum_{k=0}^{\infty} \|d^k\|^2 < \infty$ implies $\sum_{k=0}^{\infty} \|\nabla f(x^k)\|^2 < \infty$. Consequently, $\nabla f(x^k) \to 0$ as $k \to \infty$.

Let $\bar{x}$ be an accumulation point of $\{x^k\}$. Then there exists a subsequence $x^{k_j} \to \bar{x}$ as $j \to \infty$. The continuity of ∇f implies $\nabla f(x^{k_j}) \to \nabla f(\bar{x})$, so $\nabla f(\bar{x}) = 0$, making $\bar{x}$ a stationary point of (1). Due to the monotonicity of $\{f(x^k)\}$, we have $f(\bar{x}) = \inf_{k\in\mathbb{N}} f(x^k)$.

For (iii), assume $\{x^k\}$ has an isolated accumulation point $\bar{x}$. For any subsequence $x^{k_j} \to \bar{x}$, the preceding arguments show $\bar{x}$ is a stationary point and $\lim_{j\to\infty} \|x^{k_j+1} - x^{k_j}\| = 0$. By [10, Proposition 8.3.10], the entire sequence $\{x^k\}$ converges to $\bar{x}$, completing the proof. ■

Remark 3.5. *If the entire sequence $\{x^k\}$ generated by Algorithm 1 is bounded, then the set of its accumulation points is nonempty, closed, and connected. Indeed, by part (ii) of Theorem 3.4, we have the Ostrowski condition $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$. The conclusion then follows from [27, Theorem 28.1].*

Under some additional assumptions, the theorem below establishes the convergence rates for the iterative sequences produced by Algorithm 1.

Theorem 3.6. *Suppose that, in addition to the assumptions of Theorem 3.4, the sequence $\{x^k\}$ has an accumulation point $\bar{x}$ such that the gradient mapping ∇f is Hölder strongly metrically subregular at $(\bar{x}, 0)$ of order $p \in (0, \infty)$. The following assertions hold:*

(i) If $p \in (1, \infty)$, then the sequences $\{x^k\}$, $\{\nabla f(x^k)\}$, and $\{f(x^k)\}$ converge in a finite number of steps.

(ii) If $p = 1$, then the sequence $\{f(x^k)\}$ converges at least Q-linearly, and both $\{x^k\}$ and $\{\nabla f(x^k)\}$ converge at least R-linearly.

(iii) If $p \in (0, 1)$, then for sufficiently large $k \in \mathbb{N}$,

$$\|x^k - \bar{x}\| \le O\left(k^{\frac{p^2}{2(p-1)}}\right), \quad \|\nabla f(x^k)\| \le O\left(k^{\frac{p}{2(p-1)}}\right), \text{ and } |f(x^k) - f(\bar{x})| \le O\left(k^{\frac{p}{p-1}}\right).$$

Proof. Since $\bar{x}$ is an accumulation point of $\{x^k\}$, it follows from part (ii) of Theorem 3.4 that $\nabla f(\bar{x}) = 0$ and $f(x^k) \ge f(\bar{x})$ for all $k \in \mathbb{N}$. Note that the imposed Hölder strong metric subregularity of ∇f implies that $\bar{x}$ is an isolated accumulation point. By part (iii) of Theorem 3.4, this ensures $x^k \to \bar{x}$ as $k \to \infty$. Further, there exists $\kappa > 0$ such that

$$\|x^k - \bar{x}\| \le \kappa \|\nabla f(x^k)\|^p \quad \text{for all sufficiently large } k \in \mathbb{N}. \tag{20}$$

Since ∇f is Lipschitz continuous and $\nabla f(\bar{x}) = 0$, [12, Lemma A.11] guarantees the existence of $l > 0$ satisfying

$$0 \le f(x^k) - f(\bar{x}) = \left| f(x^k) - f(\bar{x}) - \langle \nabla f(\bar{x}), x^k - \bar{x} \rangle \right| \le \frac{l}{2} \|x^k - \bar{x}\|^2. \tag{21}$$

By Step 7 of Algorithm 1 and Lemma 3.2, we have $f(x^k) - f(x^{k+1}) \ge \sigma\zeta\tau_k \|d^k\|^2$ holds for all sufficiently large $k \in \mathbb{N}$. Combining this with (19) yields

$$\Big(f(x^k) - f(\bar{x})\Big) - \Big(f(x^{k+1}) - f(\bar{x})\Big) \ge \frac{\sigma\zeta\tau_k}{(l + \rho_k \lambda_{\max})^2} \|\nabla f(x^k)\|^2 \tag{22}$$

for k large enough. Given $\sup_{k\in\mathbb{N}} \rho_k < \infty$ and $f(x^k) - f(\bar{x}) \ge 0$, there exists $M > 0$ such that

$$\|\nabla f(x^k)\|^2 \le M \Big(f(x^k) - f(\bar{x})\Big). \tag{23}$$

Let $\alpha_k := f(x^k) - f(\bar{x})$ and define

$$\eta := \inf_{k\in\mathbb{N}} \left(\frac{2}{l\kappa^2}\right)^{\frac{1}{p}} \frac{\sigma\zeta\tau_k}{(l+\rho_k\lambda_{\max})^2}.$$

Clearly, $\alpha_k \geq 0$. By part (ii) of Theorem 3.4, we have $\inf_{k\in\mathbb{N}} \tau_k > 0$, which implies that $\eta > 0$. From (20), (21), and (22), we derive

$$\alpha_{k+1} \leq \alpha_k - \eta\alpha_k^{\frac{1}{p}}. \tag{24}$$

Applying Lemma 2.4 with $q = \frac{1}{p}$, we obtain:

(i) If $p \in (1,\infty)$, then the sequence $\{\alpha_k\}$ converges to zero in finitely many steps. By (20) and (23), $\{x^k\}, \{\nabla f(x^k)\}$ and $\{f(x^k)\}$ also converge in finitely many steps.

(ii) If $p = 1$, then $\{\alpha_k\}$ converges Q-linearly to zero. By (20) and (23), both $\{x^k\}$ and $\{\nabla f(x^k)\}$ converge at least R-linearly.

(iii) If $p \in (0,1)$, then $\{\alpha_k\}$ satisfies $\alpha_k \leq O\left(k^{\frac{p}{p-1}}\right)$ for sufficiently large $k \in \mathbb{N}$. By (20) and (23), this implies $\|x^k - \bar{x}\| \leq O\left(k^{\frac{p^2}{2(p-1)}}\right)$ and $\|\nabla f(x^k)\| \leq O\left(k^{\frac{p}{2(p-1)}}\right)$.

This completes the proof of the theorem. ■

Remark 3.7. *(a) We note that our convergence analysis differs from that in [14]. Specifically, the convergence rate analysis in [14] requires metric regularity of ∇f at a solution (which implies local strong convexity of f), whereas we impose the weaker assumption of Hölder strong metric subregularity. Our work draws inspiration from [1, 21, 33], where proximal/generalized Newton-type methods for (convex composite) minimization problems are proposed and analyzed under (Hölder) metric subregularity.*

(b) It is worth noting that in Theorem 3.4, ξ can be any real number—even a negative one—provided that $\rho_k \geq \frac{\zeta-\xi}{\lambda_{\min}}$. However, when $\xi = 0$ and $\rho_k = c\|\nabla f(x^k)\| \downarrow 0$ for some $c > 0$, there exists no positive ζ satisfying $\rho_k \geq \frac{\zeta-\xi}{\lambda_{\min}}$. In this case, we have the following result.

Theorem 3.8. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $C^{1,1}$-class objective function for problem (1) on $\mathbb{R}^n$, with $\inf f > -\infty$. Choose an initial point $x_0 \in \mathbb{R}^n$ and define the sublevel set*

$$\Omega := \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}.$$

Assume further that $\rho_k = c\|\nabla f(x^k)\|$ for some $c > 0$, and that the eigenvalues of the regularization matrices B_k in Algorithm 1 lie within $[\lambda_{\min}, \lambda_{\max}]$ where $\lambda_{\max} > \lambda_{\min} > 0$. Additionally, assume that the second-order subdifferential $\partial^2 f(x)$ is 0-lower-definite (positive semidefinite) for all $x \in \Omega$.

Then Algorithm 1 is well-defined, the sequence $\{f(x^k)\}$ is monotonically decreasing and convergent, $\|\nabla f(x^k)\| \to 0$ as $k \to \infty$, and any accumulation point of $\{x^k\}$ is a stationary point of (1).

In addition, suppose $\{x^k\}$ has an accumulation point $\bar{x}$ such that the gradient mapping ∇f is Hölder strongly metrically subregular at $(\bar{x}, 0)$ of order $p \in (0,\infty)$. Then the entire sequence $\{x^k\}$ converges to $\bar{x}$, and $\bar{x}$ is a $\frac{1+p}{p}$-order sharp minimizer of (1); that is, there exist $\epsilon, \delta \in (0,\infty)$ such that

$$\epsilon\|x - \bar{x}\|^{\frac{1+p}{p}} \leq f(x) - f(\bar{x}) \quad \text{for all } x \in \mathbb{B}_\delta(\bar{x}).$$

We also have the following assertions:

(i) If $p \in (2,\infty)$, then the sequences $\{x^k\}$, $\{\nabla f(x^k)\}$, and $\{f(x^k)\}$ converge in a finite number of steps.

(ii) If $p = 2$, then $\{f(x^k)\}$ converges at least Q-linearly, and both $\{x^k\}$ and $\{\nabla f(x^k)\}$ converge at least R-linearly.

(iii) If $p \in (0,2)$, then for all sufficiently large $k \in \mathbb{N}$,

$$\|x^k - \bar{x}\| \le O\left(k^{\frac{p^2}{2(p-2)}}\right), \quad \|\nabla f(x^k)\| \le O\left(k^{\frac{p}{2(p-2)}}\right), \text{ and } |f(x^k) - f(\bar{x})| \le O\left(k^{\frac{p}{p-2}}\right).$$

Proof. By applying Lemma 3.2 with $\xi = 0$ and $\zeta = \rho_k \lambda_{\min}$, for each $k \in \mathbb{N}$, there exists a nonzero direction $d^k \in \mathbb{R}^n$ satisfying

$$-\nabla f(x^k) \in \partial^2 f(x^k)(d^k) + \rho_k B_k d^k,$$

which shows that Algorithm 1 is well-defined. Meanwhile, we also obtain

$$\langle \nabla f(x^k), d^k \rangle \le -\rho_k \lambda_{\min} \|d^k\|^2. \tag{25}$$

Let $l > 0$ denote the Lipschitz constant of ∇f; both (19) and (21) hold. From (25) and Steps 7–10 of Algorithm 1, we derive

$$0 \le \rho_k \lambda_{\min} \sigma \tau_k \|d^k\|^2 \le -\sigma \tau_k \langle \nabla f(x^k), d^k \rangle \le f(x^k) - f(x^{k+1}). \tag{26}$$

This implies the sequence $\{f(x^k)\}$ is monotonically decreasing and convergent.

If Algorithm 1 terminates after a finite number of iterations, there is nothing to prove. Next, we assume it generates an infinite sequence $\{x^k\}$. The remainder of the proof is divided into three claims.

Claim 1: $\|\nabla f(x^k)\| \to 0$ as $k \to \infty$, and any accumulation point of $\{x^k\}$ is a stationary point of (1).

By (19) and (26), we have

$$\sum_{k=0}^{\infty} \frac{c\lambda_{\min}\tau_k}{(l + \rho_k \lambda_{\max})^2} \|\nabla f(x^k)\|^3 \le \sum_{k=0}^{\infty} \rho_k \lambda_{\min} \tau_k \|d^k\|^2 \le \frac{1}{\sigma}\left(f(x_0) - \inf_{k\in\mathbb{N}} f(x^k)\right). \tag{27}$$

Note that $\zeta = \rho_k \lambda_{\min} = c\lambda_{\min}\|\nabla f(x^k)\|$; by Lemma 3.3, either $\tau_k = 1$ or $1 > \tau_k \ge 2c\lambda_{\min}\beta(1-\sigma)\|\nabla f(x^k)\|/l$. Substituting this into (27) implies

$$\sum_{k=0}^{\infty} \frac{2c^2\lambda_{\min}^2 \beta(1-\sigma)}{l(l + c\lambda_{\max}\|\nabla f(x^k)\|)^2} \|\nabla f(x^k)\|^4 \le \sum_{k=0}^{\infty} \frac{c\lambda_{\min}\tau_k}{(l + \rho_k \lambda_{\max})^2} \|\nabla f(x^k)\|^3 < \infty.$$

This implies

$$0 \le \frac{2c^2\lambda_{\min}^2(1-\sigma)\beta}{l(l + c\|\nabla f(x^k)\|\lambda_{\max})^2} \|\nabla f(x^k)\|^4 \to 0 \quad \text{as } k \to \infty,$$

which yields $\|\nabla f(x^k)\| \to 0$. Pick any accumulation point $\bar{x}$ of $\{x^k\}$; there exists a subsequence $\{x^{k_j}\}$ such that $x^{k_j} \to \bar{x}$ as $j \to \infty$. By continuity of ∇f,

$$\nabla f(\bar{x}) = \nabla f\left(\lim_{j\to\infty} x^{k_j}\right) = \lim_{j\to\infty} \nabla f(x^{k_j}) = 0,$$

so $\bar{x}$ is a stationary point of (1). This verifies Claim 1.

Next, we suppose $\{x^k\}$ has an accumulation point $\bar{x}$ such that the gradient mapping ∇f is Hölder strongly metrically subregular at $(\bar{x}, 0)$ of order p.

Claim 2: The entire sequence $\{x^k\}$ converges to $\bar{x}$ as $k \to \infty$, and $\bar{x}$ is a $\frac{1+p}{p}$-order sharp minimizer of (1).

The Hölder strong metric subregularity of ∇f implies $\bar{x}$ is an isolated accumulation point, and there exist $\kappa, \delta > 0$ such that

$$\|x - \bar{x}\| \le \kappa \|\nabla f(x)\|^p \quad \text{for all } x \in \mathbb{B}_\delta(\bar{x}). \tag{28}$$

By the monotonicity of $\{f(x^k)\}$, we have $f(\bar{x}) < f(x^0)$. Since f is $C^{1,1}$, there exists a convex neighborhood U of $\bar{x}$ contained in Ω. Because $\partial^2 f(x)$ is positive semidefinite for all $x \in U \subset \Omega$, [6, Theorem 4.2] (while formulated for globally defined functions, its proof remains valid in the local setting) implies f is locally convex in U. For locally convex functions, stationary points are local minimizers. Together with the Hölder strong metric subregularity of ∇f, it follows from [36, Theorem 4.1] that $\bar{x}$ is a $\frac{1+p}{p}$-order sharp minimizer of (1).

For any subsequence $x^{k_j} \to \bar{x}$, [10, Proposition 8.3.10] guarantees full convergence if $\lim_{j\to\infty} \|x^{k_j+1} - x^{k_j}\| = 0$. For large j, $x^{k_j} \to \bar{x}$ and (28) gives $\|x^{k_j} - \bar{x}\| \le \kappa \|\nabla f(x^{k_j})\|^p$. By this and the convexity of f,

$$f(x^{k_j}) - f(\bar{x}) \le \langle \nabla f(x^{k_j}), x^{k_j} - \bar{x} \rangle \le \|\nabla f(x^{k_j})\| \|x^{k_j} - \bar{x}\| \le \kappa \|\nabla f(x^{k_j})\|^{1+p}. \tag{29}$$

Since $\tau_{k_j} \le 1$, combining (26) and (29) yields

$$\begin{aligned}
\|x^{k_j} - x^{k_j+1}\|^2 = \tau_{k_j}^2 \|d^{k_j}\|^2 &\le \frac{\tau_{k_j}}{c\lambda_{\min}\sigma \|\nabla f(x^{k_j})\|} \big(f(x^{k_j}) - f(x^{k_j+1})\big) \\
&\le \frac{\kappa^{\frac{1}{1+p}} \tau_{k_j}}{c\lambda_{\min}\sigma} \big(f(x^{k_j}) - f(\bar{x})\big)^{-\frac{1}{1+p}} \big((f(x^{k_j}) - f(\bar{x})) - (f(x^{k_j+1}) - f(\bar{x}))\big) \\
&\le \frac{\kappa^{\frac{1}{1+p}}}{c\lambda_{\min}\sigma} \Big(\big(f(x^{k_j}) - f(\bar{x})\big)^{\frac{p}{1+p}} - \big(f(x^{k_j+1}) - f(\bar{x})\big)^{\frac{p}{1+p}}\Big).
\end{aligned}$$

This implies $\sum_{j=0}^{\infty} \|x^{k_j} - x^{k_j+1}\|^2 < \infty$, so $\lim_{j\to\infty} \|x^{k_j+1} - x^{k_j}\| = 0$. Claim 2 is thus verified.

Claim 3: For all sufficiently large $k \in \mathbb{N}$, $\|x^k - \bar{x}\| \le O(\|\nabla f(x^k)\|^p)$, $\|\nabla f(x^k)\| \le O\big((f(x^k) - f(\bar{x}))^{\frac{1}{4}}\big)$, and there exists $\eta > 0$ such that

$$f(x^k) - f(\bar{x}) - (f(x^{k+1}) - f(\bar{x})) \ge \eta \big(f(x^k) - f(\bar{x})\big)^{\frac{2}{p}}. \tag{30}$$

By Claim 2 and the Hölder strong metric subregularity of ∇f, there exists $\kappa > 0$ such that (20) holds, so $\|x^k - \bar{x}\| \le O(\|\nabla f(x^k)\|^p)$. Since $\rho_k = c\|\nabla f(x^k)\| \to 0$, (19) together with (21) implies that $\|\nabla f(x^k)\| \le O(\|d^k\|)$ and $f(x^k) - f(\bar{x}) \le O(\|x^k - \bar{x}\|^2)$. Combining these gives

$$\big(f(x^k) - f(\bar{x})\big)^{\frac{2}{p}} \le O(\|\nabla f(x^k)\|^4). \tag{31}$$

From Claim 1, the sequence $\{\nabla f(x^k)\}$ is bounded. Note that $1 \ge \tau_k \ge 2c\|\nabla f(x^k)\|(1 - \sigma)\lambda_{\min}\beta/l$. By (26), we have

$$f(x^k) - f(\bar{x}) - (f(x^{k+1}) - f(\bar{x})) \ge \frac{2c^2\lambda_{\min}^2(1-\sigma)\beta}{l(l + c\|\nabla f(x^k)\|\lambda_{\max})^2} \|\nabla f(x^k)\|^4.$$

Together with (31), this implies $\|\nabla f(x^k)\| \le O\big((f(x^k) - f(\bar{x}))^{\frac{1}{4}}\big)$ and the existence of $\eta > 0$ satisfying (30), verifying Claim 3.

Let $\alpha_k := f(x^k) - f(\bar{x})$; by (30), we derive

$$\alpha_{k+1} \le \alpha_k - \eta \alpha_k^{\frac{2}{p}}.$$

Using arguments similar to those in Theorem 3.6, the convergence results follow directly from Lemma 2.4 with $q = \frac{2}{p}$. This completes the proof. ■

Remark 3.9. *Under the assumption that the second-order subdifferential is positive semidefinite, reference [33] conducts convergence rate analysis of the algorithm under study by incorporating the assumptions of Hölder metric subregularity and p-order semismooth* property, while reference [1] does so by adding the assumptions of strong metric subregularity and semismooth differentiability. In contrast, Theorem 3.8 establishes an explicit quantitative relationship between the algorithm's convergence rate and the order of Hölder strong metric subregularity, relying solely on the assumption of Hölder strong metric subregularity without requiring additional semismoothness conditions.*

4 Coderivative-based positive definite regularized Newton algorithm for Prox-regular functions

In this section, we consider the nonsmooth unconstrained optimization problem (1) where the objective function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is continuously prox-regular. We focus on designing and justifying a coderivative-based positive definite regularized Newton algorithm for solving the subgradient inclusion

$$0 \in \partial f(x). \tag{32}$$

Adopting the methodology from [14, 15], we convert the subgradient inclusion (32) into a gradient system by substituting f—which belongs to the class of continuously prox-regular functions—with its Moreau envelope, a function that has been shown to be of class $C^{1,1}$.

To begin, we present the definitions of Moreau envelopes and proximal mappings related to extended real-valued functions (see, e.g., [31]). Given a lower semicontinuous function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a parameter $\gamma > 0$, the *Moreau envelope* $e_{\gamma f}$ and the *proximal mapping* $\text{Prox}_{\gamma f}$ are defined respectively as

$$e_{\gamma f}(x) := \inf_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}, \tag{33}$$

$$\text{Prox}_{\gamma f}(x) := \mathop{\arg\min}_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}. \tag{34}$$

These concepts have been extensively studied in variational analysis and optimization, serving as effective tools for regularizing and approximating nonsmooth functions. Recall that I denotes the identity operator, and a function f is prox-bounded if it is bounded below by a quadratic function on $\mathbb{R}^n$. For convenience, for $\gamma \in (0, \infty)$, we define

$$U_\gamma := \text{rge}(I + \gamma \partial f). \tag{35}$$

The following result regarding prox-regular functions (see [14, Lemma 6.3]) plays a crucial role in the subsequent justification of the algorithm.

Lemma 4.1. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be prox-bounded on $\mathbb{R}^n$ and continuously prox-regular at $\overline{x}$ for $\overline{v} \in \partial f(\overline{x})$ with modulus $r > 0$. The following assertions hold for all $\gamma \in (0, r^{-1})$; if f is convex, the parameter γ can be arbitrarily chosen from $(0, \infty)$ with $U_\gamma = \mathbb{R}^n$:*

(i) The Moreau envelope $e_{\gamma f}$ is of class $C^{1,1}$ on U_γ (defined in (35)), which contains a neighborhood of $\overline{x} + \gamma\overline{v}$. Moreover, $\overline{x}$ solves the subgradient inclusion (32) if and only if $\nabla e_{\gamma f}(\overline{x}) = 0$.

(ii) The proximal mapping $\mathrm{Prox}_{\gamma f}$ is single-valued, monotone, and Lipschitz continuous on U_γ, satisfying $\mathrm{Prox}_{\gamma f}(\overline{x} + \gamma\overline{v}) = \overline{x}$.

(iii) The gradient of $e_{\gamma f}$ is given by

$$\nabla e_{\gamma f}(x) = \frac{1}{\gamma}\big(x - \mathrm{Prox}_{\gamma f}(x)\big) = \big(\gamma I + \partial f^{-1}\big)^{-1}(x) \quad \text{for all } x \in U_\gamma. \tag{36}$$

It follows from Lemma 4.1 that using the Moreau envelope (33) enables transforming the solution of the subgradient inclusion (32) into finding the stationary point of the $C^{1,1}$ optimization problem $\min_{x\in\mathbb{R}^n} e_{\gamma f}(x)$.

To apply the corresponding results for $C^{1,1}$ problems established in Sects. 3, we need to characterize the generalized Hessian of the Moreau envelope $e_{\gamma f}$; see [14, Lemma 6.4].

Lemma 4.2. *In the setting of Lemma 4.1, for any $\gamma \in (0, r^{-1})$, $x \in U_\gamma$, and $y = \nabla e_{\gamma f}(x)$ we have the equivalence*

$$(u, w) \in \mathrm{gph}(D^*\nabla e_{\gamma f})(x, y) \iff (u - \gamma w, w) \in \mathrm{gph}\,\partial^2 f(x - \gamma y, y).$$

Next, we present the relationships between the lower-definiteness and Hölder strong metric subregularity of the original objective f and the corresponding properties of the Moreau envelope $e_\gamma f$.

Lemma 4.3. *Let $\xi \in \mathbb{R}$ and $p \in (0, 1]$. Under the assumptions of Lemma 4.1, for any $\gamma \in (0, r^{-1})$, let $\bar{y} = \nabla e_{\gamma f}(\bar{x})$. Then the following hold:*

(i) If $\partial^2 f(\bar{x} - \gamma\bar{y}, \bar{y})$ is ξ-lower-definite, then $\partial^2 e_{\gamma f}(\bar{x})$ is $\frac{\xi}{1+2\xi\gamma}$-lower-definite.

(ii) If the subdifferential ∂f is Hölder strongly metrically subregular at $(\bar{x} - \gamma\bar{y}, \bar{y})$ of order p, then $\nabla e_{\gamma f}$ is Hölder strongly metrically subregular at $(\bar{x}, \bar{y})$ of order p.

Proof. (i) Assume that $\partial^2 f(\bar{x}-\gamma\bar{y}, \bar{y})$ is ξ-lower-definite. For any $(u, w) \in \mathrm{gph}\partial^2 e_{\gamma f}(\bar{x}, \bar{y}) = \mathrm{gph}(D^*\nabla e_{\gamma f})(\bar{x}, \bar{y})$, Lemma 4.2 implies that $(u - \gamma w, w) \in \mathrm{gph}\,\partial^2 f(\bar{x} - \gamma\bar{y}, \bar{y})$. By the ξ-lower-definiteness assumption, we have

$$\langle w, u - \gamma w\rangle = \langle w, u\rangle - \gamma\|w\|^2 \geq \xi\|u - \gamma w\|^2 = \xi\left(\|u\|^2 - 2\gamma\langle u, w\rangle + \gamma^2\|w\|^2\right).$$

Rearranging terms yields

$$(1 + 2\xi\gamma)\langle w, u\rangle \geq \xi\|u\|^2 + \gamma(1 + \gamma\xi)\|w\|^2 \geq \xi\|u\|^2,$$

showing that $\partial^2 e_{\gamma f}(\bar{x})$ is $\frac{\xi}{1+2\xi\gamma}$-lower-definite.

(ii) By part (i) of Lemma 4.1, $e_{\gamma f} \in C^{1,1}(U_\gamma)$. Let $L > 0$ denote the Lipschitz constant of $\nabla e_{\gamma f}$. Suppose ∂f is Hölder strongly metrically subregular at $(\bar{x} - \gamma\bar{y}, \bar{y})$ of order p, so there exist $\kappa > 0$ and $\delta \in (0, 1/L)$ such that

$$\|u - (\bar{x} - \gamma\bar{y})\| \le \kappa\, d(\bar{y}, \partial f(u))^p \quad \text{for all } u \in \mathbb{B}_\delta(\bar{x} - \gamma\bar{y}). \tag{37}$$

Take any $x \in \mathbb{B}_{\frac{\delta}{1+\gamma L}}(\bar{x})$ and set $y = \nabla e_{\gamma f}(x)$. By part (iii) of Lemma 4.1, $y \in \partial f(x - \gamma y)$. The triangle inequality gives

$$\|x - \gamma y - (\bar{x} - \gamma\bar{y})\| \le \|x - \bar{x}\| + \gamma\|y - \bar{y}\| \le (1 + \gamma L)\|x - \bar{x}\| \le \delta,$$

so $x - \gamma y \in \mathbb{B}_\delta(\bar{x} - \gamma\bar{y})$. Applying (37) yields

$$\|x - \gamma y - (\bar{x} - \gamma\bar{y})\| \le \kappa\, d(\bar{y}, \partial f(x - \gamma y))^p \le \kappa\|y - \bar{y}\|^p.$$

Since $p \in (0, 1]$ and $\delta \in (0, 1/L)$, we have $\|y - \bar{y}\| = \|\nabla e_{\gamma f}(x) - \nabla e_{\gamma f}(\bar{x})\| \le L\|x - \bar{x}\| < 1$, hence $\|y - \bar{y}\| \le \|y - \bar{y}\|^p$. Using the triangle inequality again:

$$\begin{aligned}
\|x - \bar{x}\| &\le \|x - \gamma y - (\bar{x} - \gamma\bar{y})\| + \gamma\|y - \bar{y}\| \\
&\le \kappa\|y - \bar{y}\|^p + \gamma\|y - \bar{y}\| \\
&\le \max\{\kappa, \gamma\}\|y - \bar{y}\|^p \\
&= \max\{\kappa, \gamma\}\|\nabla e_{\gamma f}(x) - \nabla e_{\gamma f}(\bar{x})\|^p,
\end{aligned}$$

which shows that $\nabla e_{\gamma f}$ is Hölder strongly metrically subregular at $(\bar{x}, \bar{y})$ of order p. ■

We are now ready to describe and analyze the proposed coderivative-based positive definite regularized Newton algorithm for solving the subgradient inclusion (32) with a prox-regular objective function.

Algorithm 2: Coderivative-based positive definite regularized Newton algorithm for Prox-regular functions

Require: $x^0 \in \mathbb{R}^n, \gamma \in (0, r^{-1}), \sigma, \beta \in (0, 1),$ and $e_{\gamma f}$ as in (33)

for $k = 0, 1, \ldots$ **do**
  if $x^k = \text{Prox}_{\gamma f}(x^k)$ **then**
    Stop and return x^k
  end if
  Set $v^k := \frac{1}{\gamma}\big(x^k - \text{Prox}_{\gamma f}(x^k)\big)$
  Choose $B_k \succ 0, \rho_k \ge 0$ and $d^k \in \mathbb{R}^n \setminus \{0\}$ such that
  $-v^k \in \partial^2 f(x^k - \gamma v^k, v^k)(\gamma v^k + (I + \gamma\rho_k B_k)d^k) + \rho_k B_k d^k$
  Set $\tau_k = 1$
  while $e_{\gamma f}(x^k + \tau_k d^k) > e_{\gamma f}(x^k) + \sigma\tau_k\langle v^k, d^k\rangle$ **do**
    Set $\tau_k := \beta\tau_k$
  end while
  Set $x^{k+1} := x^k + \tau_k d^k$
end for

Observe that, for convex functions $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, we can choose γ in Algorithm 2 arbitrarily from $(0, \infty)$ with $U_\gamma = \mathbb{R}^n$ due to Lemma 4.1.

We now move on to formulating and proving the key result of this section, which concerns the well-posedness and convergence of the proposed algorithm for prox-regular functions.

Theorem 4.4. *Let $\zeta > 0$, $\xi \in \mathbb{R}$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be the objective function of problem* (1) *with* $\inf f > -\infty$. *Choose an initial point $x_0 \in \mathbb{R}^n$ and define the sublevel set*

$$\Omega_\gamma := \{x \in \mathbb{R}^n \mid e_{\gamma f}(x) \leq e_{\gamma f}(x_0)\}.$$

Assume that f is prox-bounded on $\mathbb{R}^n$ and continuously prox-regular on Ω_γ with modulus $r > 0$, and that the second-order subdifferential $\partial^2 f(x - \gamma y, y)$ is ξ-lower-definite for all $x \in \Omega_\gamma$, where $y = \nabla e_{\gamma f}(x)$ and $\gamma \in (0, r^{-1})$. Suppose further that the eigenvalues of the regularization matrices B_k in Algorithm 2 lie within $[\lambda_{min}, \lambda_{max}]$ with $\lambda_{max} > \lambda_{min} > 0$.

Then Algorithm 2 either terminates at a stationary point, or if $\sup_{k\in\mathbb{N}} \rho_k < \infty$ and $\inf_{k\in\mathbb{N}} \rho_k \geq \frac{\zeta(1+2\xi\gamma)-\xi}{\lambda_{min}(1+2\xi\gamma)}$, it generates sequences $\{x^k\} \subseteq \Omega_\gamma$, $\{d^k\}$, and $\{\tau_k\}$ satisfying:

(i) The sequence $\{e_{\gamma f}(x^k)\}$ is monotonically decreasing and convergent.

(ii) $\sum_{k=0}^{\infty} \|\nabla e_{\gamma f}(x^k)\|^2 < \infty$ and (16) *holds. In particular, if $\bar{x}$ is an accumulation point of $\{x^k\}$, then $\bar{x}$ solves the subgradient inclusion* (32) *and satisfies $e_{\gamma f}(\bar{x}) = \inf_{k\in\mathbb{N}} e_{\gamma f}(x^k)$.*

(iii) If $\{x^k\}$ has an isolated accumulation point $\bar{x}$, then the entire sequence $\{x^k\}$ converges to $\bar{x}$ as $k \to \infty$, where $\bar{x}$ solves the subgradient inclusion (32).

Suppose additionally that $\{x^k\}$ has an accumulation point $\bar{x}$ such that the subgradient mapping ∂f is Hölder strongly metrically subregular at $(\bar{x}, 0)$ of order $p \in (0, 1]$. Then:

(a) If $p = 1$, the sequence $\{e_{\gamma f}(x^k)\}$ converges at least Q-linearly, and both $\{x^k\}$ and $\{\nabla e_{\gamma f}(x^k)\}$ converges at least R-linearly.

(b) If $p \in (0, 1)$, then for all sufficiently large $k \in \mathbb{N}$,

$$\|x^k - \bar{x}\| \leq O\left(k^{\frac{p^2}{2(p-1)}}\right) \quad \text{and} \quad \|\nabla e_{\gamma f}(x^k)\| \leq O\left(k^{\frac{p}{2(p-1)}}\right).$$

Proof. By Lemmas 4.1 and 4.2, the conditions in Steps 5 and 6 of Algorithm 2 can be equivalently rewritten as

$$v^k = \nabla e_{\gamma f}(x^k) \quad \text{and} \quad -\nabla e_{\gamma f}(x^k) \in \partial^2 e_{\gamma f}(x^k)(d^k) + \rho_k B_k d^k.$$

Thus, Algorithm 2 reduces to Algorithm 1 with $f := e_{\gamma f}$. By part (i) of Lemma 4.1, solving the subgradient inclusion $0 \in \partial f(x)$ for the prox-regular functions under consideration using Algorithm 1 is equivalent to solving the gradient system $\nabla e_\lambda \varphi(x) = 0$ for the $C^{1,1}$ function $e_\lambda \varphi$ via Algorithm 2 with the indicated parameter choices.

Under the given assumptions, part (i) of Lemma 4.3 implies that $\partial^2 e_{\gamma f}(x)$ is $\frac{\xi}{1+2\xi\gamma}$-lower-definite for all $x \in \Omega_\gamma$. Consequently, assertions (i), (ii), and (iii) follow directly from Lemma 4.1 and Theorem 3.4. And then $\bar{y} = \nabla e_{\gamma f}(\bar{x}) = 0$.

Suppose additionally that $\{x^k\}$ has an accumulation point $\bar{x}$ such that the subgradient mapping ∂f is Hölder strongly metrically subregular at $(\bar{x}, 0)$ of order $p \in (0, 1]$. By part (ii) of Lemma 4.3, $\nabla e_{\gamma f}$ is Hölder strongly metrically subregular at $(\bar{x}, 0)$ of order p. The convergence rate results (a) and (b) then follow from Theorem 3.6. ■

5 Application in composite optimization

In this section, we consider the convex composite optimization problem given by

$$\text{minimize } f(x) := \varphi(x) + \psi(x), \quad x \in \mathbb{R}^n, \tag{38}$$

where $\varphi : \mathbb{R}^n \to \mathbb{R}$ is a convex smooth function, and the regularizer $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex and extended-real-valued.

Problems of the form (38) frequently arise in various applied fields, including machine learning, compressed sensing, and image processing. To develop new regularized Newton methods for solving (38), we first recall the concept known as the *forward-backward envelope*, introduced by Patrinos and Bemporad [29] for convex composite optimization problems.

Let $f = \varphi + \psi$ be as in (38), and let $\gamma > 0$. The forward-backward envelope (FBE) of f with parameter γ is defined as

$$f_\gamma(x) := \inf_{y\in\mathbb{R}^n} \left\{ \varphi(x) + \langle \nabla\varphi(x), y - x\rangle + \psi(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}. \tag{39}$$

By the construction of the Moreau envelope in (33), the forward-backward envelope can be expressed as

$$f_\gamma(x) = \varphi(x) - \frac{\gamma}{2}\|\nabla\varphi(x)\|^2 + e_{\gamma\psi}(x - \gamma\nabla\varphi(x)). \tag{40}$$

The forward-backward envelope (FBE) has been widely adopted in the development of efficient algorithms for solving nonsmooth optimization problems (see, e.g., [29, 32, 34]). The results presented below, adapted from [29, 32], summarize the core properties of the FBE for convex composite extended-real-valued functions—properties that are indispensable for deriving the main results of this section.

Lemma 5.1. *Let $f = \varphi + \psi$ be as in* (38)*, and let $\gamma > 0$. Suppose φ is C^2-smooth on $\mathbb{R}^n$, with $\nabla\varphi$ Lipschitz continuous on $\mathbb{R}^n$ and modulus $\ell > 0$. Then:*

(*i*) *The FBE f_γ of f is C^1-smooth on $\mathbb{R}^n$, with gradient*

$$\nabla f_\gamma(x) = \gamma^{-1}\left(I - \gamma\nabla^2\varphi(x)\right)\left(x - \mathit{Prox}_{\gamma\psi}(x - \gamma\nabla\varphi(x))\right), \quad x \in \mathbb{R}^n. \tag{41}$$

Moreover, the set of optimal solutions to (38) *coincides with the stationary points of f_γ, i.e.,*

$$\operatorname{argmin} f = \{x \in \mathbb{R}^n \mid \nabla f_\gamma(x) = 0\} \quad \text{for all } \gamma \in (0, 1/\ell).$$

(*ii*) *Let $\varphi(x) := \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + \alpha$, where $A \in \mathbb{R}^{n\times n}$ is a positive-semidefinite symmetric matrix, $b \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$. For all $\gamma \in (0, 1/\ell)$, the FBE f_γ is convex, and its gradient ∇f_γ is globally Lipschitz continuous on $\mathbb{R}^n$ with modulus $L := 2\left(1 - \gamma\lambda_{\min}(A)\right)/\gamma$.*

To better study the application of the regularized Newton method from Section 3 to Lasso problems, we consider in this section only the case where ψ in (38) is a quadratic function, in which case problem (38) takes the form

$$\text{minimize } f(x) := \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + \alpha + \psi(x), \quad x \in \mathbb{R}^n, \tag{42}$$

where $A \in \mathbb{R}^{n\times n}$ is a positive-semidefinite symmetric matrix, $b \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$.

It is worth emphasizing that problems of the form (42) are significant in their own right. Moreover, they frequently emerge in various efficient numerical algorithms, including sequential quadratic programming (SQP) methods [12], augmented Lagrangian methods [18], proximal Newton methods [16, 21], and constrained quadratic optimization [26].

Lemma 5.1 implies that using the forward-backward envelope (39) enables transforming the nonsmooth composite optimization problem (38) into the unconstrained problem $\min_{x\in\mathbb{R}^n} f_\gamma(x)$. The explicit expressions for f_γ in (40) and its gradient in (41) allow us to apply the results from Sects. 3 to the class of nondifferentiable convex problems (42). To proceed, we need to compute the generalized Hessian of the FBE (see [13, Proposition 4]).

Lemma 5.2. *Let $f = \varphi + \psi$ be as in (42), and let $\gamma > 0$ be such that $B := I - \gamma A$ is positive-definite. For any $\bar{x} \in \mathbb{R}^n$, $\bar{d} \in \mathbb{R}^n$, and $\bar{u} := \bar{x} - \gamma(A\bar{x} + b)$, the following equivalence holds:*

$$\bar{z} \in \partial^2 f_\gamma(\bar{x})(\bar{d}) \iff B^{-1}\bar{z} - A\bar{d} \in \partial^2\psi\left(\mathrm{Prox}_{\gamma\psi}(\bar{u}), \frac{1}{\gamma}\left(\bar{u} - \mathrm{Prox}_{\gamma\psi}(\bar{u})\right)\right)(\bar{d} - \gamma B^{-1}\bar{z}).$$

Below we present the relationship between Hölder strongly metrically subregularity associated with the objective function f and the FBE f_γ.

Lemma 5.3. *Let $f = \varphi + \psi$ be as in (42), where $A \in \mathbb{R}^{n\times n}$ is a positive-semidefinite symmetric matrix. Let $p \in (0, 1]$ and $\gamma > 0$ be such that $B := I - \gamma A$ is positive-definite. Then, if ∂f is Hölder strongly metrically subregular at $(\mathrm{Prox}_{\gamma\psi}(\bar{u}), \bar{y})$ of order p, it follows that ∇f_γ is Hölder strongly metrically subregular at $(\bar{x}, \bar{y})$ of order p, where $\bar{x} \in \mathbb{R}^n$, $\bar{u} = \bar{x} - \gamma(A\bar{x} + b)$, and $\bar{y} = \nabla f_\gamma(\bar{x})$.*

Proof. For any $x \in \mathbb{R}^n$, let $u_x := x - \gamma(Ax + b) = Bx - \gamma b$ and $y_x := \nabla f_\gamma(x)$. By Lemma 5.1, we have

$$y_x = \gamma^{-1}B\left(x - \mathrm{Prox}_{\gamma\psi}(u_x)\right).$$

Since $B := I - \gamma A$ is positive-definite, it follows that $x - \gamma B^{-1}y_x = \mathrm{Prox}_{\gamma\psi}(u_x)$. Given the convexity of ψ, convex analysis implies that

$$\frac{1}{\gamma}\left(u_x - \mathrm{Prox}_{\gamma\psi}(u_x)\right) \in \partial\psi\left(x - \gamma B^{-1}y_x\right) = \partial\psi\left(\mathrm{Prox}_{\gamma\psi}(u_x)\right).$$

Noting that

$$A\mathrm{Prox}_{\gamma\psi}(u_x) + b + \frac{1}{\gamma}\left(u_x - \mathrm{Prox}_{\gamma\psi}(u_x)\right) = \frac{1}{\gamma}\left(\gamma A\mathrm{Prox}_{\gamma\psi}(u_x) + Bx - \mathrm{Prox}_{\gamma\psi}(u_x)\right) = y_x,$$

we conclude

$$y_x \in \nabla\varphi\left(\mathrm{Prox}_{\gamma\psi}(u_x)\right) + \partial\psi\left(\mathrm{Prox}_{\gamma\psi}(u_x)\right) = \partial f\left(\mathrm{Prox}_{\gamma\psi}(u_x)\right).$$

In particular, $\bar{y} = y_{\bar{x}} \in \partial f\left(\mathrm{Prox}_{\gamma\psi}(\bar{u})\right)$.

By the assumption of Hölder strongly metrically subregularity, there exist $\kappa > 0$ and $\delta \in (0, 1)$ such that

$$\|w - \mathrm{Prox}_{\gamma\psi}(\bar{u})\| \le \kappa\, d\left(\bar{y}, \partial f(w)\right)^p \quad \text{for all } w \in B\left(\mathrm{Prox}_{\gamma\psi}(\bar{u}), \delta\right). \tag{43}$$

By part (ii) of Lemma 4.1, $\mathrm{Prox}_{\gamma\psi}$ is globally Lipschitz continuous on $\mathbb{R}^n$. Let $l > 0$ denote its Lipschitz modulus. For any $x \in B\left(\bar{x}, \frac{\delta}{l(1+\lambda_{\max}(A))}\right)$, the triangle inequality gives

$$\|\mathrm{Prox}_{\gamma\psi}(u_x) - \mathrm{Prox}_{\gamma\psi}(\bar{u})\| \le l\|u_x - \bar{u}\| = l\|x - \bar{x} + \gamma(Ax - A\bar{x})\| \le l\left(1 + \lambda_{\max}(A)\right)\|x - \bar{x}\| < \delta.$$

From (43), it follows that

$$\|\operatorname{Prox}_{\gamma\psi}(u_x) - \operatorname{Prox}_{\gamma\psi}(\bar{u})\| \le \kappa\, d\left(\bar{y}, \partial f\left(\operatorname{Prox}_{\gamma\psi}(u_x)\right)\right)^p \le \kappa \|\nabla f_\gamma(x) - \nabla f_\gamma(\bar{x})\|^p. \tag{44}$$

On the other hand, we have

$$\begin{aligned}\|\operatorname{Prox}_{\gamma\psi}(u_x) - \operatorname{Prox}_{\gamma\psi}(\bar{u})\| &= \|x - \gamma B^{-1} y_x - \bar{x} + \gamma B^{-1}\bar{y}\| \\ &\ge \|x - \bar{x}\| - \gamma \|B^{-1}(y_x - \bar{y})\| \\ &\ge \|x - \bar{x}\| - \gamma \lambda_{\max}(B^{-1}) \|y_x - \bar{y}\| \\ &\ge \|x - \bar{x}\| - \frac{\gamma}{1 - \gamma\lambda_{\max}(A)} \|y_x - \bar{y}\|^p,\end{aligned}$$

where the last inequality uses $B := I - \gamma A$ and $p \in (0, 1]$.

Combining this with (44) implies that ∇f_γ is Hölder strongly metrically subregular at $(\bar{x}, \bar{y})$ of order p. ■

We are now ready to describe and justify the proposed regularized Newton method for solving the convex composite optimization problem (42), where we set $B_k \equiv B := I - \gamma A$.

Algorithm 3: Coderivative-based regularized Newton algorithm for convex composite Functions

Require: $x^0 \in \mathbb{R}^n, \gamma > 0$ such that $B := I - \gamma A \succ 0, \sigma, \beta \in (0,1)$, and f_γ as in (40)

for $k = 0, 1, \ldots$ **do**
  if $\nabla f_\gamma(x^k) = 0$ **then**
    Stop and return x^k
  end if
  Set $u^k := x^k - \gamma(Ax^k + b)$, $v^k := \operatorname{Prox}_{\gamma\psi}(u^k)$
  Choose $\rho_k \ge 0$ and $d^k \in \mathbb{R}^n \setminus \{0\}$ such that

$$-\frac{1}{\gamma}(x^k - v^k) \in \partial^2\psi\left(v^k, \frac{1}{\gamma}(u^k - v^k)\right)\left(x^k - v^k + (1 + \gamma\rho_k)d^k\right) + \rho_k d^k + Ad^k$$

  Set $\tau_k = 1$
  while $f_\gamma(x^k + \tau_k d^k) > f_\gamma(x^k) + \sigma\tau_k \langle \nabla f_\gamma(x^k), d^k\rangle$ **do**
    Set $\tau_k := \beta\tau_k$
  end while
  Set $x^{k+1} := x^k + \tau_k d^k$
end for

The explicit forms of the sequences $\{v^k\}$ and $\{d^k\}$ in Algorithm 3 are determined by the particular structures of the regularizers ψ, which are effectively defined in practical models of the Lasso problem (see, for instance, Sect. 6). The theorem below establishes both the well-posedness and the convergence rate of Algorithm 3 when applied to the class of convex composite optimization problems (42).

Theorem 5.4. *Let $\zeta > 0$, $\xi \in \mathbb{R}$, and $f = \varphi + \psi$ be the objective function of problem* (42) *with $\inf f > -\infty$. Let $A \in \mathbb{R}^{n\times n}$ be a positive-semidefinite symmetric matrix and $\gamma \in (0, 1/\lambda_{max}(A))$.*

Then, for any initial point $x_0 \in \mathbb{R}^n$, Algorithm 3 either terminates at a stationary point, or if $\sup_{k\in\mathbb{N}} \rho_k < \infty$ and $\inf_{k\in\mathbb{N}} \rho_k \ge \frac{\zeta}{\lambda_{min}}$, it generates sequences $\{x^k\}$, $\{d^k\}$, and $\{\tau_k\}$ satisfying:

(i) The sequence $\{f_\gamma(x^k)\}$ is monotonically decreasing and convergent.

(ii) $\sum_{k=0}^{\infty} \|\nabla f_\gamma(x^k)\|^2 < \infty$ and (16) holds. In particular, if $\bar{x}$ is an accumulation point of $\{x^k\}$, then $\bar{x}$ solves optimization problem (42) and satisfies $f_\gamma(\bar{x}) = \inf_{k\in\mathbb{N}} f_\gamma(x^k)$.

(iii) If $\{x^k\}$ has an isolated accumulation point $\bar{x}$, then the entire sequence $\{x^k\}$ converges to $\bar{x}$ as $k \to \infty$, where $\bar{x}$ solves optimization problem (42).

Suppose additionally that $\{x_k\}$ has an accumulation point $\bar{x}$ such that the subgradient mapping ∂f is Hölder strongly metrically subregular at $(\text{Prox}_{\gamma\psi}(\bar{u}), 0)$ of order $p \in (0,1]$, where $\bar{u} = \bar{x} - \gamma(A\bar{x} + b)$. Then:

(a) If $p = 1$, the sequence $f_\gamma(x^k)$ converges at least Q-linearly, and both $\{x^k\}$ and $\{\nabla f_\gamma(x^k)\}$ converges at least R-linearly.

(b) If $p \in (0,1)$, then for all sufficiently large $k \in \mathbb{N}$,

$$\|x^k - \bar{x}\| \le O\left(k^{\frac{p^2}{2(p-1)}}\right) \quad \text{and} \quad \|\nabla f_\gamma(x^k)\| \le O\left(k^{\frac{p}{2(p-1)}}\right).$$

Proof. By part (ii) of Lemma 5.1, solving optimization problem (42) is equivalent to solving the gradient system $\nabla f_\gamma(x) = 0$. Using Lemma 5.1 and 5.2, the conditions in Steps 5 and 6 of Algorithm 3 can be equivalently rewritten as

$$\nabla f_\gamma(x^k) = \gamma^{-1} B\left(x^k - v^k\right) \quad \text{and} \quad -\nabla f_\gamma(x^k) \in \partial^2 f_\gamma(x^k)(d^k) + \rho_k B d^k.$$

Thus, Algorithm 3 reduces to Algorithm 1 with $f := f_\gamma$ and $B_k \equiv B := I - \gamma A$.

Since $A \in \mathbb{R}^{n\times n}$ is positive-semidefinite and $\gamma \in (0, 1/\lambda_{\max}(A))$, the matrix $B := I - \gamma A$ is positive-definite, and the eigenvalues of regularization matrices B_k in Algorithm 3 lie within $[1 - \gamma\lambda_{\max}(A), 1]$. By part (ii) of Lemma 5.1, the FBE f_γ is convex with ∇f_γ globally Lipschitz continuous on $\mathbb{R}^n$. From Chieu et al. [6, Theorem 3.2], $\partial^2 f_\gamma(x)$ is 0-lower-definite for all $x \in \mathbb{R}^n$. Consequently, assertions (i), (ii), and (iii) follow directly from and Theorem 3.4. And then $\bar{y} = \nabla f_\gamma(\bar{x}) = 0$.

If ∂f is Hölder strongly metrically subregular at $(\text{Prox}_{\gamma\psi}(\bar{u}), 0)$ of order p, Lemma 5.3 implies that ∇f_γ is also Hölder strongly metrically subregular at $(\bar{x}, 0)$ of order p. The stated result then follows directly from Theorems 3.6. ■

6 Numerical experiments

This section specifies Algorithm 3 for the basic Lasso problem below, then presents numerical experiments on this problem and comparisons with some coderivative-based generalized Newton method. The basic *Lasso problem*—also called the ℓ^1-regularized least squares problem—was introduced by Tibshirani [35]. It has since been widely studied and applied in statistics, machine learning, image processing, etc., which is formulated as:

$$\text{minimize } f(x) := \frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_1 \quad \text{subject to } x \in \mathbb{R}^n, \tag{45}$$

where $A \in \mathbb{R}^{m\times n}$, $\mu > 0$, $b \in \mathbb{R}^m$, and $\|\cdot\|_1, \|\cdot\|_2$ are standard p-norms on $\mathbb{R}^n$. The Lasso problem (45) is a convex composite optimization problem (42), as it can be written as minimizing $f(x) = \varphi(x) + \psi(x)$ with

$$\varphi(x) := \frac{1}{2}\langle A^T Ax, x\rangle - \langle A^T b, x\rangle + \frac{1}{2}\|b\|_2^2 \quad \text{and} \quad \psi(x) := \mu\|x\|_1. \tag{46}$$

Notably, $A^T A$ is symmetric and positive-semidefinite, and (45) always has an optimal solution [35].

For the function $\psi(x) := \mu\|x\|_1$, both its proximal mapping and generalized Hessian admit explicit computations (see, e.g., [14, Proposition 8.1]). The proximal mapping $\mathrm{Prox}_{\gamma\psi}$ is given by

$$(\mathrm{Prox}_{\gamma\psi}(x))_i = \begin{cases} x_i - \mu\gamma & \text{if } x_i > \mu\gamma, \\ 0 & \text{if } -\mu\gamma \le x_i \le \mu\gamma, \\ x_i + \mu\gamma & \text{if } x_i < -\mu\gamma, \end{cases} \quad \text{for } i = 1, 2, \ldots, n. \tag{47}$$

For $(x, y) \in \mathrm{gph}\, \partial\psi$ and $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, the generalized Hessian of ψ is calculated by

$$\partial^2\psi(x, y)(v) = \left\{ w \in \mathbb{R}^n \;\middle|\; \left(\frac{1}{\mu} w_i, -v_i\right) \in G\left(x_i, \frac{1}{\mu} y_i\right), \, i = 1, \ldots, n \right\}, \tag{48}$$

where the set-valued mapping $G : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ is specified as

$$G(t, p) := \begin{cases} \{0\} \times \mathbb{R} & \text{if } t \neq 0 \text{ and } p \in \{-1, 1\}, \\ \mathbb{R} \times \{0\} & \text{if } t = 0 \text{ and } p \in (-1, 1), \\ (\mathbb{R}_+ \times \mathbb{R}_-) \cup (\{0\} \times \mathbb{R}) \cup (\mathbb{R} \times \{0\}) & \text{if } t = 0 \text{ and } p = -1, \\ (\mathbb{R}_- \times \mathbb{R}_+) \cup (\{0\} \times \mathbb{R}) \cup (\mathbb{R} \times \{0\}) & \text{if } t = 0 \text{ and } p = 1, \\ \emptyset & \text{otherwise.} \end{cases} \tag{49}$$

To implement Algorithm 3, we derive explicit forms for the sequences $\{v^k\}$ and $\{d^k\}$ as follows: $u^k = x^k - \gamma\left(A^T A x^k + A^T b\right)$ and, for $i = 1, 2, \ldots, n$, the sequence v^k follows from (47) that

$$\left(v^k\right)_i = \begin{cases} (u^k)_i - \mu\gamma & \text{if } (u^k)_i > \mu\gamma, \\ 0 & \text{if } -\mu\gamma \le (u^k)_i \le \mu\gamma, \\ (u^k)_i + \mu\gamma & \text{if } (u^k)_i < -\mu\gamma. \end{cases}$$

To determine d^k, using the expression for $\partial^2\psi$ from (48) and (49), we arrive at the following conditions:

$$\begin{cases} \left(-\frac{1}{\gamma}(x^k - v^k) - (\rho_k I + A^T A) d^k\right)_i = 0 & \text{if } (v^k)_i \neq 0, \\ \left(x^k - v^k + (1 + \gamma\rho_k) d^k\right)_i = 0 & \text{if } (v^k)_i = 0. \end{cases}$$

Our proposed algorithm is compared with three coderivative-based generalized Newton method for solving the LASSO problem, namely the coderivative-based generalized damped Newton method (GDNM), the coderivative-based generalized regularized Newton method (GRNM) in [13], and the generalized regularized Newton method with a positive definite regularization term (GRNM-PD) in [33].

All numerical experiments were conducted in MATLAB on a desktop computer configured with a 12th Gen Intel® Core™ i7-12700 processor (2.10 GHz), an Intel® UHD Graphics 770, an NVIDIA GeForce RTX 3060 graphics card, and 32 GB of random access memory (RAM). For the LASSO problem in our numerical tests, the regularization parameter and step size parameter are set to $\mu = 10^{-3}$ and $\gamma = 10^{-2}$, respectively, with the corresponding numerical results reported in Table 1. All algorithms are initialized at the zero point $x^0 := 0$ across all experiments, and the accuracy of the approximate optimal solution x^k for (45) is evaluated by the relative KKT residual η_k proposed in [18]:

$$\eta_k := \frac{\left\| x^k - \text{Prox}_{\mu\psi}\left(x^k - A^T A x^k + A^T b\right) \right\|_2}{1 + \|x^k\|_2 + \|Ax^k - b\|_2}.$$

Algorithms terminate when either $\eta_k < 10^{-6}$ or the maximum iterations of 10,000 is reached.

Table 1: Solving LASSO problems on test instances

Problem	Measure	Algorithm			
$m; n$		GDNM	GRNM	GRNM-PD	PDRNM
1000; 1000	Time	341.43s	2434.51s	1450.61s	896.46s
	Iter	1552	6196	6422	1268
200; 200	Time	0.07s	29.95s	27.29s	0.32s
	Iter	18	4869	4510	58
1000; 20	Time	0.00s	1.01s	0.00s	0.00s
	Iter	1	526	4	1
1000; 500	Time	0.05s	47.79s	78.57s	0.05s
	Iter	2	704	1183	2
20; 1000	Time	1803.63s	95.04s	82.05s	100.98s
	Iter	10000	1345	1375	1515
500; 1000	Time	1477.59s	113.31s	105.52s	110.23s
	Iter	10000	823	813	876

Different from the way that ρ_k is directly generated by $\|\nabla f_\gamma(x^k)\|$ in the GDNM, GRNM and GRNM-PD algorithms, ρ_k in our proposed Algorithm 3 can be preassigned, which can effectively save the computation time. In the numerical experiments, when the data matrix $A \in \mathbb{R}^{m\times n}$ is overdetermined $m > n$, PDRNM is competitive with the other tested algorithms except GDNM. When the data matrix $A \in \mathbb{R}^{m\times n}$ is underdetermined $m < n$, the GDNM algorithm is not applicable, while the performance of PDRNM is comparable to that of the other tested algorithms.

7 Conclusion

This paper focuses on $C^{1,1}$ optimization problems and investigates a coderivative-based positive definite regularized Newton method. Under the assumption of Hölder strong metric subregularity, we quantify the precise relationship between the convergence rate of the algorithm and the order of Hölder strong metric subregularity. When the objective function is a prox-regular function, a corresponding algorithm has also been developed via its Moreau envelope. To explore the application of the proposed algorithm in Lasso problems, we also consider a special class of composite convex optimization problems. Numerical experiments show that appropriately selecting the regularization matrix is beneficial for improving the convergence speed of the algorithm. In the future, we plan to further accelerate the convergence speed of the algorithm under appropriate assumptions.

Statements and Declarations

- Data Availability Statement: The data that support the findings of this study will be made available by the authors upon publication.
- Competing Interests: The authors declare no conflicts of interest.

- Funding: This work was supported by the National Natural Science Foundation of the People's Republic of China [Grant 12261109 and Grant12571331] and the Applied Basic Research Program of Yunnan Province [Grant 202301AT070080].

Acknowledgments

The authors would like to express their appreciation to the reviewers and the editor for their time and comments.

References

[1] Aragón-Artacho, F.J., Mordukhovich, B.S., Pérez-Aros, P. Coderivative-based semi-Newton method in nonsmooth difference programming. *Math. Program.* **213**, 385-432 (2025)

[2] Bauschke, H.H., Combettes, P.L. Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd edn. Springer, New York (2017)

[3] Borwein, J.M., Li, G., Yao, L. Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets. *SIAM J. Optim.* **24**, 498–527 (2014)

[4] Bonnans, J.F. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Appl. Math. Optim.* **29**, 161–186 (1994)

[5] Beck, A. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB.* SIAM, Philadelphia, PA (2014)

[6] Chieu, N., Chuong, T., Yao, J.C., Yen, N. Characterizing convexity of a function by its Fréchet and limiting second-order subdifferentials. *Set-Valued Var. Anal.* **19**(1), 75–96 (2011)

[7] Cibulka, R., Dontchev, A.L., Kruger, A.Y. Strong metric subregularity of mappings in variational analysis and optimization. *J. Math. Anal. Appl.* **457**, 1247–1282 (2018)

[8] Dan, H., Yamashita, N., Fukushima, M. Convergence properties of the inexact Levenberg–Marquardt method under local error bound conditions. *Optim. Meth. Softw.* **17**, 605–626 (2002)

[9] Dontchev, A.L., Rockafellar, R.T. *Implicit Functions and Solution Mappings: A View from Variational Analysis*, 2nd edn. Springer, New York (2014)

[10] Facchinei, F., Pang, J.-S. Finite-Dimensional Variational Inequalities and Complementarity Problems, I, II. Springer, New York (2003)

[11] Gfrerer, H., Outrata, J.V. On a semismooth* Newton method for solving generalized equations. *SIAM J. Optim.* **31**, 489–517 (2021)

[12] Izmailov, A.F., Solodov, M.V. Newton-Type Methods for Optimization and Variational Problems. Springer, New York (2014)

[13] Khanh, P.D., Mordukhovich, B.S., Phat, V.T., Tran, D.B. Globally convergent coderivative-based generalized Newton methods in nonsmooth optimization. *Math. Program.* **205**, 373–429 (2024)

[14] Khanh, P.D., Mordukhovich, B.S., Phat, V.T. A generalized Newton method for subgradient systems. *Math. Oper. Res.* **48**, 1811–1845 (2023)

[15] Khanh, P.D., Mordukhovich, B.S., Phat, V.T., Tran, D.B. Generalized damped Newton algorithms in nonsmooth optimization via second-order subdifferentials. *J. Global Optim.* **86**, 93–122 (2023)

[16] Lee, J.D., Sun, Y., Saunders, M.A. Proximal Newton-type methods for minimizing composite functions. *SIAM J. Optim.* **24**, 1420–1443 (2014)

[17] Li, D.H., Fukushima, M., Qi, L., Yamashita, N. Regularized Newton methods for convex minimization problems with singular solutions. *Comput. Optim. Appl.* **28**, 131–147 (2004)

[18] Li, X., Sun, D., Toh, K.-C. A highly efficient semismooth Newton augmented Lagrangian method for solving Lasso problems. *SIAM J. Optim.* **28**, 433–458 (2018)

[19] Mordukhovich, B.S. Sensitivity analysis in nonsmooth optimization. In: Field, D.A., Komkov, V. (eds) *Theoretical Aspects of Industrial Design*, pp. 32–46. *SIAM Proc. Appl. Math.* **58**. Philadelphia, PA (1992)

[20] Mordukhovich, B.S., Sarabi, M.E. Generalized Newton algorithms for tilt-stable minimizers in nonsmooth optimization. *SIAM J. Optim.* **31**, 1184–1214 (2021)

[21] Mordukhovich, B.S., Yuan, X., Zheng, S., Zhang, J. A globally convergent proximal Newton-type method in nonsmooth convex optimization. *Math. Program.* **198**, 899–936 (2023)

[22] Mordukhovich, B.S. Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications. Springer, Berlin (2006)

[23] Mordukhovich, B.S. Variational Analysis and Applications. Springer, Cham (2018)

[24] Nam, N.M. Coderivatives of normal cone mappings and Lipschitzian stability of parametric variational inequalities. *Nonlinear Anal.* **73**, 2271–2282 (2010)

[25] Nesterov, Yu. Lectures on Convex Optimization, 2nd edn. Springer, Cham (2018)

[26] Nocedal, J., Wright, S. Numerical Optimization. Springer, New York (2006)

[27] Ostrowski, A.M. Solution of Equations and Systems of Equations, 2nd edn. Academic Press, Cambridge (1966)

[28] Ouyang, W., Li, L. Hölder strong metric subregularity and its applications to convergence analysis of inexact Newton methods. *J. Ind. Manag. Optim.* **17**(1), 169–184 (2021)

[29] Patrinos, P., Bemporad, A. Proximal Newton methods for convex composite optimization. In: IEEE Conference on Decision and Control, pp. 2358–2363 (2013)

[30] Poliquin, R.A., Rockafellar, R.T. Tilt stability of a local minimum. *SIAM J. Optim.* **8**, 287–299 (1998)

[31] Rockafellar, R.T., Wets, R.J.-B. Variational Analysis. Springer, Berlin (1998)

[32] Stella, L., Themelis, A., Patrinos, P. Forward-backward quasi-Newton methods for nonsmooth optimization problems. *Comput. Optim. Appl.* **67**, 443–487 (2017)

[33] Shi, Z.J., Chao, M.T. Generalized Newton Method with Positive Definite Regularization for Nonsmooth Optimization Problems with Nonisolated Solutions. *J. Optim. Theory Appl.* **201**, 396–432 (2024)

[34] Stella, L., Themelis, A., Patrinos, P. Forward-backward envelope for the sum of two nonconvex functions: further properties and nonmonotone linesearch algorithms. *SIAM J. Optim.* **28**, 2274–2303 (2018)

[35] Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc.* **58**, 267–288 (1996)

[36] Zheng, X.Y., Ng, K.F. Hölder stable minimizers, tilt stability, and Hölder metric regularity of subdifferentials. *SIAM J. Optim.* **25**, 416–438 (2015)

[37] Zhang, B., Zhou, C., Zhu, J. Convergence analysis of the generalized Douglas-Rachford splitting method under Hölder subregularity assumptions. *J. Ind. Manag. Optim.* **19**, 5060–5077 (2023)