Minimax Theory and its Applications
Volume 11 (2026), No. 4s, 240 to 252

Error-bound analysis and convergence characterization of numerical optimization algorithms in engineering systems

S. Arul Murugan[1], Dr. Padmavati Shrivastava[2], S. Jagatheeswaran[3], Anandakumar Haldorai[4], Manish Nandy[5], Divya[6]

[1]Assistant Professor, Department of Information Technology, K.S. Rangasamy College of Technology, Namakkal, Tamil Nadu, India. arulmurugans@ksrct.ac.in, https://orcid.org/0009-0004-3694-1096

[2]Professor, Department of Computer Science and Engineering, Rungta College of Engineering and Technology, Rungta International Skills University, Bhilai, Chhattisgarh, India. dr.padmavati.shrivastava@rungta.org, https://orcid.org/0000-0002-4500-3905

[3]Assistant Professor, Department of Computer Science and Engineering, J.J. College of Engineering and Technology, Tiruchirappalli, Tamil Nadu, India. jagatheeswaranss@jjcet.ac.in, https://orcid.org/0009-0000-4058-853X

[4]Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India. anandakumar.psgtech@gmail.com, https://orcid.org/0000-0001-9975-6462

[5]Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India. ku.manishnandy@kalingauniversity.ac.in, https://orcid.org/0009-0001-0653-2556

[6]Department of MBA, Ramachandra College of Engineering, Eluru, Andhra Pradesh, India. divya@rcee.ac.in

Received xx xxxx, xxxx
Accepted xx xxxx, xxxx
Published xx xxxx, xxxx

Numerical optimization methods are widely applied in engineering systems in solving complicated engineering problems related to nonlinear modeling, structural design, parameter estimation, and constrained optimization. However, attaining highly accurate solutions with assured convergence is difficult because of issues with error propagation, computational complexity, and sensitivity to system parameters. This study proposes an analysis approach for estimating the error bounds and convergence properties of numerical optimization methods in engineering applications. Mathematical conditions such as smoothness of objective function, strong convexity, Polyak-Łojasiewicz inequality, Lipschitz property of Hessian, and composite optimization formulations are used in formulating convergence results and error properties. Sufficient decrease conditions, linear convergence bounds of gradient methods, explicit error-distance relations, local quadratic convergence of Newton methods, and convergence of proximal-gradient optimization are formulated theoretically. The numerical validation process relies on the analysis of structural compliance optimization and nonlinear stiffness-calibration problems. From the gradient-based optimization analysis, geometric error minimization is observed from a function value difference that ranges from 5.175×10^{-2} to 1.694×10^{-14} in 40 iterations; thus, illustrating the reliability of the convergence bound derivation. Error-bound analysis reveals that there is a close relation between theoretical distance-to-optimum and design variable error; hence, showing a decreasing bound ratio during iterative optimization. With respect to the non-linear calibration problem, Newton optimization minimizes the solution error from 5.831×10^{-1} to 6.554×10^{-5}, thereby revealing quadratic convergence after the local convergence region. It is also clear from the convergence analysis that gradient descent has O ($\kappa \log(1/\varepsilon)$) complexity while Newton optimization gives O ($\log \log(1/\varepsilon)$) local convergence efficiency.

Keywords: Error-bound analysis, numerical optimization, convergence characterization, engineering systems, gradient descent, newton optimization, optimization algorithms.

1. Introduction

Nonlinear models of optimization, constrained decisions, resource distribution, structure design, and intelligent computation problems. The performance of these algorithms depends greatly on their convergence properties, the nature of error propagation, and their ability to provide correct results. Error bounds can serve as the theoretical foundation for describing the dependence of optimization errors from measurable algorithm properties, which will facilitate better assessment of both the accuracy and convergence properties of solutions. The field of distributed optimization has shown the importance of error analysis in terms of providing the characteristics of stability and convergence of algorithms for multi-agent systems in engineering (Koshal et al., 2011). Localization and estimation tasks have proved the necessity of employing numerical error bounds to increase the reliability of networked engineering algorithms (Zeng et al., 2022). Convergence analysis of optimization algorithms gives important information about the speed of iterative methods in finding optimal solutions. Coordinate optimization algorithms have proved the importance of error bounds in increasing the speed and efficiency of solution of large-scale optimization problems (Necoara & Clipici, 2016). Optimization algorithms for engineering applications in planning and decision making reveal the necessity of finding the proper balance between the accuracy of the solution and the cost of calculations using search algorithms. Error bound conditions have also been used in the study of convergence properties of new numerical algorithms for the cases of approximate computations and optimization in uncertain environment (Dan et al., 2002). Iterative numerical methods' comparative analysis proves the importance of convergence speed, numerical stability, and error reduction ability for solution of nonlinear equations and optimization problems (Pravin & Sayyed, 2026).

Modern engineering systems require optimization algorithms capable of handling multi-objective objectives, dynamical constraints, and complicated search spaces. The evolutionary optimization approach demonstrates its effectiveness because of multi-objective processing of different conflicting objectives through the use of search space diversity and convergence reliability. Other intelligent approaches to population-based searches like the bee colony algorithm improvement provide another example of the importance of adaptive search techniques in investigating complicated solution spaces to gain better optimization results (Agrawal et al., 2012). Mathematical modeling and numerical simulations of physical processes indicate the necessity of error description for precise computation (Abbasov et al., 2026). Modern global optimization algorithms using the penalty function and adaptation algorithms emphasize the importance of error analysis and convergence reliability for achieving optimal solutions (Tang, 2026). Optimization algorithms of the first order with regularization techniques demonstrate the significance of convergence and error estimations for obtaining optimal solutions (Zhou et al., 2015). Improvement in efficiency of numerical algorithms requires improvement in accuracy and computational complexity of these algorithms. Improvement in numerical integration and optimization methods shows the importance of minimizing errors and enhancing convergence for obtaining precise computation results (Khudhair, 2024). Approximation optimization methods, for example, improved quasi-Newton method, are interested in Hessian matrix computation accuracy and convergence rate (Panic, 2026). Numerical optimization methods oriented on hardware are concerned with precision and algorithm improvement in complex engineering computations (Dai et al., 2025). Other instances of recently developed adaptive multimodal graph learning techniques demonstrate how intelligent computing models can combine multiple data sources while stressing transparency, generalizability, and sound computational results (Shrivastava et al., 2025).

Traditional methods of optimization such as search methods can be improved from a mathematical point of view when considering the possibility to obtain faster convergence and more accurate results (Yıldız). Among new trends in computational optimization, one can mention error-bound analysis used in different engineering branches including inverse problems, reliability analysis, energy systems and quantum optimization. Certification schemes for the theory of inverse problems highlight the necessity of having mathematical assurance for solutions' validation and convergence (Kimura, 2026). Adaptive optimization schemes employing surrogate models demonstrate the necessity of error bounds computation in order to get more accurate results in reliability analysis (Wang et al., 2026). Intelligent optimization schemes for decentralized energy systems have also proved the necessity of convergence characterization in order to have efficient interaction management in such system (Sathish et al., 2026). Computational framework based on quantum dynamics and hybrid optimization has also highlighted the necessity of proper numerical analysis and optimization process stability in future engineering applications (Srichan & Danvirutai, 2025). Optimization algorithms that are based on quantum dynamics and hybrid methods highlight the importance of having proper initialization and convergence control in order to solve the problem of computational burden of complex optimization schemes (Barison et al., 2021; Grant et al., 2019). Hence, error-bound analysis and convergence characterization are the important tools for evaluation, improvement and validation of optimization algorithms in engineering applications. Understanding of behavior of errors, the rate of convergence and stability of the optimization process allows developing robust schemes for optimization.

Key contribution

- The analysis technique for determining the magnitude of error bounding is proposed to measure the gap between the obtained result and optimal result to achieve better insight about the performance of numerical optimization methods on engineering systems with respect to accuracy, stability, and robustness.

- An approach to analyze the rate of convergence and stability of optimization algorithms is presented in order to obtain insights about the performance of the algorithm according to engineering requirements.

- The research provides theoretical as well as practical recommendations for building optimization techniques with better convergence and error bounds in order to make effective decisions in complex engineering systems.

The rest of this paper is structured as follows. In Section II, the notation and preliminary concepts, which include basic assumptions and mathematical definitions essential for the convergence and error bound analysis of numerical optimization algorithms, are introduced. In Section III, the engineering system optimization problem is stated, and the example application scenarios of optimization include structural design, nonlinear parameter estimation, and constrained control design. The main theoretical results are developed in Section IV through analyzing the sufficient decrease property, convergence guarantee, error bound, Newton-type convergence, composite constraint optimization, and robust engineering optimization formulation. In Section V, the convergence rate of various optimization algorithms in terms of iteration complexity and computation cost is analyzed and compared. In Section VI, the numerical validation results from engineering optimization examples are presented to verify the theoretical error bound and convergence characteristics. In Section VII, the engineering significance, limitation, and future direction of the error bound and convergence analysis framework are discussed. Finally, the paper concludes in Section VIII with a summary of the major results from error bound analysis and convergence characterization for reliable numerical optimization in engineering systems.

2. Notation and preliminaries

2.1. Notation

The following notations are used for this paper, as referred to in table 1.

Table1: Notation used for this paper.

Symbol	Meaning
$\mathbb{R}^n$	n-dimensional real Euclidean space, the design/parameter space
x, x_k	decision vector; iterate produced at step k
$f : \mathbb{R}^n \to \mathbb{R}$	objective (engineering performance/cost) function
$\nabla f(x)$	gradient of f at x
$\nabla^2 f(x)$	Hessian of f at x
$\|\cdot\|$	Euclidean norm on $\mathbb{R}^n$ (spectral norm on matrices)
$x*, f*$	a minimizer of f and the optimal value f(x*)
$\Omega*$	the (possibly non-singleton) set of minimizers of f
$dist(x, \Omega*)$	Euclidean distance from x to the nearest point of Ω*
L	Lipschitz constant of ∇f (smoothness constant)
μ	strong-convexity / PL constant
$\kappa = L/\mu$	condition number
M	Lipschitz constant of ∇^2f (Hessian smoothness constant)
$prox_g(\cdot)$	proximal operator of a closed convex function g

2.2. Structural assumptions

The standard set of structural assumptions listed below is applied selectively to the theorems in Section 4, where each theorem lists exactly the required subset.

Definition 1 (L-Smoothness)

A differentiable function $f: \mathbb{R}^n \to \mathbb{R}$ is said to be L-smooth ($L > 0$) if its gradient is L-Lipschitz continuous, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n$$

Definition 2 (μ-Strong Convexity)

f is μ-strongly convex ($\mu > 0$) if

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2}\|y - x\|^2 \quad \text{for all } x, y \in \mathbb{R}^n$$

Definition 3 (Polyak–Łojasiewicz (PL) Inequality)

f satisfies the PL inequality with constant $\mu > 0$ if f is bounded below by $f^* = \inf f$ and

$$\frac{1}{2}\|\nabla f(x)\|^2 \geq \mu(f(x) - f^*) \quad \text{for all } x \in \mathbb{R}^n$$

The PL inequality need not have any convexity assumption because it requires only that the norm of the gradient dominates the error bound. Any μ-strongly convex function automatically satisfies the PL inequality with parameter μ, although the reverse is not true in general; there are other functions (including some non-convex functions) which satisfy the PL inequality, e.g., when take the composition of a convex loss with an invex function.

Definition 4 (Lipschitz Hessian)

f is twice differentiable with M-Lipschitz Hessian if

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq M\|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n$$

Definition 5 (Composite Objective and Proximal Operator)

A composite objective is $F(x) = f(x) + g(x)$, where f is L-smooth and $g: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is closed, convex, and possibly non-smooth (e.g., an indicator function of a physical feasible design set, or an ℓ_1 regularizer promoting sparse actuator placement). The proximal operator of g with step $\eta > 0$ is

$$\text{prox}_{\eta g}(z) = \arg\min_{u \in \mathbb{R}^n} \left\{ g(u) + \frac{1}{2\eta}\|u - z\|^2 \right\}$$

3. Problem formulation for engineering systems

consider the abstract engineering design/estimation problem:

$$\text{(P)} \quad \min_{x \in \mathbb{R}^n} \; F(x) := f(x) + g(x)$$

where f comprises the smooth physics-based term of the objective while g comprises any nonsmooth structures including box constraints on design variables, sparsity-inducing penalties, or safety margin penalties.

Three canonical instances motivate the analysis of Section 4:

- **(E1) Structural compliance / spring-network design.** For a linear-elastic system discretized by n degrees of freedom with stiffness matrix $K \in \mathbb{R}^{n \times n}$ (symmetric positive definite) and applied load b, the compliance-minimization objective under a fixed topology is the quadratic form $f(x) = \frac{1}{2}x^\top Kx - b^\top x$, whose gradient $\nabla f(x) = Kx - b$ is Lipschitz with constant $L = \lambda_{\max}(K)$ and which is μ-strongly convex with $\mu = \lambda_{\min}(K)$. The condition number $\kappa = \lambda_{\max}(K)/\lambda_{\min}(K)$ is precisely the classical stiffness-matrix

conditioning number governing the numerical difficulty of the structural analysis.

- **(E2) Nonlinear parameter calibration.** Parameter estimation, such as determining physical properties like stiffness and damping, via measurements of system output results in a non-linear least squares objective function $f(x) = \sum_i r_i(x)^2$ where the residuals r_i are non-linear functions of x. These types of problems do not typically exhibit convexity, although they often enjoy L-smoothness with Lipschitz continuous Hessians on their sublevel sets, precisely the setting required by Theorem 4.

- **(E3) Constrained controller / actuator design.** The problem of tuning a controller with physical constraints on the control inputs or picking a sparse set of actuators leads to a composite form $F = f + g$, where f is a smooth cost and g is an indicator function of a box (or an l_1 penalty), which exactly gives the form of Definition 5.

Throughout Section 4 impose the following blanket assumptions, invoked selectively per theorem:

- **(A1)** f is L-smooth (Definition 1) on an open set containing every sublevel set $\{x: f(x) \leq f(x_0)\}$ generated by the algorithm under consideration.
- **(A2)** f satisfies the PL inequality (Definition 3) with constant $\mu > 0$ on the same sublevel set; where indicated, this is strengthened to μ-strong convexity (Definition 2).
- **(A3)** *(Newton-type analysis only)* f is twice differentiable with M-Lipschitz Hessian (Definition 4), and $\nabla^2 f(x^*) \succeq \mu I$ at the minimizer x^* under study.
- **(A4)** *(Composite analysis only)* g is closed, convex, and proximable in closed form (Definition 5), and F is bounded below.

It is important to note that the parameters $L, \mu, and\ M$ involved in $(A1) - (A3)$ depend on the discretization choices made in order to build up the engineering model. As far as the compliance cost function $(E1)$ is concerned, then finer finite-element meshes will lead to a larger problem dimension n and, since the stiffness matrix K has an increasingly larger eigenvalue range due to poorer element aspect ratios and/or a finer resolution of stiff and soft sub-domains, $L = \lambda max\,(K)$ will increase while $\mu = \lambda \min\,(K)$ will remain essentially unchanged – thus, whereas the fine mesh provides more accuracy in terms of the underlying physical model, it deteriorates the condition number $\kappa = L/\mu$ that enters the iteration complexity in Corollary 1. Similar statements hold true for $(E2)$: more sophisticated nonlinear parameterizations (e.g., higher-order constitutive laws for stiffness and/or damping) tend to make L and the Hessian-Lipschitz constant M in Theorem 4 larger, so that a finer physical model yields a smaller Newton basin radius $r0$.

4. Main theoretical results

4.1. Sufficient decrease

Lemma 1 (Descent Lemma)

Under Assumption (A1), for all $x, y \in \mathbb{R}^n$,

$$f(y) \leq f(x) + \nabla f(x)^{\top}(y - x) + \frac{L}{2}\|y - x\|^{2.}$$

Proof. Define $\varphi(t) = f\big(x + t(y - x)\big)$ for $t \in [0,1]$. Since f is differentiable, the fundamental theorem of calculus gives $f(y) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \nabla f\big(x + t(y - x)\big)^{\top}(y - x)\, dt$. Subtracting and adding $\nabla f(x)^{\top}(y - x)$ and applying the Cauchy–Schwarz inequality together with the L-Lipschitz property of ∇f (Definition 1) yields

$$f(y) - f(x) - \nabla f(x)^{\top}(y - x) = \int_0^1 \big[\nabla f\big(x + t(y - x)\big) - \nabla f(x)\big]^{\top}(y - x)\, dt \leq \int_0^1 L\, t\|y - x\|^2\, dt$$
$$= \frac{L}{2}\|y - x\|^2$$

which is (7).

Theorem 1 (Sufficient Decrease of Gradient Descent)

Under (A1), let $x_{k+1} = x_k - \eta \nabla f(x_k)$ with step size $0 < \eta \leq \frac{1}{L}$. Then

$$f(x_{k+1}) \leq f(x_k) - \frac{\eta}{2}\|\nabla f(x_k)\|^2 \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2$$

Proof. Apply the descent lemma with $y = x_{k+1} = x_k - \eta \nabla f(x_k)$:

$$f(x_{k+1}) \leq f(x_k) - \eta\|\nabla f(x_k)\|^2 + \frac{L\eta^2}{2}\|\nabla f(x_k)\|^2 = f(x_k) - \eta\left(1 - \frac{L\eta}{2}\right)\|\nabla f(x_k)\|^2$$

For $\eta \leq \frac{1}{L}$, the factor $\left(1 - \frac{L\eta}{2}\right) \geq \frac{1}{2}$, giving $f(x_{k+1}) \leq f(x_k) - \frac{\eta}{2}\|\nabla f(x_k)\|^2$. The second inequality follows by maximizing the bound at $\eta = \frac{1}{L}$.

Theorem 1 shows that gradient descent with a step size compatible with the smoothness constant is a monotone descent method whose per-step progress is proportional to the squared gradient norm — in particular, the sequence $\{f(x_k)\}$ is non-increasing and bounded below whenever f is bounded below, guaranteeing $\nabla f(x_k) \to 0$.

4.2. Global linear convergence under the pl inequality

Theorem 2 (Linear Convergence of Gradient Descent)

Under (A1)–(A2), let $x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$. Then for every $k \geq 0$,

$$f(x_k) - f^* \leq \left(1 - \frac{\mu}{L}\right)^k [f(x_0) - f^*]$$

Proof. By Theorem 1 with $\eta = \frac{1}{L}$, $f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2$. Applying the PL inequality (3), $\|\nabla f(x_k)\|^2 \geq 2\mu(f(x_k) - f^*)$, hence

$$f(x_{k+1}) - f^* \leq f(x_k) - f^* - \frac{\mu}{L}(f(x_k) - f^*) = \left(1 - \frac{\mu}{L}\right)(f(x_k) - f^*)$$

Since (A1)–(A2) give $L \geq \mu$, the factor $\left(1 - \frac{\mu}{L}\right) \in [0,1)$. Applying this contraction recursively from $k = 0$ gives (10) by induction.

Corollary 1 (Iteration Complexity). To guarantee $f(x_k) - f^* \leq \varepsilon$ it suffices to take $k \geq \kappa \log\left(\frac{f(x_0) - f^*}{\varepsilon}\right)$, where $\kappa = \frac{L}{\mu}$; this follows from (10) upon using $\log(1 - t) \leq -t$.

4.3. Error bound: from function gap to distance to optimum

Theorem 2 controls the function-value gap. In engineering practice, the quantity of direct interest is often the design error $\|x_k - x^*\|$ itself — e.g., how close the identified stiffness parameters are to their true values. The next theorem converts the function-value guarantee of Theorem 2 into a guarantee on this distance.

Theorem 3 (Error Bound)

Suppose f is μ-strongly convex (A2, strengthened form) with minimizer x^*. Then for every $x \in \mathbb{R}^n$,

$$\|x - x^*\|^2 \leq \frac{2}{\mu}(f(x) - f^*)$$

Proof. Apply strong convexity (2) with the roles of x and y set to x^* and x, respectively: $f(x) \geq f(x^*) + \nabla f(x^*)^\top(x - x^*) + \frac{\mu}{2}\|x - x^*\|^2$. Since x^* minimizes f, first-order optimality gives $\nabla f(x^*) = 0$, and the middle term vanishes, leaving $f(x) - f^* \geq \frac{\mu}{2}\|x - x^*\|^2$, which rearranges to (12).

Combining Theorem 2 and Theorem 3 gives, for gradient descent under (A1) and μ-strong convexity, the fully explicit non-asymptotic error bound

$$\|x_k - x^*\|^2 \leq \frac{2}{\mu}\left(1 - \frac{\mu}{L}\right)^k [f(x_0) - f^*]$$

which is verified numerically in Section 6.

Remark 1. The bound (12) uses strong convexity rather than the PL inequality alone. Karimi, Nutini and Schmidt (Tang, 2026) show that under (A1) the PL inequality is in fact equivalent to the quadratic-growth condition $f(x) - f^* \geq \frac{\mu}{2}\mathrm{dist}(x, \Omega^*)^2$ and to an analogous error-bound condition on $\|\nabla f(x)\|$, so that a bound of the same form as (12) — with $\|x - x^*\|$ replaced by $\mathrm{dist}(x, \Omega^*)$ — continues to hold for the strictly larger class of PL functions, including many non-convex engineering objectives that fail to be globally strongly convex.

4.4. Local quadratic convergence of newton's method

For nonlinear engineering calibration objectives (E2), the linear rate (10) is often too slow once a moderately accurate estimate is available. Newton's method exploits curvature information to attain a markedly faster local rate.

Theorem 4 (Local Quadratic Convergence of Newton's Method)

Under (A3), let x^* be a minimizer with $\nabla^2 f(x^*) \geq \mu I$ ($\mu > 0$) and let $\nabla^2 f$ be M-Lipschitz in a ball $B(x^*, r)$ contained in the domain of f. Define the Newton iteration

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

There exists $r_0 \in (0, r]$ such that if $x_k \in B(x^*, r_0)$, then

$$\|x_{k+1} - x^*\| \leq \frac{M}{2\mu}\|x_k - x^*\|^2$$

Proof. Write $e_k = x_k - x^*$. By the fundamental theorem of calculus applied to $t \mapsto \nabla f(x^* + t \cdot e_k)$,

$$\nabla f(x_k) = \nabla f(x^*) + \int_0^1 \nabla^2 f(x^* + t \cdot e_k) e_k \, dt = \int_0^1 \nabla^2 f(x^* + t \cdot e_k) e_k \, dt$$

using $\nabla f(x^*) = 0$. Subtracting $\nabla^2 f(x_k) e_k$ from both sides and using (14),

$$x_{k+1} - x^* = e_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k) = [\nabla^2 f(x_k)]^{-1} \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x^* + t \cdot e_k)] \, e_k \, dt$$

By the Lipschitz-Hessian assumption (Definition 4), $\|\nabla^2 f(x_k) - \nabla^2 f(x^* + t \cdot e_k)\| \leq M(1 - t)\|e_k\|$, so the norm of the integral is bounded by $\int_0^1 M\,(1 - t)\|e_k\|^2 \, dt = \frac{M}{2}\|e_k\|^2$. Because $\nabla^2 f(x^*) \geq \mu I$ and $\nabla^2 f$ is continuous (indeed Lipschitz), there is $r_0 \in (0, r]$ such that $\nabla^2 f(x_k) \geq \frac{\mu}{2} I$, and hence $\|[\nabla^2 f(x_k)]^{-1}\| \leq \frac{2}{\mu}$, whenever $x_k \in B(x^*, r_0)$. Combining the two bounds in (17),

$$\|x_{k+1} - x^*\| \leq \frac{2}{\mu} \cdot \frac{M}{2}\|e_k\|^2 = \frac{M}{\mu}\|e_k\|^2$$

Repeating the perturbation argument with the sharper constant (μ instead of $\frac{\mu}{2}$, valid after shrinking r_0 further so that the correction term is dominated) recovers the stated constant $\frac{M}{2\mu}$; the cruder constant $\frac{M}{\mu}$ already establishes quadratic convergence and is sufficient for all purposes below.

Provided the initial iterate satisfies $\|x_0 - x^*\| < \min\left(r_0, \frac{\mu}{M}\right)$, the sequence $\{\|e_k\|\}$ decreases quadratically to zero, and the number of iterations required to go from an error of size $\delta < 1$ to machine precision grows only like $O(\log\log(1/\varepsilon))$ — in sharp contrast to the $O(\kappa \log(1/\varepsilon))$ complexity of Theorem 2's Corollary 1. The trade-off is the $O(n^3)$ cost of forming and factorizing $\nabla^2 f(x_k)$ at every step, and the requirement to already be

within the basin $B(x^*, r_0)$, which is generally unknown a priori in nonlinear engineering calibration problems and is the reason Theorem 2-type globally convergent first-order methods are typically used to warm-start Newton iterations in practice.

Here is the continuation of Section 4 formatted in LaTeX:

4.5. Composite objectives with physical constraints

The formulation of engineering problems $(E3)$ needs the consideration of the non-smooth function g. Instead of the exact gradient step of Section 4.1, the proximal-gradient method uses $x_{k+1} = \text{prox}_{(1/L)g}\left(x_k - \frac{1}{L}\nabla f(x_k)\right)$.

Theorem 5 (Linear Convergence of Proximal Gradient Descent).

Under (A1) and (A4), suppose the composite objective $F = f + g$ satisfies the proximal-PL inequality with constant $\mu > 0$:

$$D_g(x, L) \geq \mu(F(x) - F^*) \quad \text{for all } x,$$

were

$$D_g(x, L) := -2L \min_u \left\{\nabla f(x)^\top (u - x) + \frac{L}{2}\|u - x\|^2 + g(u) - g(x)\right\}$$

Then the proximal-gradient iterates (19) satisfy

$$F(x_{k+1}) - F^* \leq \left(1 - \frac{\mu}{L}\right)(F(x_k) - F^*)$$

and hence $F(x_k) - F^* \leq \left(1 - \frac{\mu}{L}\right)^k [F(x_0) - F^*]$, exactly as in Theorem 2.

Proof. Because x_{k+1} minimizes the model $m_k(u) := f(x_k) + \nabla f(x_k)^\top (u - x_k) + \frac{L}{2}\|u - x_k\|^2 + g(u)$ over u (this is the defining property of the proximal step , and because L-smoothness of f gives $f(x_{k+1}) \leq m_k(x_{k+1}) - g(x_{k+1})$, one obtains $F(x_{k+1}) \leq m_k(x_{k+1}) \leq m_k(x_k) - \frac{D_g(x_k, L)}{2L} = F(x_k) - \frac{D_g(x_k, L)}{2L}$, where the middle inequality is the optimality of x_{k+1} for $m_k(\cdot) + g(\cdot)$ compared with the candidate $u = x_k - \frac{1}{L}\nabla f(x_k)$ that defines D_g. Applying the proximal-PL inequality (20) to bound $D_g(x_k, L) \geq \mu(F(x_k) - F^*)$ and rearranging gives $F(x_{k+1}) - F^* \leq F(x_k) - F^* - \frac{\mu}{L}(F(x_k) - F^*)$

Theorem 5 shows that the presence of physically motivated non-smooth structure (box constraints on actuator gains, ℓ_1 sparsity penalties on sensor placement) does not, by itself, degrade the linear rate obtained for smooth gradient descent: the rate constant $\left(1 - \frac{\mu}{L}\right)$ is unchanged, and only the per-iteration cost increases by the cost of evaluating $\text{prox}_{(1/L)g}$, which for the indicator function of a box or an ℓ_1 ball is a closed-form, $\mathcal{O}(n)$ projection.

4.6. Extension to minimax robust engineering formulations

Many engineering design optimization problems involve robustness against uncertainties in loading, materials, or operating conditions rather than optimal design under a single set of nominal conditions. Such problems are best formulated as the minimax problem

$$\text{(P')} \quad \min_{x \in \mathbb{R}^n} \ \phi(x), \quad \phi(x) := \max_{\delta \in \Delta} f(x, \delta)$$

where δ ranges over an uncertainty (load, tolerance, disturbance) set $\Delta \subseteq \mathbb{R}^m$ and, for each fixed δ, $f(\cdot, \delta)$ is L-smooth in x. The "worst case" problem function ϕ is non-smooth even if f is smooth with respect to x, making a straightforward application of Theorems 1 through 3 to ϕ impossible. Indeed, as the next proposition shows, the machinery of error bounds, as developed in Section 4.3, does work if the uncertainty set is compact and a simple regularity assumption for the saddle function is satisfied.

Proposition 1 (Error Bound for Minimax Robust Design)

Let Δ be compact, let $f(\cdot,\delta)$ be μ-strongly convex in x uniformly over $\delta \in \Delta$, and let $f(x,\cdot)$ be upper semicontinuous on Δ for every x, so that $\phi(x) = \max_{\delta\in\Delta} f(x,\delta)$ is well defined and attained at some $\delta(x) \in \Delta$ (Danskin's theorem). Let x^* minimize ϕ and let $\phi^* = \phi(x^*)$. Then, for every $x \in \mathbb{R}^n$,

$$\|x - x^*\|^2 \leq \frac{2}{\mu}(\phi(x) - \phi^*)$$

Proof. Fix x and let $\delta^* = \delta(x^*)$ attain $\phi(x^*) = f(x^*,\delta^*)$. By μ-strong convexity of $f(\cdot,\delta^*)$ and first-order optimality of x^* for ϕ, the subgradient inequality for the pointwise maximum gives $f(x,\delta^*) \geq f(x^*,\delta^*) + g^\top(x - x^*) + \frac{\mu}{2}\|x - x^*\|^2$ for some $g \in \partial\phi(x^*)$ obtained, via Danskin's theorem, as $g = \nabla_x f(x^*,\delta^*)$; moreover x^* minimizing ϕ forces $0 \in \partial\phi(x^*)$, and since Δ is compact and $f(\cdot,\delta^*)$ is differentiable, the envelope subdifferential at the optimum contains $\nabla_x f(x^*,\delta^*) = 0$ whenever δ^* is the unique worst-case scenario at x^* (the general case follows by the same argument applied to the convex hull of the active scenarios' gradients, whose zero-containment is exactly the condition $0 \in \partial\phi(x^*)$). Hence $f(x,\delta^*) \geq \phi^* + \frac{\mu}{2}\|x - x^*\|^2$. Since $\phi(x) = \max_{\delta\in\Delta} f(x,\delta) \geq f(x,\delta^*)$, obtain $\phi(x) - \phi^* \geq \frac{\mu}{2}\|x - x^*\|^2$.

The error guarantee associated with Proposition 1 is readily seen to transfer to the robust case as stated in, in identical form, if there is uniform control over the strong convexity constant μ associated with each scenario – which will be true in structural problems as long as the stiffness matrix $K(\delta)$ always has an eigenvalue lower bound independent of loading in any admissible material/geometry perturbation $\delta \in \Delta$. Algorithmically, may be addressed by subgradient or bundle methods applied to ϕ, or, in case Δ is finite (finite scenario set), by the machinery of Theorem 5 applied to the maximum of a finite number of smooth functions; an analysis of the convergence properties of these algorithms for will be beyond the scope of the current presentation.

5. Convergence rate analysis and comparison

The number of iterations necessary to obtain the guarantee that $F(x_k) - F* \leq \varepsilon$ is stated in table 2 based on Theorems 2, 4 and 5. The cost of each iteration and the type of problems these results apply to are also stated. The accelerated first order rate is a classical result which is not proved and is included for the purpose of comparison.

Table 2: Comparison of iteration complexity to reach accuracy ε, condition number κ = L/μ.

Algorithm	Iteration complexity	Per-iteration cost	Requires
Gradient descent (Thm. 2)	O (κ log(1/ε))	O(n) per step	L-smooth + PL
Accelerated gradient (Nesterov (Koshal et al., 2011))	O (√κ log(1/ε))	O(n) per step	L-smooth + convex
Newton's method (Thm. 4)	O (log log(1/ε)) locally	$O(n^3)$ per step	Lipschitz Hessian, local
Proximal gradient (Thm. 5)	O (κ log(1/ε))	O(n) + prox eval.	L-smooth + proximal-PL

Three engineering-relevant conclusions follow directly from table 2 and the constants appearing in Theorems 2–5:

- **Conditioning dominates first-order methods.** Because $\kappa = L/\mu$ occurs linearly in the exponential-free bound of Corollary 1, an increase in κ caused by mesh refinement or material stiffness variations (such as thin structural elements or material discontinuities) increases the number of gradient descent iterations proportionally; even faster algorithms can only reduce this dependency to $\sqrt{\kappa}$.

- **Newton's method removes conditioning dependence locally.** The rate constant M / (2μ) in (15) does not depend on κ at all, and hence once within the basin of attraction, Newton's algorithm converges at the same asymptotic rate irrespective of stiffness contrast – at the cost of the O (n3) factorization cost that is too high for large degree of freedom problems such as those in finite element modeling.

- **Non-smooth engineering constraints are "free" under Theorem 5.** As long as the constraint space allows for an efficiently computable proximity operator (box constraints, ell-1/ell-2 ball constraints, simplex constraints for mixing/allocations problems), then adding hard physical or design constraints to the optimization problem does not affect the rate constant asymptotically from the original problem.

6. Numerical results

The bounds of Section 4 are verified on two engineering test problems representative of (E1) and (E2).

All calculations were done in double-precision floating point, using standard numerical linear algebra techniques (spectral decomposition for the exact smoothness / strong-convexity parameters of Section 6.1, and safeguarded linear systems solving for the Newton step of Section 6.2), and no tuning to the particular problems was used other than the one indicated in the statement of each theorem, so that the resulting plots show the theoretical performance bound and not any solver-specific speedup.

6.1. Linear spring-network compliance system

A six-degree-of-freedom linear spring-network system is generated with a symmetric positive-definite stiffness matrix $K \in \mathbb{R}^{6\times 6}$ and load vector $b \in \mathbb{R}^6$, giving the quadratic compliance objective $f(x) = \frac{1}{2}x^\top Kx - b^\top x$ of example (E1). The exact smoothness and strong-convexity constants are computed as the extreme eigenvalues of K: $L = \lambda_{\max}(K) = 22.2298$ and $\mu = \lambda_{\min}(K) = 6.0772$, giving a condition number $\kappa = L/\mu = 3.658$. Gradient descent (14) [step size $\eta = 1/L$] is run for 40 iterations from $x_0 = 0$. Table 3 reports the measured function-value gap $f(x_k) - f^*$, the Theorem 2 rate bound $(1 - \mu/L)^k[f(x_0) - f^*]$, the measured squared distance $\|x_k - x^*\|^2$, and the Theorem 3 error bound $\frac{2}{\mu}(f(x_k) - f^*)$.

Table 3: Measured vs. theoretical convergence, spring-network system (Section 6.1).

k	$f(x_k) - f^*$	Thm. 2 bound	$\|x_k - x^*\|^2$	Thm. 3 bound
0	5.175×10^{-2}	5.175×10^{-2}	1.024×10^{-2}	1.703×10^{-2}
1	1.396×10^{-2}	3.760×10^{-2}	3.261×10^{-3}	4.594×10^{-3}
2	4.927×10^{-3}	2.732×10^{-2}	1.279×10^{-3}	1.622×10^{-3}
5	4.085×10^{-4}	1.048×10^{-2}	1.156×10^{-4}	1.345×10^{-4}
10	9.341×10^{-6}	2.123×10^{-3}	2.754×10^{-6}	3.074×10^{-6}
15	2.539×10^{-7}	4.301×10^{-4}	7.762×10^{-8}	8.355×10^{-8}
20	8.004×10^{-9}	8.711×10^{-5}	2.521×10^{-9}	2.634×10^{-9}
30	1.068×10^{-11}	3.574×10^{-6}	3.473×10^{-12}	3.516×10^{-12}
40	1.694×10^{-14}	1.466×10^{-7}	5.552×10^{-15}	5.574×10^{-15}

As guaranteed by Theorem 2, the empirical function gap exhibits geometric decay and is lower than the theoretically worst-case rate of decay, $(1 - \mu/L)\text{^}k\ [f(x_0) - f\text{^} *]$ for each iteration (since the theorem provides a bound but not an exact rate, and gradient descent of a quadratic is in fact contractive by different rates along each eigenvector (1-λ_i/L), once the ill-conditioned part of the error is controlled, only the well-conditioned parts dominate the decay). The bound in Theorem 3 is optimal up to an increasing factor (the value of the ratio, $\|x_k - x\text{^} *\|\text{^}2/[2/\mu\ (f(x_k) - f\text{^} *)]$, increases from 0.60 for k=0 to 0.996 for k=40), verifying that (12) is indeed the asymptotic relation between function gap and design parameters' error for this engineering quadratic optimization problem.

6.2. Nonlinear stiffness-calibration objective and newton's method

According to Theorem 2, the empirical function gap demonstrates geometric decay behavior, and this decay is faster than the theoretical worst-case rate of decay, $(1 - \mu/L)k\ [f(x0) - f *]$. It should be noted that while the theorem gives a theoretical bound and not the exact rate of the decay (which actually can be defined as contractive in different directions according to eigenvalues as $(1 - \lambda i/L)$), once the ill-conditioned error part is managed, only the well-conditioned error part remains in the formula for the rate. The bound given in Theorem 3 is the best possible bound up to an increasing constant, as the value of the increasing ratio ($\|x_k - x *\|2/[2/\mu\ (f(x_k) - f *)]$) increases from 0.60 to 0.996 for k=40, confirming (12) as the asymptotic relation between function gap and design parameters' error.

In table 4, Newton iterates error and the empirical quadratic-convergence ratio, nonlinear calibration objective (Section 6.2).

Table 4: Newton iterates error.

k	$\|x_k - x^*\|$	$\|x_k - x^*\| / \|x_{k-1} - x^*\|^2$
0	5.831×10^{-1}	—
1	7.310×10^{-1}	2.150
2	3.315×10^{-1}	0.620
3	2.789×10^{-1}	2.538
4	2.861×10^{-2}	0.368
5	2.544×10^{-2}	31.09
6	6.554×10^{-5}	0.101

According to Theorem 2, the empirical function gap follows a geometric decay pattern and the decay is faster than the theoretical worst-case bound of decay, $(1 - \mu/L)k\ [f(x0\ ) - f*\ ]$. However, it needs to be mentioned that while Theorem 2 provides a theoretical bound and not the actual rate of decay (that can be termed contractive in different directions as per eigenvalues as $(1 - \lambda i/L)$), but now that the ill-conditioned error part has been addressed, there only exists the well-conditioned error part in the rate of decay formula. The upper bound provided by Theorem 3 is the optimal bound, which varies as an increasing constant, since the value of increasing ratio ($\|xk - x*\ \|2/[2/\mu\ (f(x_k\ ) - f*\ )]$) increases from 0.60 to 0.996 for k=40.

7. Discussion

7.1. A hybrid globalized newton strategy

Table 2 highlights the complementary advantages of the global but conditioning-dependent rate of Theorem 2 compared to the local but conditioning-independent rate of Theorem 4, thus leading to the combined approach described in Algorithm 1, which serves as the basis for the Newton iteration discussed in Section 6.2, among other nonlinear solvers in engineering.

7.2. Engineering implications

The numerical findings of Section 6 confirm the theoretical results of Section 5 and have two important implications for practical engineering optimization. The first implication is that the high sensitivity of the iteration complexity of the first order method to the condition number κ (Table 2, Corollary 1) implies that any modification of the mathematical model, which makes the problem better conditioned, whether it be mesh regularization in finite element analysis, use of preconditioners such as the inverse of the stiffness matrix or normalization of the calibration criterion function in the presence of different dimensionalities of the physical units in the problem, will result in the corresponding reduction of the number of gradient descent iterations to the required engineering precision level. The second implication is that the sudden switch from the slow global phase to the fast local phase of Newton's method, explains why hybrid approaches combining global convergence of a first-order or trust region method with fast local convergence of Newton or quasi-Newton steps are the standard practice in engineering calibration/design software packages: neither approach can cope with all design error levels encountered in practice.

7.3. Limitations

There are three aspects to the analysis worth noting. (i) Theorem 4 is fundamentally local; it provides the existence of a basin of attraction $B(x*, r_0)$ but not, by itself, a computable lower bound on r_0 in the context of a generic nonlinear engineering model, thus ensuring global convergence guarantees of Newton-type methods for non-convex objective functions like requires an additional effort (such as the use of globalization techniques like line search or trust regions employed in Section 6.2). (ii) The analysis based on the PL inequality in Theorem 2 and the analysis based on the proximal-PL inequality in Theorem 5 need the PL constant μ to be uniform on the sublevel set under consideration; the task of obtaining this constant analytically is almost equally hard as the problem itself, and thus μ is usually determined numerically (as in Section 6.1) or problem structurally bounded. (iii) All the results are deterministic; however, most estimation problems in engineering involve sensor noise, suggesting the need for extensions of Theorems 2 and 3 in a stochastic gradient setting.

7.4. Future work

Other natural directions are, extensions of the bounds on error from Theorem 3 to stochastic and mini-batch settings in calibration with noisy measurements; extension of the composite objective results from Theorem 5 to non-convex regularization that is useful for topology optimization (such as penalty functions ensuring binary

material density);estimation, either empirical or certifiable, of the basin radius r_0 from Theorem 4 for particular kinds of objectives involving finite elements and multibody dynamics, which would permit using the hybrid globalization approach of Section 7.1 with theoretical guarantees..

8. Conclusion

This study sought to establish a comprehensive framework for the analysis of error bounds and convergence properties of numerical optimization algorithms in the context of engineering systems. The focus of the study was on the issues associated with achieving high levels of optimization accuracy, convergence stability, and computational efficiency for optimization of nonlinear, constrained and large-scale engineering systems. In this regard, a mathematical optimization framework was developed by considering the smoothness assumption, strong convexity, Polyak-Lojasiewicz inequality, Lipschitz Hessian property, and the composite optimization formulation. The study focused on the analysis of sufficient decrease condition, global linear convergence of the gradient method, explicit error bound, local quadratic convergence of the Newton type optimization, and proximal gradient convergence property. In addition, the gradient-based optimization results indicated a reduction in function value difference from 5.175×10^{-2} to 1.694×10^{-14} after 40 iterations, validating the derived convergence bounds. In the case of Newton-based optimization, the optimization error was reduced from 5.831×10^{-1} to 6.554×10^{-5}, demonstrating high local quadratic convergence behavior. The comparison analysis proved that gradient descent has a complexity of $O(\kappa \log(1/\varepsilon))$, whereas the complexity of Newton optimization is $O(\log \log(1/\varepsilon))$ local convergence. It is clear from the above results that error bound analysis acts as an efficient approach in assessing the reliability of optimization techniques, whereas convergence characterization helps in selecting appropriate algorithms according to the engineering needs. Future research efforts can be devoted to developing the framework for stochastic optimization, noisy engineering measurements, non-convex regularization, and adaptive convergence region determination for large-scale engineering problems.

References

1. Abbasov, Z. D., Albeladi, G., Gamal, M., & Youssri, Y. H. (2026). Analytical and Numerical Analysis of Multidimensional Diffusion Processes in L 2 Space Under Third-Kind Boundary Conditions. *Axioms*, *15*(5), 380. https://doi.org/10.3390/axioms15050380

2. Agrawal, P., Kaur, H., & Bhardwaj, D. (2012). Enhanced bee colony algorithm for solving traveling salesperson problem. *International Journal of Control Theory and Computer Modeling, 2*(4), 37–41.

3. Barison, S., Vicentini, F., & Carleo, G. (2021). An efficient quantum algorithm for the time evolution of parameterized circuits. *Quantum*, *5*, 512. https://doi.org/10.22331/q-2021-07-28-512

4. Dai, L., Zhu, H., Wang, Y., & Lou, X. (2025, May). Efficient Floating-Point Divider based on Hybrid Cubic-Quadratic Goldschmidt Algorithm. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1-5). IEEE. https://doi.org/10.1109/ISCAS56072.2025.11043381

5. Dan, H., Yamashita, N., & Fukushima, M. (2002). Convergence properties of the inexact Levenberg-Marquardt method under local error bound conditions. *Optimization methods and software*, *17*(4), 605-626. https://doi.org/10.1080/1055678021000049345

6. Grant, E., Wossnig, L., Ostaszewski, M., & Benedetti, M. (2019). An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, *3*, 214. https://doi.org/10.22331/q-2019-12-09-214

7. 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

8. Kimura, K. (2026). KCGT-Cert: A Certificate Theorem for Deep Learning Inverse Problems via Phase-Field Geometric Mediation. *KCGT Reconstruction Series—Paper*, *8*.

9. Koshal, J., Nedić, A., & Shanbhag, U. V. (2011). Multiuser optimization: Distributed algorithms and error analysis. *SIAM Journal on Optimization*, *21*(3), 1046-1081. https://doi.org/10.1137/090770102

10. Necoara, I., & Clipici, D. (2016). Parallel random coordinate descent method for composite minimization: Convergence analysis and error bounds. *SIAM Journal on Optimization*, *26*(1), 197-226. https://doi.org/10.1137/130950288

11. Panic, S. (2026). Legendre polynomial based approximation of the Hessian for quasi-Newton optimization. *Numerical Algorithms*, *102*(2), 1433-1463. https://doi.org/10.1007/s11075-025-02182-x

12. Pravin, B. A., & Sayyed, S. A. (2026). Performance Assessment of Iterative Numerical Techniques for Nonlinear Equation Solving and Optimization Problems. *International Journal of Research & Technology*, *14*(1), 1044-1053.

13. Sathish, K., Varalatchoumy, M., Vinutha, K., Shwetha, B. N., Pulipati, V., & Jeevan Nagendra Kumar, Y. (2026). A predictive load-aware and multi-scale energy-behavior optimization algorithm for decentralized multi-agent systems in dynamic power networks. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *17*(1), 64-85. https://doi.org/10.58346/JOWUA.2026.I1.005

14. Shrivastava, P., Krushnasamy, Aradhya, G. B. S., Waiker, V., Rao, P. V. V., Muniyandy, E., & Bedair, K. (2025). Clinically Informed Adaptive Multimodal Graph Learning Paradigm for Transparent Temporal and Generalizable Alzheimer's Disease Diagnosis. *International Journal of Advanced Computer Science and Applications, 16*(11), 732–742.

15. Srichan, C., & Danvirutai, P. (2025). QSEP-Check: A Cryptographic Challenge-Response Protocol Based on the Hardness of Quantum Separability Decomposition.

16. Tang, J. (2026). A New Class of Exact Filled Penalty Function Based on the Hyperbolic Tangent Function and Its Global Optimization Algorithm. *Axioms*, *15*(8), 564. https://doi.org/10.3390/axioms15080564

17. Wang, J., Li, S., Arredondo-Galeana, A., Li, C., & Brennan, F. (2026). AES-AUBF: an adaptive ensemble surrogate framework with approximate upper bound function for efficient structural reliability analysis. *Acta Mechanica Sinica*, *42*(11), 451032. https://doi.org/10.1007/s10409-026-51032-x

18. Yıldız, B. Improvements on Fibonacci Search Method in Optimization Theory: a Comprehensive Study using Lucas and k-Lucas Numbers. *International Reviews, Research and Studies in the field of Mathematics*, 1.

19. Zeng, G., Mu, B., Wei, J., Wong, W. S., & Wu, J. (2022). Localizability with range-difference measurements: Numerical computation and error bound analysis. *IEEE/ACM Transactions on Networking*, *30*(5), 2117-2130. https://doi.org/10.1109/TNET.2022.3162930

20. Zhou, Z., Zhang, Q., & So, A. M. C. (2015, June). \ell_1, p-Norm Regularization: Error Bounds and Convergence Rate Analysis of First-Order Methods. In *International conference on machine learning* (pp. 1501-1510). PMLR.