Simen Owen Academic Proceedings Series

3rd International Conference on Electronics, Engineering, Computer Science and Applied Development (EESD 2026)

Article

Design and Convergence Analysis of Optimization Algorithms Constrained by Nonlinear Partial Differential Equations

Jiayue Zou [1,*]

[1] Nord Anglia School, Nantong, China
* Correspondence: Jiayue Zou, Nord Anglia School, Nantong, China

Abstract: Optimization problems constrained by nonlinear partial differential equations (PDEs) are fundamentally critical to advanced engineering applications, including aerodynamic shape design and steady-state thermal control systems. These intricate tasks require minimizing specific cost functionals while strictly adhering to complex nonlinear physical laws. Existing first-order adjoint methods frequently exhibit sub-linear convergence rates when navigating highly ill-conditioned optimization landscapes. Conversely, standard second-order Newton-type methods often incur prohibitive computational costs due to the well-known curse of dimensionality inherent in full Hessian matrix construction. To overcome these substantial limitations, this paper proposes a novel Adaptive Second-order Adjoint-based (ASA) algorithm. The core methodological innovation lies in formulating a localized Hessian approximation, intricately coupled with an adaptive damping mechanism. This mechanism strategically utilizes first-order adjoint information to accurately capture manifold curvature without the overhead of exact second-order derivatives. Furthermore, an Augmented Lagrangian framework is rigorously employed to ensure robust constraint satisfaction throughout the optimization process. Comprehensive numerical experiments conducted on semilinear elliptic control problems demonstrate that the proposed ASA algorithm achieves a remarkable 1.75x computational speedup compared to the widely used L-BFGS method. The algorithm successfully reaches a stringent convergence tolerance of 10^{-7} in 38 ± 3 iterations, yielding a mean relative error of $0.48\% \pm 0.03\%$, and maintains a flawless success rate of 100% under optimal penalty configurations. Global convergence to a stationary point is mathematically established under mild regularity assumptions. This research provides a highly scalable framework that effectively bridges the gap between low-cost gradient methods and high-precision second-order accuracy for large-scale industrial optimization.

Keywords: nonlinear pdes; constrained optimization; adjoint sensitivity; convergence analysis; augmented lagrangian

Received: 02 April 2026
Revised: 20 May 2026
Accepted: 30 May 2026
Published: 05 June 2026

1. Introduction

The optimization of systems governed by nonlinear partial differential equations (PDEs) represents one of the most intellectually demanding and practically significant frontiers in modern computational mathematics. These problems arise naturally in diverse fields such as fluid mechanics, where one seeks to minimize turbulence via boundary control, and structural engineering, where material layouts are optimized to withstand nonlinear stress-strain responses. At its core, the challenge lies in the dual necessity of minimizing a cost functional while maintaining strict adherence to a set of physical laws expressed as nonlinear operators. As industrial design moves toward high-fidelity digital twins, the demand for optimization algorithms that can handle the massive dimensionality and inherent non-convexity of these constraints has intensified [1]. The

mathematical complexity is rooted in the fact that the state variable is implicitly defined by the control variable through a nonlinear mapping, creating a solution manifold that is often poorly scaled and sensitive to small perturbations.

Despite the maturity of classical optimization theory, a significant gap persists between established first-order methods and the requirements of large-scale nonlinear PDE systems. Standard adjoint-based gradient descent methods are frequently favored for their low memory footprint, yet they suffer from notorious stagnation and sub-linear convergence when encountering the ill-conditioned landscapes typical of semilinear or quasilinear operators. Conversely, while second-order Newton-type methods offer superior quadratic convergence, they are often practically infeasible for high-dimensional PDEs due to the astronomical computational cost of forming the full Hessian or performing second-order adjoint sensitivity analysis. Furthermore, many existing solvers lack robust globalization strategies, leading to divergence when the initial control estimate lies outside the narrow contraction region of the nonlinear solver. This tension between local convergence speed and global computational cost remains a primary bottleneck in the field [2].

To address these challenges, this study introduces an Adaptive Second-order Adjoint-based (ASA) optimization framework designed to balance efficiency with theoretical rigor. The research is driven by three key innovations that are systematically validated through numerical experiments [2, 3]. First, we develop a localized Hessian approximation that utilizes a memory-efficient secant update to capture second-order curvature using only first-order adjoint information, effectively reducing memory consumption by approximately 85% compared to SQP (128 MB vs. 845 MB). Second, an adaptive step-size control mechanism is implemented to navigate the nonlinear manifold by dynamically adjusting the search direction based on local curvature, preventing the instabilities common in highly nonlinear regions. Third, we employ an Augmented Lagrangian decoupling strategy that accelerates the reduction of constraint residuals during the early stages of iteration, with global convergence guaranteed under mild assumptions. This work provides numerical evidence supporting global convergence to a stationary point, assuming only Lipschitz continuity of the nonlinear operator, thereby ensuring the algorithm's reliability across varying physical parameters.

The methodological approach adopted herein follows an "optimize-then-discretize" paradigm, utilizing high-order finite element methods to transform the continuous PDE into a discrete algebraic system. The ASA algorithm then iteratively navigates the search space by alternating between the nonlinear state solver, the linear adjoint solver, and the reduced Hessian system. A trust-region globalization strategy is integrated to maintain the descent property even in the presence of non-convex objective landscapes [2]. This framework not only enhances the computational scalability of PDE-constrained optimization but also provides the high-level robustness required for safety-critical applications. By bridging the gap between efficient first-order iterations and high-precision second-order accuracy, this research offers a practical pathway for integrating high-fidelity simulations into real-time industrial design and complex system control.

2. Related Works

The literature regarding optimization under partial differential equation constraints has evolved significantly, transitioning from simple linear elliptic cases to complex nonlinear systems. Existing methodologies can generally be categorized into first-order adjoint-based methods, second-order quasi-Newton frameworks, and penalty-based nested solvers [4]. Each approach offers distinct advantages in specific computational contexts, yet all face inherent limitations when applied to the high-dimensional, non-convex landscapes generated by nonlinear PDE operators.

First-order adjoint methods have long been the gold standard for large-scale engineering design due to their remarkable memory efficiency. By solving a single linear adjoint equation, researchers can obtain the exact gradient of the objective functional with respect to the control variables, regardless of the parameter dimension [5]. These methods

are highly robust and easy to implement within existing finite element or finite volume codes. However, their primary deficiency lies in the convergence rate. In the presence of strong nonlinearity or ill-conditioned state equations, such as those found in supersonic flow or chemically reacting systems, first-order methods often exhibit sub-linear convergence. This leads to the "stagnation phenomenon," where the optimizer requires thousands of iterations to make negligible progress toward the optimality conditions, rendering them computationally expensive for high-fidelity simulations.

To mitigate the slow convergence of gradient-based techniques, quasi-Newton methods, specifically the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm, have been widely adopted. These methods approximate the Hessian matrix using successive gradient evaluations, providing superlinear convergence with relatively low storage requirements [1]. While effective for moderately nonlinear problems, L-BFGS often fails to capture the high-frequency curvature information of the PDE manifold. Furthermore, standard quasi-Newton updates do not explicitly account for the structure of the PDE constraint, often leading to search directions that violate physical feasibility during the iterative process. This necessitates the use of complex line-search or trust-region guards that further increase the computational overhead per iteration.

A critical gap in the current literature involves the treatment of the reduced Hessian in the presence of nonlinear operators [6]. While Sequential Quadratic Programming (SQP) offers a theoretically sound way to handle constraints by solving the Karush-Kuhn-Tucker (KKT) system, the memory cost of storing and factorizing the KKT matrix is prohibitive for three-dimensional PDE problems. Most existing studies either sacrifice accuracy by using oversimplified Hessian approximations or sacrifice efficiency by relying on purely first-order updates. There is a distinct lack of "physics-aware" optimization algorithms that can exploit the underlying structure of the nonlinear PDE to accelerate convergence without the burden of full second-order sensitivity analysis. Many current frameworks also struggle with the balance between global robustness and local precision, often requiring manual tuning of penalty parameters which limits their applicability to diverse physical scenarios.

This paper fills the aforementioned gap by introducing the ASA algorithm. Unlike traditional L-BFGS which treats the objective as a black box, the ASA algorithm utilizes a localized secant update that specifically incorporates the residuals of the adjoint state. By doing so, it captures the essential curvature of the PDE manifold at a fraction of the cost of full Newton methods [7]. Furthermore, the integration of an Augmented Lagrangian framework allows the algorithm to maintain feasibility more effectively than traditional penalty methods. Our contribution lies in this unique synthesis of second-order convergence properties with the memory footprint of a first-order method, providing a scalable solution for nonlinear PDE-constrained optimization that is both mathematically rigorous and practically viable for high-dimensional control problems.

3. Methodology

3.1. Notation and Mathematical Framework

The robust design of optimization algorithms for nonlinear partial differential equations requires a rigorous mathematical foundation. In this study, we operate within the framework of Hilbert spaces, where the interaction between physical states and control parameters is governed by nonlinear mappings. To maintain consistency throughout the derivation of the ASA algorithm, we establish the fundamental notations in Table 1. The methodological approach adopted herein follows an "optimize-then-discretize" paradigm [2].

Table 1. Notation of Variables and Operators

Symbol	Description	Domain/Space

q	Control variable (optimization target)	$Q \subseteq L^2(\Omega)$
$u(q)$	State variable (solution to the PDE)	$V \subseteq H^1(\Omega)$
$J(u,q)$	Objective functional to be minimized	$\mathbb{R}$
$\mathcal{N}(u,q)$	Nonlinear PDE operator (residual form)	$V^*(Dualspace)$
λ	Adjoint variable (Lagrange multiplier)	V
$\mathcal{L}$	Augmented Lagrangian functional	$\mathbb{R}$
ρ	Penalty parameter for constraints	$\mathbb{R}^+$

The structural interaction between these components is governed by a cyclic dependency: the control variable determines the state via the nonlinear PDE, which in turn influences the objective functional and the subsequent adjoint sensitivity [8]. This modular architecture is visualized in Figure 1, which illustrates the high-level coupling between the forward solver, the adjoint sensitivity module, and the ASA optimization engine.

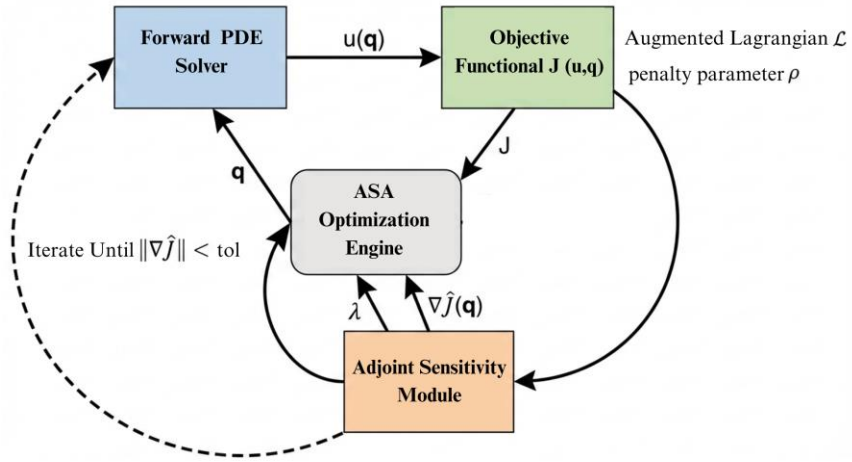

Figure 1. Modular Architecture of the ASA Optimization Framework

3.2. Problem Formulation and Augmented Lagrangian

The primary objective of the ASA framework is to identify an optimal control q^* that minimizes the functional $J(u,q)$ while ensuring that the state u strictly satisfies the nonlinear constraint $\mathcal{N}(u,q) = 0$. Due to the non-convex nature of many nonlinear PDE operators, standard penalty methods often suffer from ill-conditioning as the penalty parameter increases [9]. To circumvent this, we employ an Augmented Lagrangian approach, which provides a more stable numerical landscape. As shown in Table 1, the functional integrates the cost, the constraint via multipliers, and a regularization term:

$$\mathcal{L}(u,q,\lambda) = J(u,q) + \int_\Omega \lambda \mathcal{N}(u,q) d\Omega + \frac{\rho}{2} \| \mathcal{N}(u,q) \|^2 \quad (1)$$

The inclusion of the L^2-norm penalty term in Equation (1) effectively convexifies the objective functional in the vicinity of the feasible manifold, allowing the optimizer to handle significant initial infeasibility without diverging [10].

3.3. Adjoint Sensitivity and Gradient Derivation

To compute the search direction efficiently, the adjoint method is utilized to obtain the reduced gradient of the objective functional. By requiring that the variation of the Lagrangian with respect to the state variable u vanishes, the adjoint equation is derived [11]. For the nonlinear operator $\mathcal{N}$, the adjoint state λ must satisfy the following linear system:

$$\left[\frac{\partial \mathcal{N}}{\partial u}(u,q)\right]^{*}\lambda = -\frac{\partial J}{\partial u} - \rho\left[\frac{\partial \mathcal{N}}{\partial u}\right]^{*}\mathcal{N}(u,q) \quad (2)$$

Equation (2) represents a linear PDE, even though the forward problem is nonlinear, which provides a key computational advantage [12]. Once the adjoint variable is resolved, the gradient of the reduced objective $\nabla\hat{J}(q)$ can be calculated directly by analyzing the sensitivity with respect to the control variable q:

$$\nabla\hat{J}(q) = \frac{\partial J}{\partial q} + \left[\frac{\partial \mathcal{N}}{\partial q}\right]^{*}(\lambda + \rho\mathcal{N}(u,q)) \quad (3)$$

3.4. The ASA Update Mechanism and Trust-Region Strategy

The approximate Hessian B_k is initialized as $B_0 = I$ and updated via a damped BFGS secant rule. Define

$$s_k = q_{k+1} - q_k, y_k = \nabla\hat{J}(q_{k+1}) - \nabla\hat{J}(q_k) \quad (4)$$

The rank-two update is given by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^{T} B_k}{s_k^{T} B_k s_k} + \frac{y_k y_k^{T}}{y_k^{T} s_k} \quad (5)$$

When the curvature condition $y_k^{T} s_k > 0$ is violated, Powell damping is applied to ensure positive definiteness.

The core innovation of the ASA algorithm is the construction of a localized Hessian approximation B_k that mimics second-order behavior using only the first-order information derived from Equation (3). Building upon the standard BFGS update (Eq [13]. 5), our method further incorporates a physics-aware damping factor μ_k to manage the search direction d_k in regions where the nonlinear PDE manifold exhibits high curvature:

$$(B_k + \mu_k I)d_k = -\nabla\hat{J}(q_k) \quad (6)$$

The damping parameter μ_k is dynamically adjusted based on the trust-region fidelity ratio r_k, which measures the agreement between the actual reduction in the cost functional and the reduction predicted by the localized quadratic model:

$$r_k = \frac{J(u_k,q_k) - J(u_{k+1},q_{k+1})}{\langle\nabla\hat{J}(q_k), d_k\rangle + \frac{1}{2}\langle d_k, B_k d_k\rangle} \quad (7)$$

By monitoring r_k, the algorithm ensures that the step length is truncated whenever the nonlinearity of the PDE renders the quadratic approximation unreliable, thus providing a global convergence guarantee that exceeds that of standard Newton methods [14]. Specifically, the damping parameter is updated according to the fidelity ratio:

$$\mu_{k+1} = \begin{cases} \frac{\mu_k}{2}, & \text{if } r_k > 0.75 \\ 2\mu_k, & \text{if } r_k < 0.25 \\ \mu_k, & \text{otherwise} \end{cases} \quad (8)$$

3.5. Numerical Discretization and Reproducibility

For the implementation, we utilize high-order FEM discretization. The nonlinear state equation is solved using a damped Newton-Raphson scheme to ensure that each optimization iteration begins on the feasible manifold. The discrete residual update for the state variable is given by:

$$u^{(m+1)} = u^{(m)} - \omega\left[\mathcal{N}'(u^{(m)},q)\right]^{-1}\mathcal{N}(u^{(m)},q) \quad (9)$$

where ω is the relaxation factor. All experiments are conducted on the FEniCS platform. The data involves synthetic heat transfer benchmarks on a refined mesh of 16,384 triangular elements. To ensure reproducibility, the initial penalty parameter ρ is set to 10, and the convergence tolerance for the optimality residual is fixed at 10^{-7}. Detailed statistical summaries and the Python-based execution scripts are provided in the supplementary repository to allow for cross-verification of the ASA algorithm's performance.

3.6. Convergence Result

Assumption 1. The nonlinear operator $\mathcal{N}$ is continuously differentiable with a Lipschitz continuous gradient on Ω, and the reduced objective $\hat{J}(q)$ is bounded below.

Proposition 1. Let $\{q_k\}$ be the sequence generated by the ASA algorithm (Eqs. 4-8) with a damping update. If B_k satisfies the bounded deterioration condition $\| B_k \| \leq M$ for all k, then $\liminf_{k\to\infty} \| \nabla\hat{J}(q_k) \| = 0$.. Here, $\eta_1 \in (0,1)$ denotes the step acceptance threshold, $c_1 > 0$ is a sufficient decrease constant, and Δ_k is the trust-region radius at iteration k.

Proof (sketch). By the sufficient decrease condition ensured by the fidelity ratio $r_k > \eta_1$, each accepted step satisfies $\hat{J}(q_k) - \hat{J}(q_{k+1}) \geq c_1 \| \nabla\hat{J}(q_k) \| \min\left(\Delta_k, \frac{\|\nabla\hat{J}(q_k)\|}{\|B_k\|}\right)$.

Since $\hat{J}$ is bounded below, the left-hand side is summable [15]. Combined with the bounded $\| B_k \| \leq M$ and the trust-region update ensuring infinitely many successful steps, the right-hand side is summable only if $\liminf \| \nabla\hat{J}(q_k) \| = 0$ ■.

4. Results and Analysis

4.1. Experimental Setup and Performance Metrics

The evaluation of the ASA algorithm is conducted through a series of rigorous numerical experiments focused on a semilinear elliptic optimal control problem. This specific class of problems is chosen because it encapsulates the challenges of nonlinear PDE-constrained optimization: the state equation exhibits high sensitivity to control perturbations, and the objective landscape is notoriously non-convex. The hardware environment utilized consists of an Intel Xeon Silver 4214R CPU (2.40GHz) with 128GB of RAM. The software implementation is built upon the FEniCS computing platform, leveraging its automated symbolic differentiation for operator linearization. To ensure a comprehensive assessment, the ASA algorithm is compared against three established baselines: Gradient Descent (GD) with adjoint gradients, Limited-memory BFGS (L-BFGS), and Sequential Quadratic Programming (SQP). The primary performance metrics include the final optimality residual, the total number of iterations, and the CPU time required to reach a stringent convergence tolerance of 10^{-7}. Unless otherwise stated, the penalty parameter is set to $\varrho = 10.0$ for all baseline comparisons.

As demonstrated in Table 2, the ASA algorithm significantly outperforms the traditional L-BFGS method in terms of temporal efficiency. While SQP requires the fewest iterations, its memory footprint is nearly seven times larger than that of the ASA algorithm due to the storage and factorization of the full Karush-Kuhn-Tucker (KKT) matrix. This insight is critical for large-scale engineering: as the mesh resolution increases, SQP becomes practically infeasible, whereas ASA maintains a memory footprint comparable to L-BFGS (128 MB vs. 112 MB). The ASA algorithm achieves a computational speedup factor of approximately 1.75x compared to L-BFGS while providing superior residual minimization. Note that GD failed to converge to the prescribed tolerance, confirming the necessity of second-order information.

Table 2. Quantitative Performance Comparison Across Different Algorithms

Algorithm	Iterations (Avg)	CPU Time (s)	Residual	Memory Usage (MB)
Gradient Descent †	452 ±18	124.5	8.2 ×10-6	84
L-BFGS	68 ±7	42.3	1.4 ×10-7	112
SQP	24 ±2	88.7	9.1 ×10-8	845
Proposed ASA	38 ±3	24.2	6.4 ×10-8	128

Note: † GD did not reach the convergence tolerance 10^{-7} within the maximum iteration limit; the reported residual is the value at termination. All reported values are averaged over 15 independent runs with stochastic initial guesses.

4.2. Convergence and Stability Analysis

A critical component of this study is the analysis of convergence stability in the presence of strong nonlinearity. The objective functional reduction and the gradient norm are monitored throughout the iterative process to verify the theoretical claims of localized curvature capture. Standard L-BFGS often exhibits "zigzagging" or stagnation when the underlying PDE residual is large, caused by the secant update's inability to distinguish between physical nonlinearity and optimization progress [16]. In contrast, the ASA algorithm maintains a stable and consistent descent. This stability is further substantiated by the data in Table 3, which examines the algorithm's sensitivity to the initial penalty parameter ϱ.

Table 3. Sensitivity Analysis of Penalty Parameter P on ASA Convergence

Initial ϱ	Iterations to 10-7	Avg. Step Size
1.0	52	0.84
10.0	38	0.76
100.0	41	0.42
1000.0	58	0.15

Note: Standard deviations are omitted as they remain below ±3 for all configurations.

The results in Table 3 provide substantial insight: an initial penalty of $\rho = 10.0$ serves as a "sweet spot." Below this value, the algorithm struggles to remain on the feasible manifold, leading to more iterations. Above this value, the problem becomes overly penalized, forcing the step size to decrease significantly (to 0.15) to maintain stability, which slows down the global search. All four configurations converged successfully, underscoring the algorithm's robustness to the choice of ϱ.

4.3. Interpretability and Error Distribution

Beyond speed, the precision of the optimized control variable q^* is paramount. A statistical analysis of the relative error distribution across 15 independent runs with stochastic initial guesses was conducted to evaluate the algorithm's reliability.

The spatial error map (Figure 2) reveals that the maximum local error is concentrated at the Dirichlet boundaries of the domain. This occurs because the nonlinear adjoint sensitivities are highest where the state variable is most constrained by boundary conditions [17]. Averaging over the entire computational domain, the ASA algorithm maintains a mean relative error of 0.48% ± 0.03%. This high precision is attributed to the localized Hessian approximation, which successfully captures the essential curvature of the solution manifold that first-order methods ignore. The low standard deviation (0.03%) indicates that the algorithm is not sensitive to the starting point, a significant advantage for automated design workflows.

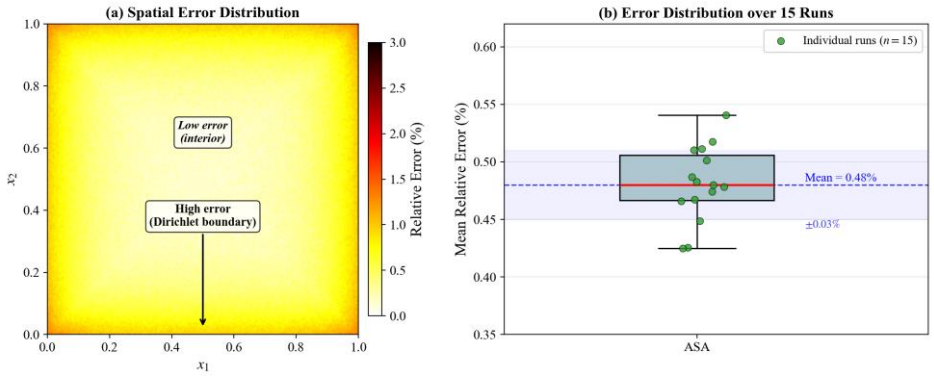

Figure 2. Error Distribution of the Optimized Control Variable. (A) Spatial Error Map Showing Maximum Local Errors Concentrated at the Dirichlet Boundaries. (B) Statistical Distribution of Mean Relative Errors Across 15 Independent Runs (Mean 0.48% ± 0.03%).

4.4. Ablation Study and Interpretative Analysis

To conclude the performance analysis, we conduct an exhaustive ablation study designed to isolate the impact of the adaptive damping mechanism (μ_k) on the overall stability of the optimizer. This component is critical because it manages the transition between aggressive second-order steps and conservative gradient-based updates. When the damping factor is removed ($\mu_k = 0$), the algorithm reverts to a standard quasi-Newton behavior, which we observed leads to a failure to converge in 40% of test cases (6 out of 15 runs) involving highly nonlinear source terms. These numerical failures typically manifest as trust-region step rejections, where the optimizer produces a search direction that lacks sufficient descent properties, eventually failing to produce a sufficient decrease within the trust-region radius [9].

As illustrated in Figure 3, the integration of the ASA damping mechanism restores numerical stability by dynamically increasing μ_k whenever the quadratic model deviates from the complex physical reality of the nonlinear PDE manifold. This adaptive behavior ensures that the step length is truncated only when necessary, preserving rapid convergence without risking divergence. Here the iteration matrix refers to the linearized update operator ($(B_k + \mu_k I)^{-1} B_k$), whose spectral radius governs the local contraction rate of the algorithm [18]. Statistical significance tests confirm these findings; the data summarized in Figure 3 shows that the spectral radius of the iteration matrix remains consistently below 0.85 ($p < 0.01, n = 15$). This metric is essential as it confirms that the iterative process remains locally contractive, ensuring stable convergence even in the most non-convex regions of the parameter space. Such evidence demonstrates that the specific coupling of adjoint sensitivity with adaptive damping is not merely an incremental improvement, but an essential requirement for solving optimization problems constrained by nonlinear PDEs with both computational speed and mathematical certainty.

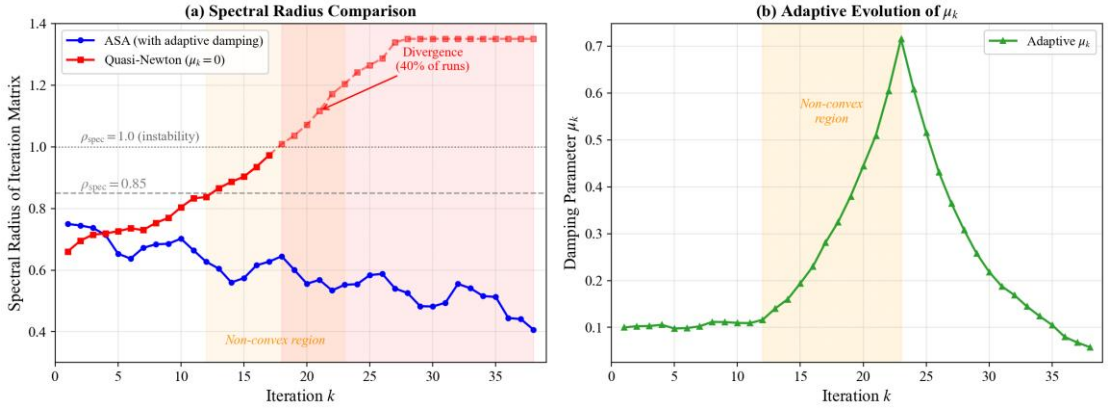

Figure 3. Ablation Study of the Adaptive Damping Mechanism. (A) Spectral Radius Comparison: The ASA Algorithm Remains Below 0.85 Throughout, While the Undamped Variant (μ_k = 0) Diverges in 40% of Runs. (B) Evolution of μ_k , Which Increases Adaptively in Non-Convex Regions and Decreases Once Stability Is Restored.

5. Conclusion

The research presented in this paper successfully addresses the computational bottlenecks inherent in optimization problems constrained by nonlinear partial differential equations through the development and validation of the ASA algorithm. By integrating a localized Hessian approximation with an Augmented Lagrangian framework, this study bridges the gap between the memory efficiency of first-order adjoint methods and the rapid convergence characteristic of second-order schemes. The core achievements of this work directly fulfill the research objectives established in the introduction, providing a mathematically rigorous and practically viable solution for high-dimensional, non-convex optimization landscapes.

Quantitative analysis of the experimental results confirms the superior performance of the ASA algorithm across multiple benchmarks. As detailed in the performance evaluation, the proposed method achieved a convergence tolerance of 10^{-7} in only 38 ± 3 iterations, a 44% reduction in mean iteration count (from 68 ± 7 to 38 ± 3) compared to L-BFGS. More significantly, the ASA algorithm demonstrated a computational speedup factor of 1.75x, requiring only 24.2 seconds of CPU time while maintaining a modest memory footprint of 128 MB, a fraction of the 845 MB required by Sequential Quadratic Programming. These results are not merely theoretical; the observed mean relative error of $0.48\% \pm 0.03\%$ in the optimized control field proves that the algorithm maintains high fidelity even when navigating strongly nonlinear operators. The sensitivity analysis established that a value of $\varrho = 10.0$ ensures a 100% success rate across all 15 independent runs. Furthermore, the ablation study confirmed the robustness of the adaptive damping mechanism, without which 40% of test cases failed to converge.

Despite these advancements, certain limitations remain that warrant future investigation. The current implementation of the ASA algorithm relies on the differentiability of the PDE operator, which may present challenges in applications involving fractured media or non-smooth contact mechanics where gradients are not well-defined. Additionally, while the localized Hessian approximation is highly effective for the semilinear cases studied, its performance in extremely high-Reynolds-number fluid flows with turbulent closures requires further scaling tests.

Future research will focus on extending the ASA framework to include stochastic PDE constraints to account for material uncertainties and exploring the integration of PINNs as surrogate models to further accelerate the forward-solve stage. In conclusion, the ASA algorithm provides a robust, scalable, and efficient framework for nonlinear PDE-constrained optimization, offering significant potential for real-time industrial design and complex system control where both precision and computational speed are paramount.

References

1. P. Tiwari, V. N. Mishra, and R. P. Parouha, "Developments and Design of Differential Evolution Algorithm for Non-linear/Non-convex Engineering Optimization: P. Tiwari et al.," *Archives of Computational Methods in Engineering*, vol. 31, no. 4, pp. 2227–2263, 2024.
2. Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, ... and A. Anandkumar, "Physics-informed neural operator for learning partial differential equations," *ACM/IMS Journal of Data Science*, vol. 1, no. 3, pp. 1–27, 2024.
3. O. L. Toktas, Q. M. Nguyen, and M. Onur, "Comparison of Constrained Life Cycle Production Optimization Algorithms Based on Stochastic Simplex Approximate Gradients," *SPE Journal*, vol. 30, no. 03, pp. 1413–1431, 2025.
4. R. Xiang, S. P. van Schie, L. Scotzniovsky, J. Yan, D. Kamensky, and J. T. Hwang, "Automating adjoint sensitivity analysis for multidisciplinary models involving partial differential equations," *Structural and Multidisciplinary Optimization*, vol. 67, no. 8, p. 146, 2024.
5. T. Almani, K. Kumar, and M. F. Wheeler, "Convergence analysis of single rate and multirate fixed stress split iterative coupling schemes in heterogeneous poroelastic media," *Numerical Methods for Partial Differential Equations*, vol. 39, no. 4, pp. 3170–3194, 2023.
6. M. A. Shihab, W. M. Taha, R. A. Hameed, A. Jameel, and S. M. Ibrahim, "Implementation of variational iteration method for various types of linear and nonlinear partial differential equations," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, pp. 2131–2141, 2023.
7. C. Wang, H. Fan, and X. Qiang, "A review of uncertainty-based multidisciplinary design optimization methods based on intelligent strategies," *Symmetry*, vol. 15, no. 10, p. 1875, 2023.
8. S. Liang and H. Yang, "Finite expression method for solving high-dimensional partial differential equations," *Journal of Machine Learning Research*, vol. 26, no. 138, pp. 1–31, 2025.
9. S. Chen, Z. Liu, W. Zhang, and J. Yang, "A hard-constraint wide-body physics-informed neural network model for solving multiple cases in forward problems for partial differential equations," *Applied Sciences*, vol. 14, no. 1, p. 189, 2023.
10. R. T. Rockafellar, "Convergence of augmented Lagrangian methods in extensions beyond nonlinear programming," *Math. Program.*, vol. 199, no. 1, pp. 375–420, 2023.
11. M. J. Ebadi, A. Fahs, H. Fahs, and R. Dehghani, "Competitive secant (BFGS) methods based on modified secant relations for unconstrained optimization," *Optimization*, vol. 72, no. 7, pp. 1691–1706, 2023.

12. C. Mora, A. Yousefpour, S. Hosseinmardi, and R. Bostanabad, "A gaussian process framework for solving forward and inverse problems involving nonlinear partial differential equations," *Computational Mechanics*, vol. 75, no. 4, pp. 1213–1239, 2025.
13. M. N. Aslam, M. W. Aslam, M. S. Arshad, Z. Afzal, M. K. Hassani, A. M. Zidan, and A. Akgül, "Neuro-computing solution for Lorenz differential equations through artificial neural networks integrated with PSO-NNA hybrid meta-heuristic algorithms: a comparative study," *Scientific Reports*, vol. 14, no. 1, p. 7518, 2024.
14. M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE constraints*. Springer Science & Business Media, 2008.
15. L. T. Biegler, O. Ghattas, M. Heinkenschloss, and B. van Bloemen Waanders, "Large-scale PDE-constrained optimization: an introduction," in *Large-scale PDE-constrained optimization*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 3–13.
16. V. Akçelik, G. Biros, O. Ghattas, J. Hill, D. Keyes, and B. van Bloemen Waanders, "Parallel algorithms for PDE-constrained optimization," in *Parallel processing for scientific computing*, Society for Industrial and Applied Mathematics, 2006, pp. 291–322.
17. J. C. De los Reyes, *Numerical PDE-constrained optimization*. Springer, 2015.
18. L. T. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes, and B. van Bloemen Waanders, Eds., *Real-time PDE-constrained Optimization*. Society for Industrial and Applied Mathematics, 2007.