

Hybrid Fractional-Order Optimization in High-Dimensional Engineering Systems: A Rigorous Convergence-Guaranteed Mathematical Framework

J. Devagnanam¹, Gandhikota Umamahesh², Ms.Swati Meshram³, Ritesh Kumar Kushwaha⁴, Chandrashekara A C⁵, Rasha Ahmed Hamid Ahmed⁶

¹Assistant Professor, Department of computer Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur-603203, Tamil Nadu, devagnaj@srmist.edu.in

²Assistant professor, Department of CSE, Aditya University, Surampalem,

³Designation Assistant Professor, P.G.Department of Computer Science, SNDT Women's University, Mumbai, Maharashtra

⁴Department of Electronics and Communication Engineering, Chandigarh University, Mohali, India

⁵associate Professor, Department Of Mathematics, Government First Grade College, Chamarajanagara, Karnataka State -571313, Chamarajanagara University

⁶Lecturer, Department of Mathematics, University college of AdDair, Jazan University, Jazan, KSA

Abstract:

Classical integer-order gradient methods update each iterate using only the instantaneous local gradient, discarding the trajectory's history and, in high-dimensional, multimodal engineering optimization landscapes, this local-memoryless behavior frequently produces premature convergence to poor local optima. Fractional-order calculus, particularly the Caputo derivative, offers a mathematically principled mechanism for injecting long-range memory into the optimization trajectory by replacing the integer-order derivative with a non-local, history-weighted convolution operator. This paper reviews the convergence theory and hybrid algorithmic designs that have emerged around fractional-order optimization for high-dimensional engineering systems, synthesizing formal convergence-rate results for Caputo fractional gradient descent (CFGD) under convex, strongly convex, and non-convex smoothness assumptions with the metaheuristic hybridization literature that embeds fractional-order memory into population-based optimizers such as differential evolution and the Young's Double-Slit Experimental optimizer. Particular attention is given to the mathematical mechanism connecting the fractional order α to convergence behavior: theoretical and empirical evidence converges on the finding that fractional orders above unity accelerate convergence while orders below unity improve steady-state accuracy, with a domain-dependent optimum typically located near $\alpha \approx 0.65$ for highly multimodal engineering benchmarks. Comparative tables map the reviewed convergence theorems to their underlying smoothness assumptions and guarantee type, cross-reference hybrid fractional-metaheuristic algorithms against the specific engineering optimization problem each was validated on, and set reported statistical performance, convergence-speed improvement, error reduction, and significance testing, against the baseline algorithms each hybrid method outperformed. The review concludes that while rigorous convergence guarantees for fractional gradient methods are now well established for smooth convex and non-convex settings, extending these formal guarantees to the population-based hybrid metaheuristics that currently dominate high-dimensional engineering practice remains the field's principal open mathematical problem.

Keywords: Fractional-Order Calculus, Caputo Derivative, Convergence Analysis, High-Dimensional Optimization, Hybrid Metaheuristics, Gradient Descent, Engineering Optimization, Non-Convex Optimization.

I. Introduction

High-dimensional engineering optimization problems, truss topology design, photovoltaic parameter identification, dynamic resource allocation, and constrained structural optimization, share a common mathematical pathology: their objective landscapes are frequently non-convex, highly multimodal, and populated with numerous local optima that trap integer-order gradient-based methods before they reach a globally competitive solution [1]. Classical gradient descent updates the iterate using only the first-order derivative evaluated at the current point, a memoryless local rule whose convergence behavior is well characterized under Lipschitz-smoothness assumptions but which, precisely because it discards trajectory history, has no mechanism for distinguishing a genuine local minimum from a shallow saddle region in a high-dimensional, multimodal landscape [9]. The classical descent-lemma argument establishes that if the gradient ∇f is Lipschitz continuous with constant L , gradient descent with a fixed step size $t \leq 1/L$ is guaranteed to converge, with convergence rate $O(1/k)$ for convex objectives and finite-iteration convergence to a stationary point under additional regularity conditions [10], [11]. These guarantees, however, are guarantees of convergence to a stationary point, not to a global or even a high-quality local optimum, a distinction that becomes practically decisive in the high-dimensional engineering settings this paper addresses.

Fractional-order calculus offers a mathematically principled alternative. Where the integer-order derivative is inherently local, the Caputo fractional derivative is defined as a convolution of the function's derivative history against a power-law memory kernel, giving the resulting fractional gradient direction an explicit dependence on the optimization trajectory's past rather than solely its present state [4]. A foundational treatment of Caputo fractional gradient descent (CFGD) established that each search direction generated by the generic CFGD update is the steepest descent direction of a local smoothing of the original objective function, providing the first rigorous variational interpretation of why fractional-order memory alters, and in many empirical settings improves, the geometry the optimizer actually descends [4]. Building on this foundation, a systematic convergence analysis proved convergence results for fractional gradient descent across smooth and strongly convex, smooth and convex, and smooth and non-convex settings, establishing novel inequalities connecting fractional derivatives to integer derivatives specifically because properties like smoothness, convexity, and strong convexity are conventionally expressed in terms of first derivatives and are otherwise not directly applicable to a fractional-order operator [2].

A parallel and largely independent research program has explored fractional-order memory not within gradient-based methods but within population-based metaheuristics, motivated by the same high-dimensional, multimodal engineering problems. A recent Fractional-Order Differential Evolution (FCDE) framework introduced Caputo fractional derivatives directly into the mutation and update equations of standard Differential Evolution, arguing that most existing metaheuristic methods use integer-order update dynamics, which do not take into account the memory-dependent and non-local properties that are typically found in real-world systems, a simplification that makes it harder for them to adapt their searches, stay converged, and escape local optima, especially in high-dimensional or constrained contexts [3]. Tested directly against truss topology design, photovoltaic maximum-power-point-tracking control, and dynamic resource allocation, three canonical high-dimensional engineering optimization problems, FCDE achieved a mean error reduction of up to 38% on highly multimodal functions relative to ten state-of-the-art baseline algorithms, with statistical significance confirmed at $p < 0.01$ in Wilcoxon signed-rank tests [3].

This paper reviews the gradient-based convergence theory for fractional-order optimization alongside the hybrid metaheuristic literature that applies fractional-order memory to high-dimensional engineering problems, within a single comparative framework organized around the mathematical mechanism, guarantee type, and empirically validated engineering application of each reviewed method.

Research Objectives

- To review the formal convergence-rate theory established for Caputo fractional gradient descent under convex, strongly convex, and non-convex smoothness assumptions.
- To synthesize the hybrid fractional-metaheuristic literature applying fractional-order memory to high-dimensional, constrained engineering optimization problems.
- To examine the mathematical relationship between the fractional order parameter α and observed convergence speed and solution accuracy.
- To evaluate the scalability and numerical-stability challenges fractional-order methods face as problem dimensionality grows into the hundreds or thousands of variables.
- To identify the open problem of extending rigorous convergence guarantees from gradient-based fractional methods to the population-based hybrid metaheuristics that currently dominate empirical engineering practice.

II. Literature Review

2.1 Mathematical Foundations of Fractional-Order Gradient Descent

The Caputo fractional derivative is the derivative definition most commonly adopted in fractional-order optimization because, unlike the Riemann-Liouville definition, it permits the use of conventional integer-order initial conditions, a property that has "nice analytic properties" specifically suited to initial-value optimization problems [2]. The foundational Caputo Fractional Gradient Descent (CFGD) algorithm defines a generic update requiring three tunable parameters, and its central theoretical result establishes that each direction the algorithm generates is the steepest descent direction of a local smoothing of the original objective function, providing a variational, rather than purely heuristic, justification for the algorithm's behavior [4]. This local-smoothing interpretation is significant because it explains, within a single unifying mathematical picture, why fractional-order methods empirically exhibit greater robustness to shallow local minima than integer-order gradient descent: the smoothing implicit in the fractional operator effectively regularizes the optimization landscape before the descent step is taken.

Building rigorously on this foundation, a systematic convergence analysis established novel inequalities connecting fractional derivatives to integer derivatives, necessary because standard smoothness, convexity, and strong-convexity assumptions are conventionally stated in terms of the gradient (first derivative), and without a formal bridge between the two derivative notions, applying these assumptions to a fractional-order method would be mathematically unjustified [2]. Using these bridging inequalities, the analysis proved convergence rates for a fractional gradient descent method structurally related to the Adaptive Terminal Caputo Fractional Gradient Descent (AT-CFGD) method, with a specific linear convergence-rate theorem established for the single-dimensional case across a range of hyperparameter domains [2]. The proof technique for the non-convex smooth case proceeds by bounding the objective decrease at each step using the L-smoothness descent inequality applied to the fractional search direction, yielding a per-step decrease bound of the form:

$$f(x_{t+1}) - f(x_t) \leq (-\eta_t(1 - \lambda_t K_1 - |\lambda_t| K_2) + (L\eta_t^2/2)(|1 - \lambda_t K_1| + |\lambda_t| K_2)^2) \|\nabla f(x_t)\|^2$$

where η_t is the step size, K_1 and K_2 are constants bounding the deviation of the fractional direction from the true gradient, and λ_t is a fractional-order weighting parameter, a bound that reduces exactly to the classical integer-order descent lemma when the fractional-deviation terms K_1 , K_2 vanish [2]. This structural reduction is mathematically important: it confirms that the fractional-order framework strictly generalizes, rather than replaces, classical gradient-descent convergence theory, so any convergence guarantee proved for the fractional method degrades gracefully to the well-established integer-order guarantee in the appropriate limit.

2.2 Adaptive Stepsize and Dimension-Aware Extensions

A separate line of research has extended fractional gradient methods specifically to address high-dimensional function optimization, the central concern of this paper. The quasi fractional order gradient descent method (QFOGDM) extended the fractional order gradient method to high-dimensional function optimization problems specifically, introducing an adaptive stepsize mechanism and establishing convergence through three separate theorems rather than a single unified proof, reflecting the added complexity adaptive step-size selection introduces into the convergence argument [5]. Numerical validation across fifteen unconstrained optimization benchmarks found that QFOGDM achieved better convergence rate and accuracy relative to fixed-stepsizes alternatives, with the method further applied successfully to parameter identification in linear discrete deterministic systems [5]. A related generalization of the gradient method using a fractional-order gradient direction (FHGD) established a specific, practically important design principle governing the relationship between the fractional order and convergence behavior: the fractional order greater than 1 provides faster convergence speed, less than 1 gives better steady state performance, and equal to 1 reduces the FHGD to standard heavy-ball gradient descent, formally characterizing the fractional order as a tunable interpolation parameter between convergence-speed and steady-state-accuracy objectives [6].

The Effective Dimension Aware Fractional-Order Stochastic Gradient Descent (2SED-FOSGD) framework addressed the specific challenge of applying fractional-order memory in high-dimensional stochastic settings, where the fixed-exponent fractional order used in earlier deterministic CFGD formulations does not account for the effective dimensionality of the underlying optimization landscape [7]. Under standard smoothness and bounded-gradient assumptions, the framework established rigorous convergence guarantees for the proposed method, with the geometry-based regularization achieving faster and more stable convergence compared to naive fixed-exponent fractional stochastic gradient descent, and experiments across diverse benchmarks demonstrating consistent improvements in both test accuracy and training loss relative to the fixed-exponent baseline [7]. This dimension-aware adaptation is directly relevant to high-dimensional engineering optimization because it addresses a specific failure mode of naive fractional methods: a fixed fractional order tuned for low-dimensional problems does not automatically transfer favorably to the hundred- or thousand-dimensional landscapes characteristic of large-scale engineering design problems.

2.3 Scalable Fractional Gradient Methods for Large-Scale Systems

A recognized practical limitation of Caputo fractional gradient descent is that the fractional derivative's convolution structure requires, in principle, storing and reweighting the entire gradient history at every iteration, an $O(t)$ per-step memory and computation cost that grows unfavorably as the number of iterations t increases, a serious obstacle for the large-scale engineering systems this paper addresses. A recent scalable reformulation addressed this bottleneck directly, reinterpreting Caputo-based fractional gradient descent as a nonlocal convolution operator acting on the gradient history and introducing a dyadic hierarchical decomposition and compression (DHDC) scheme that approximates the full convolution using a compressed, logarithmic-memory representation [8]. The associated convergence theorem established that under standard L -smoothness assumptions, the resulting method still exhibits monotone descent and linear convergence, provided that the approximation error introduced by memory compression remains controlled, with the error itself formally shown to be governed by two factors: the variation of the Caputo kernel within each compressed bin and the accumulated gradient magnitude over that bin [8]. Empirical validation across battery state estimation, stochastic signal reconstruction, and image classification found the compressed method consistently demonstrates improved stability compared to classical gradient descent and better robustness than truncated fractional schemes, while significantly reducing computational and memory costs, directly addressing the scalability requirement for high-dimensional engineering deployment [8].

2.4 Hybrid Fractional-Order Metaheuristics for Engineering Optimization

While the gradient-based fractional methods reviewed in Sections 2.1–2.3 provide the strongest available formal convergence guarantees, population-based metaheuristics remain the dominant practical tool for the non-differentiable, discontinuous, or black-box objective functions common in engineering design, and a substantial

literature has accordingly hybridized fractional-order memory into these metaheuristic frameworks. The Fractional-Order Differential Evolution (FCDE) framework added a tunable non-local search mechanism that adaptively balances exploration and exploitation throughout the optimization process by exploiting the long-term memory and heredity properties fractional calculus formalizes, embedding Caputo fractional derivatives directly into the mutation and update equations of standard Differential Evolution [3]. Tested on the CEC 2022 and CEC 2025 benchmark suites across dimensions up to 1000D, together with three real-world engineering problems, truss design, photovoltaic MPPT control, and dynamic resource allocation, FCDE recorded 17 wins out of 30 functions on CEC 2025 at 100 dimensions and 14 wins out of 30 on CEC 2022 at 30 dimensions against ten state-of-the-art competing algorithms including JADE, LSHADE, CMA-ES, and MRFO, with an average improvement of 12.8% in convergence speed and statistical significance confirmed at $p < 0.01$ across the Wilcoxon comparisons [3]. Critically for the mathematical framework this paper addresses, ablation testing found that search performance is sensitive to the fractional order $\alpha \in [0.3, 1]$, with the best convergence patterns appearing at $\alpha = 0.65$, a specific numerical finding that directly corroborates the theoretical speed-versus-accuracy trade-off formally established for gradient-based fractional methods in Section 2.2 [3], [6].

A related hybrid approach applied fractional-order memory to a physically inspired optimizer rather than to differential evolution specifically. The Fractional-Order Boosted Hybrid Young's Double-Slit Experimental optimizer (FYDSE) addressed the observation that the base YDSE optimizer, although striking a good balance between global and local searches, does not converge as fast as it should and is prone to fall into local optima on complex or high-dimensional problems [13]. FYDSE introduced a fractional-order strategy into the dark-edge position update specifically to ensure more efficient use of the search potential of a single neighborhood space while reducing the possibility of trapping in a local optimum, combined with a piecewise chaotic-mapping initialization strategy to improve the diversity of the initial population, and was validated specifically on truss topology engineering optimization problems [13]. A separate memory-enhanced fractional dung-beetle optimizer (MFO-DBO) addressed a related population-diversity failure mode directly, noting that although integer-order chaotic-map strategies have been widely used in metaheuristic variants, they often fail to ensure diversity and uniform distribution in high-dimensional spaces, impairing global exploration capabilities and producing sluggish convergence rates, a problem the fractional-order logistic chaotic map initialization strategy was specifically designed to address for photovoltaic parameter identification [12].

2.5 Research Gap

While the gradient-based fractional-order literature reviewed in Sections 2.1–2.3 has established formal, theorem-level convergence guarantees under convex, strongly convex, and non-convex smoothness assumptions, and the hybrid metaheuristic literature reviewed in Section 2.4 has established comparably strong empirical evidence for fractional-order memory's benefit on genuinely high-dimensional (up to 1000D), constrained engineering benchmarks, no work reviewed here establishes a formal convergence guarantee for a fractional-order metaheuristic of the FCDE or FYDSE type; their reported guarantees are exclusively statistical (Wilcoxon significance) rather than analytic. This asymmetry, rigorous convergence theorems for differentiable gradient-based fractional methods on the one hand, purely empirical validation for the population-based hybrid methods that dominate practical high-dimensional engineering optimization on the other, constitutes the central mathematical gap this review identifies.

III. Methodology

This study adopts a qualitative, descriptive, and comparative research design synthesizing peer-reviewed and archival literature on fractional-order calculus, convergence theory, and hybrid metaheuristic optimization. Reviewed works were compared along four axes: algorithmic family (gradient-based fractional method or population-based hybrid metaheuristic); guarantee type (formal convergence theorem, monotone-descent proof, or empirical statistical validation only); smoothness assumptions invoked (convexity, strong convexity, L-smoothness, or none); and problem dimensionality range validated (low-dimensional benchmark, high-dimensional benchmark up to 1000D, or applied engineering problem).

IV. Results and Discussion

4.1 Convergence Theorems Mapped to Assumptions and Guarantee Type

Table 1 maps the principal convergence-theoretic results reviewed onto their underlying smoothness assumptions and the strength of guarantee each establishes.

Table 1. Fractional-Order Convergence Results Mapped to Assumptions and Guarantee Type

Method	Smoothness Assumption	Guarantee Type	Convergence Rate	Key Reference
Generic CFGD	Local smoothing interpretation	Variational (steepest-descent of smoothed objective)	Not rate-quantified	Caputo fractional derivative-based algorithm [4]
Fractional GD (convex/strongly convex/non-convex)	L-smoothness + (strong) convexity	Formal theorem-level convergence rate	Linear (strongly convex, 1D case)	Convergence Analysis of Fractional GD [2]
QFOGDM (adaptive stepsize)	High-dimensional unconstrained smoothness	Three-theorem convergence proof	Improved rate vs. fixed stepsize	Quasi fractional order GD [5]
FHGD	Standard gradient Lipschitz continuity	Speed/accuracy trade-off characterization	$\alpha > 1$ faster; $\alpha < 1$ more accurate	Generalization of gradient method [6]
2SED-FOSGD	Bounded gradient + L-smoothness	Rigorous convergence guarantee (stochastic)	Faster, more stable vs. naive FOSGD	Effective Dimension Aware FOSGD [7]
Compressed CFGD (DHDC)	L-smoothness	Monotone descent + linear convergence (bounded error)	Linear, error-controlled	Scalable Caputo FGD [8]

4.2 Hybrid Metaheuristics Mapped to Engineering Application and Reported Performance

Table 2 consolidates the hybrid fractional-order metaheuristics reviewed against the specific high-dimensional engineering problem each was validated on and the statistical performance reported.

Table 2. Hybrid Fractional-Order Metaheuristics: Engineering Validation and Reported Performance

Hybrid Method	Base Algorithm	Engineering Validation	Dimensionality	Reported Performance	Key Reference
FCDE	Differential Evolution	Truss design, PV-MPPT control, dynamic resource allocation	Up to 1000D (CEC 2022/2025)	12.8% avg. convergence speed gain; 38% max. error reduction; $p < 0.01$	Fractional calculus-inspired metaheuristic [3]

FYDSE	Young's Double-Slit Experimental optimizer	Truss topology engineering optimization	High-dimensional	Improved global/local search balance; reduced local-optima trapping	Fractional-Order Boosted Hybrid YDSE [13]
MFO-DBO	Dung Beetle Optimization	Photovoltaic parameter identification	High-dimensional, multimodal	Enhanced convergence stability and solution quality	Memory Enhanced FO-DBO [12]
NMSOLMFO/BFGSOLMFO	Moth-Flame Optimization	Single/double-diode PV models; CEC2014 (30D)	30D+	Outperforms majority of compared methods in accuracy and convergence speed	Metaheuristic fractional Laplacian discretization [14]

4.3 The Fractional-Order Parameter α : A Cross-Method Speed–Accuracy Trade-Off

A consistent quantitative pattern recurs across both the gradient-based and metaheuristic branches of the reviewed literature regarding the fractional order α itself. The gradient-based theoretical result that fractional order greater than 1 provides faster convergence speed, less than 1 gives better steady state performance, and equal to 1 reduces the method to standard descent [6] is corroborated empirically by the metaheuristic literature's finding that FCDE's search performance is sensitive to the fractional order $\alpha \in [0.3, 1]$, with the best convergence patterns appearing at $\alpha = 0.65$, an intermediate value consistent with the theoretical trade-off between exploration speed and steady-state solution accuracy [3]. This convergence of theoretical and empirical evidence across structurally distinct algorithmic families, continuous gradient descent and discrete population-based evolution, strongly suggests that the speed-accuracy trade-off governed by α is a structural property of fractional-order memory itself rather than an artifact specific to either algorithmic family, a finding with direct practical implication: engineers deploying fractional-order methods on new high-dimensional problems can use α as a principled, interpretable tuning parameter rather than an opaque hyperparameter requiring exhaustive grid search.

4.4 Scalability and the Memory-Cost Bottleneck

The reviewed literature also converges on a shared practical obstacle: the Caputo fractional derivative's convolution structure inherently requires weighting the full gradient history, imposing an $O(t)$ per-iteration memory and computation cost that threatens scalability precisely in the high-dimensional, long-horizon optimization runs this paper's engineering applications require [8]. The dyadic hierarchical decomposition and compression scheme's central contribution, formally proving that monotone descent and linear convergence are preserved under bounded compression error, directly targets this scalability bottleneck and represents the most mathematically complete current answer to the tension between fractional memory's optimization benefit and its computational cost [8]. No equivalent formal compression-error analysis currently exists for the hybrid metaheuristic methods reviewed in Section 2.4, which instead manage the same underlying cost through fixed-order, non-adaptive fractional weighting schemes, a further instance of the theory-practice asymmetry identified as this review's central research gap.

V. Future Research Directions

Three priorities emerge from the reviewed literature. First, extending the formal convergence theorems established for gradient-based fractional methods (Section 2.1–2.3) to the population-based hybrid metaheuristics that

currently dominate practical high-dimensional engineering deployment (Section 2.4) represents the field's most consequential open mathematical problem; a first step would be adapting the local-smoothing variational interpretation of CFGD [4] to characterize the effective objective landscape a fractional-order mutation operator such as FCDE's actually searches. Second, the dimension-aware adaptive fractional-order mechanisms developed for stochastic gradient descent [7] should be extended to the metaheuristic setting, replacing the fixed $\alpha \in [0.3, 1]$ range tuned empirically in FCDE [3] with a theoretically grounded, dimension- and landscape-adaptive selection rule. Third, the scalable compression techniques developed for Caputo gradient descent [8] should be benchmarked directly against the fixed-memory fractional metaheuristics reviewed in Section 2.4 on shared high-dimensional engineering benchmarks (CEC 2022/2025, truss design, PV parameter identification) to establish whether compression-based scalability gains transfer across algorithmic families.

VI. Conclusion

This review has examined the mathematical foundations and hybrid algorithmic applications of fractional-order optimization to high-dimensional engineering systems, synthesizing formal convergence-rate theory for Caputo fractional gradient descent with the empirical hybrid-metaheuristic literature applying fractional-order memory to truss design, photovoltaic parameter identification, and dynamic resource allocation. The reviewed evidence establishes that fractional-order memory, formalized through the Caputo derivative's non-local convolution structure, provides both a theoretically justified mechanism for escaping shallow local optima and a practically validated performance improvement across high-dimensional engineering benchmarks up to 1000 dimensions, with convergence-speed gains averaging 12.8% and error reductions of up to 38% relative to state-of-the-art baselines. The fractional order α consistently governs a speed-versus-accuracy trade-off across both gradient-based and population-based algorithmic families, with intermediate values near $\alpha \approx 0.65$ empirically optimal for highly multimodal engineering landscapes. The principal remaining gap is that rigorous, theorem-level convergence guarantees, well established for differentiable gradient-based fractional methods under convex and non-convex smoothness assumptions, have not yet been extended to the population-based hybrid metaheuristics that dominate practical high-dimensional engineering optimization, positioning this theory-practice bridge as the central priority for future research in rigorous, convergence-guaranteed fractional-order optimization.

References:

- [1] W. F. Mbasso, S. Harrison, et al., "Fractional calculus-inspired metaheuristic algorithm for solving high-dimensional constrained optimization problems," *Evolutionary Intelligence*, Springer Nature, 2025.
- [2] "Convergence Analysis of Fractional Gradient Descent," *arXiv preprint arXiv:2311.18426*, 2024.
- [3] W. F. Mbasso et al., "Fractional calculus-inspired metaheuristic algorithm for solving high-dimensional constrained optimization problems (FCDE)," *Evolutionary Intelligence*, Springer Nature, 2025.
- [4] "A Caputo fractional derivative-based algorithm for optimization," *arXiv preprint arXiv:2104.02259*, 2021.
- [5] "A quasi fractional order gradient descent method with adaptive stepsize and its application in system identification," *Applied Mathematics and Computation*, ScienceDirect, 2020.
- [6] "Generalization of the gradient method with fractional order gradient direction," *Journal of the Franklin Institute*, ScienceDirect, 2020.
- [7] "Effective Dimension Aware Fractional-Order Stochastic Gradient Descent for Convex Optimization Problems," *arXiv preprint arXiv:2503.13764*, 2025.
- [8] "Fast and Scalable Caputo Fractional Gradient Descent via Perturbation-Preserving Memory Compression," *arXiv preprint arXiv:2607.15505*, 2026.
- [9] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
- [10] R. Tibshirani, "Gradient Descent: Convergence Analysis," Lecture Notes, Carnegie Mellon University, Statistics Dept., 2013.
- [11] Chapter 8, "Gradient Descent and Convexity," Lecture Notes, McGill University, COMP 451, n.d.
- [12] Y. Li, Z. Allen-Zhao, Y. Xu, and S. Liu, "Memory Enhanced Fractional-Order Dung Beetle Optimization for Photovoltaic Parameter Identification," *arXiv preprint arXiv:2508.06841*, 2025.

- [13] "Fractional-Order Boosted Hybrid Young's Double-Slit Experimental Optimizer for Truss Topology Engineering Optimization," *PMC / NIH*, 2024.
- [14] "A metaheuristic optimization approach to discretize the fractional order Laplacian operator without employing a discretization operator," *Swarm and Evolutionary Computation*, ScienceDirect, 2018.
- [15] R. Higuchi, P.-L. Poirion, and A. Takeda, "Fast Convergence to Second-Order Stationary Point through Random Subspace Optimization," *arXiv preprint arXiv:2406.14337*, 2024.
- [16] "High-Dimensional Composite Quantile Regression: Optimal Statistical Guarantees and Fast Algorithms," *arXiv preprint arXiv:2208.09817*, 2022.
- [17] J.-H. Park, A. J. Salgado, and S. M. Wise, "Preconditioned accelerated gradient descent methods for locally Lipschitz smooth objectives with applications to the solution of nonlinear PDEs," *arXiv preprint arXiv:2006.06732*, 2020.
- [18] K. B. Oldham and J. Spanier, *The Fractional Calculus: Theory and Applications of Differentiation and Integration to Arbitrary Order*. New York: Academic Press, 1974.
- [19] Podlubny, *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*. San Diego: Academic Press, 1999.
- [20] Y. Shin, J. Darbon, and G. E. Karniadakis, "Accelerating Gradient Descent and Adam via Fractional Gradients," *Neural Networks*, vol. 161, pp. 185–201, 2023.
- [21] Y. Chen, Q. Gao, Y. Wei, and Y. Wang, "Study on fractional order gradient methods," *Applied Mathematics and Computation*, vol. 314, pp. 310–321, 2017.
- [22] J. Wang, Y. Wen, Y. Gou, Z. Ye, and H. Chen, "Fractional-order gradient descent learning of BP neural networks with Caputo derivative," *Neural Networks*, vol. 89, pp. 19–30, 2017.
- [23] W. Wang, T. Zhou, and B. Ma, "Convergence analysis of fractional-order incremental gradient methods," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [24] Y. Storn and K. Price, "Differential Evolution – A simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.
- [25] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies (CMA-ES)," *Evolutionary Computation*, vol. 9, no. 2, pp. 159–195, 2001.
- [26] W. Zhu, J. Chen, and B. Y. Zhang, "Manta Ray Foraging Optimization (MRFO): A comparative study," *Engineering Applications of Artificial Intelligence*, vol. 87, 2020.