Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Optimization of High-Dimensional Functions Using Gradient Descent and Genetic Algorithms with Computational Complexity Analysis

Abdul Rehman Nangraj
Lecturer Institute of Mathematics and computer Science. University of Sindh Jamshoro
Corresponding author: nangraj@usindh.edu.pk

Sabira
Assistant Professor, Department of Education and Literacy
Govt. Girls Degree College Kotri
sabirakhurram@gmail.com

Ghulam Yameen Mallah
Assistant Professor, Department of Basic Science and Related Studies
Quaid-e-Awam university of Engineering Science and Technology Nawabshah Sindh
gh.yameen@quest.edu.pk

ABSTRACT

The optimization of high-dimensional functions poses significant challenges due to the curse of dimensionality, data sparsity, metric concentration, and the proliferation of saddle points in non-convex landscapes. This paper provides a comprehensive analysis of two dominant paradigms in high-dimensional optimization: gradient-based methods and genetic algorithms (GAs). Gradient descent variants, including Stochastic Gradient Descent (SGD), Momentum, Adagrad, and Adam, leverage local curvature information for efficient local refinement, while genetic algorithms employ population-based evolutionary search with selection, crossover, and mutation to explore the global search space. Special attention is given to Large-Scale Global Optimization (LSGO) techniques, particularly Cooperative Coevolution (CC) and variable interaction detection methods such as Min-Hash Decomposition (MHD), which address the exponential growth of the search space through decomposition strategies. A detailed computational complexity analysis compares the scaling behavior of batch gradient descent, SGD, Adam, and genetic algorithms, highlighting trade-offs between precision, convergence speed, and resource consumption. The study further examines hybrid approaches that integrate the global exploratory power of genetic algorithms

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

with the exploitative precision of gradient-based optimizers, such as the Adam RNA-GA framework. Benchmark results from CEC 2013 and aerodynamic shape optimization demonstrate the strengths and limitations of each paradigm. Emerging trends, including Strategic Monte Carlo Optimization (SMCO), Quality-Diversity (QD) optimization, and green optimization frameworks, point toward future directions in the field. This work underscores the necessity of hybrid and strategically guided methods for effectively navigating the complex hyper-surfaces of high-dimensional optimization problems in machine learning, engineering, and scientific computing.

Keywords: High-dimensional optimization, Gradient Descent, Genetic Algorithms, Curse of Dimensionality, Saddle Points, Cooperative Coevolution, Computational Complexity, Adam Optimizer, Large-Scale Global Optimization (LSGO), Hybrid Optimization

1. Introduction: The Convergence of Complexity and Dimension

The optimization of high-dimensional functions represents one of the most significant challenges in contemporary computational science, forming the bedrock of advancements in deep learning, bioinformatics, and complex engineering systems (Yousef & Allmer, 2023). As the dimensionality of the decision space, denoted by d, scales from hundreds to millions of parameters, the fundamental nature of the search landscape undergoes a qualitative transformation. Traditionally, optimization focused on convex surfaces where the local gradient provided an unerring path to a global minimum. However, in the high-dimensional regime, the emergence of non-convexity, data sparsity, and the proliferation of saddle points necessitates a sophisticated re-evaluation of algorithmic strategies (Dauphin et al., 2014). The dichotomy between gradient-based methods, which leverage local curvature information for rapid convergence, and genetic algorithms, which employ stochastic population-based search to maintain global diversity, defines the current state of the art. Understanding the interplay between these two paradigms requires not only a mastery of their respective mechanisms but also a rigorous analysis of their computational complexity and scaling behaviors (Jin et al., 2017).

High-dimensional global optimization (LSGO) is typically characterized by problems where d \geq 1,000. In such environments, the "curse of dimensionality" is no longer a theoretical abstraction but a dominant physical reality that dictates the feasibility of solution discovery. The volume of the search space expands exponentially with each added dimension, causing data points to become increasingly sparse and rendering traditional distance-based heuristics ineffective (Schumann & Aragón, 2021). This report provides an exhaustive review of the methodologies employed to navigate these complex hyper-surfaces, focusing on the evolutionary refinements of gradient descent and the decomposition-driven scaling of genetic algorithms, while synthesizing their performance through the lens of computational complexity (Meselhi et al., 2018).

2. The Mathematical and Statistical Topology of the High-Dimensional Frontier

The transition into high-dimensional optimization is marked by the phenomenon known as the curse of dimensionality, a term coined by Richard Bellman to describe the exponential increase in the volume of a mathematical space when extra dimensions are added (Ovbije et al., 2025). This expansion has profound implications for both the modeling of functions and the efficiency of search algorithms. In a unit hypercube, the amount of data required to maintain a constant

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

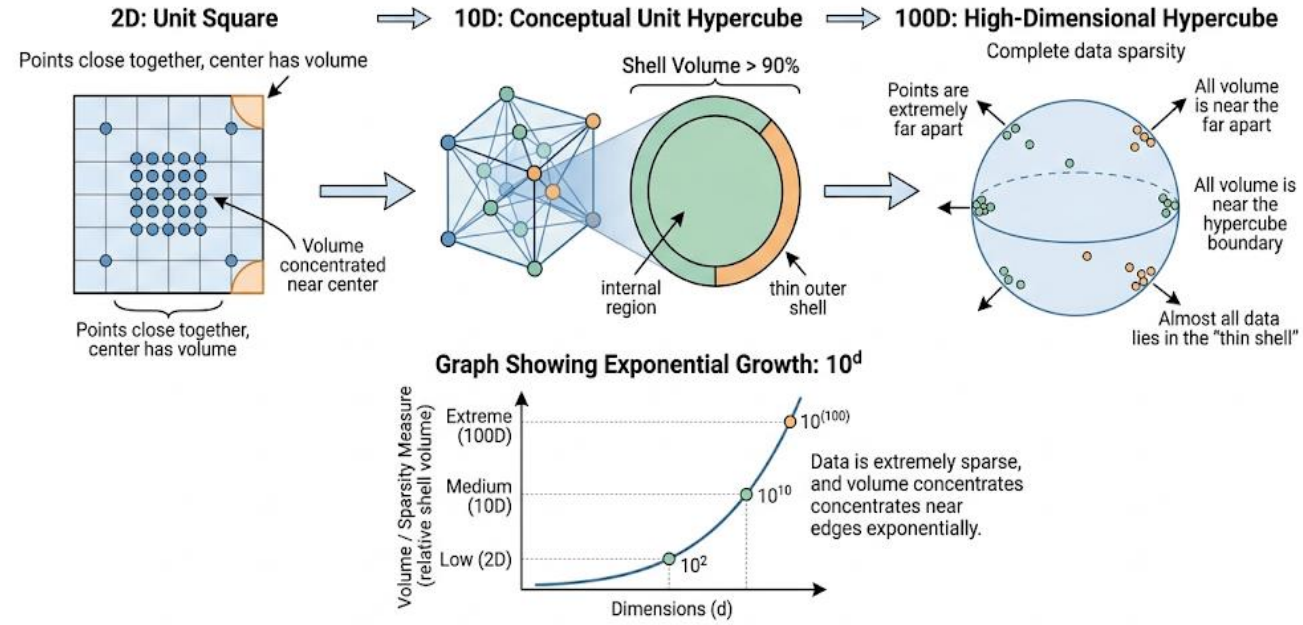
Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

sampling density grows as 10^d, a requirement that quickly exceeds any available computational or observational resource (Tailor, 2024). Figure 1 illustrates the geometric consequences of increasing dimensionality, highlighting the exponential expansion of the search space and resulting data sparsity. As shown, the dispersion of data points increases dramatically with dimension, reinforcing the challenges associated with efficient optimization.

2.1. The Statistical Curse and Data Sparsity

Beyond the combinatorial explosion of the search space, the statistical manifestation of the

Figure 1. Geometric Effects of the Curse of Dimensionality in High-Dimensional Spaces.



curse of dimensionality often referred to as the empty space phenomenon describes the progressive sparsity of data distributions as dimensionality increases (Verleysen et al., 2003). In high-dimensional settings, the volume of the space expands exponentially, causing finite datasets to occupy an increasingly negligible fraction of the domain. For example, in a 10-dimensional hypercube, a sub-cube that contains only 1% of the total volume must extend across approximately 63% of the range along each coordinate axis. This geometric property illustrates how regions that appear small in volume nevertheless span a large portion of each dimension. As a result, local averaging methods and density estimation techniques become unreliable because meaningful neighborhoods are rarely populated with sufficient data points. In optimization contexts, this implies that algorithms have limited local information to guide search directions unless they happen to operate within regions of unusually high sample density, an event that is statistically unlikely in high-dimensional spaces (Devi et al., 2016).

In addition to sparsity, high-dimensional geometry introduces another challenge known as the concentration of measure phenomenon. Under this effect, pairwise distances between data points become increasingly similar as dimensionality grows, reducing the discriminative power of distance-based similarity metrics (Radovanović et al., 2009). This behavior is commonly observed when using the Euclidean distance, which can be expressed as follows:

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

where $x = (x_1, x_2, \dots, x_n)$and $y = (y_1, y_2, \dots, y_n)$denote points in an n-dimensional space, and $d(x,y)$represents the Euclidean distance between them. As $n$increases, the relative difference between the minimum and maximum distances among points diminishes, leading to a collapse in metric informativeness. Consequently, algorithms that rely on distance comparisons—such as k-nearest neighbors' classifiers, clustering methods, or distance-based evolutionary operators—struggle to distinguish between promising and unpromising regions of the search space, thereby degrading their performance and convergence reliability in high-dimensional optimization problems (Radovanović et al., 2009).

2.2. The Proliferation of Saddle Points and High-Error Plateaus

Perhaps the most critical insight in high-dimensional non-convex optimization is the realization that the primary impediment to reaching a global minimum is not the presence of local minima, but the overwhelming abundance of saddle points. In low-dimensional spaces, our geometric intuition suggests that a function's critical points are mostly peaks or valleys. However, as the dimensionality d increases, the probability that a random critical point is a local minimum meaning all d eigenvalues of the Hessian matrix are positive is exponentially small, O(e^{-d}) (Choromanska et al., 2015).

Instead, the vast majority of critical points in high-dimensional landscapes are saddle points, where the surface curves upward in some directions and downward in others. These points are frequently surrounded by high-error plateaus where the gradient magnitude is near zero, causing first-order optimization methods to slow down dramatically (Aghajanyan, 2016). This phenomenon often creates the "illusion" of a local minimum, as the objective function ceases to improve for many iterations, even though the algorithm is simply traversing a region of low curvature (Pascanu et al., 2014).

Table 1. High-Dimensional Geometric Features and Theoretical Implications

Geometric Feature	High-Dimensional Impact	Theoretical Implication
Volume Expansion	Search space becomes intractable for random/grid search.	Necessity for directed search or structured search (Ovbije et al., 2025).
Metric Concentration	Euclidean distance becomes uninformative.	Shift toward cosine similarity or Mahalanobis distance (Tailor, 2024).
Saddle Points	Dominant critical point type; leads to slow convergence.	Plateaus provide vanishing gradients in non-optimal regions (Dauphin et al., 2014).
Hessian Spectrum	Negative eigenvalues are more probable than all-positive.	Second-order methods help detect directions of negative curvature (Dauphin et al., 2014).

3. Gradient-Based Optimization: Mechanisms of Local Refinement

Gradient descent (GD) and its variants are the primary optimization methods used for minimizing differentiable high-dimensional functions, particularly in the training of artificial neural networks. The fundamental principle of GD is to iteratively update model parameters in the direction opposite to the gradient of the objective function. This update rule is expressed as:

$\theta_{t+1} = \theta_t - \eta \nabla J(\theta_t)$

Where θ_t represents the parameter vector at iteration t, η is the learning rate, and $\nabla J(\theta_t)$ denotes the gradient of the objective function evaluated at θ_t (Ruder, 2016).

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

3.1. The Evolution from Batch to Stochastic Variants

The classical batch gradient descent algorithm computes the gradient of the loss function using the entire training dataset prior to performing a single parameter update. Although this approach yields a stable and deterministic convergence trajectory—particularly for convex optimization landscapes—it becomes computationally impractical when applied to the large-scale datasets commonly encountered in high-dimensional machine learning tasks. This limitation motivated the widespread adoption of Stochastic Gradient Descent (SGD), which updates model parameters using only a single training instance or a small subset of data known as a mini-batch (Lee et al., 2016).

The parameter update rule for gradient-based optimization can be expressed as:

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t)$$

where $\theta_t$denotes the parameter vector at iteration t, $\eta$is the learning rate controlling the step size, and $\nabla L(\theta_t)$represents the gradient of the loss function with respect to the parameters. In batch gradient descent, this gradient is computed using the full dataset, whereas in SGD and mini-batch methods, it is estimated from a subset of samples.

A defining characteristic of SGD is the high variance in its gradient estimates, which introduces stochastic fluctuations often interpreted as optimization noise into the learning process. While this stochasticity can produce a more irregular convergence trajectory, it also provides a practical advantage by enabling the algorithm to escape shallow local minima and traverse flat regions such as saddle points in non-convex optimization landscapes. Mini-batch gradient descent represents a practical compromise between the stability of batch methods and the efficiency of stochastic updates. By averaging gradients over small batches of samples, it reduces update variance relative to pure SGD while maintaining substantial computational efficiency through vectorized operations and parallel processing on modern GPU architectures (Bray & Dean, 2007).

.

3.2. Advanced Adaptive Optimizers and Momentum

The sensitivity of standard gradient descent (GD) to the choice of learning rate ηrepresents a significant limitation, particularly in complex, high-dimensional optimization landscapes where inappropriate step sizes can lead to divergence or excessively slow convergence. Modern adaptive optimization algorithms address this issue by assigning parameter-specific learning rates that evolve according to the historical behavior of gradients, thereby improving stability and convergence efficiency across heterogeneous feature spaces (Ge et al., 2015).

Momentum is one of the earliest and most influential enhancements to gradient-based optimization. Drawing inspiration from classical mechanics, the method introduces a velocity term that accumulates past gradients to smooth oscillations in directions of high curvature while accelerating movement along consistent descent directions (Xu et al., 2016). The velocity update mechanism is defined as:

$$v_t = \gamma v_{t-1} + \eta \nabla J(\theta_t)$$

where v_t represents the current velocity, γ is the momentum coefficient controlling the contribution of previous updates, and $\nabla J(\theta_t)$denotes the gradient of the objective function. By integrating historical gradient information, momentum effectively damps oscillatory behavior in steep ravines and promotes faster convergence along shallow but consistent gradients.

Adagrad extends this adaptive philosophy by modifying the learning rate according to the cumulative magnitude of past gradients. Specifically, parameters associated with infrequent features receive relatively larger updates, making the algorithm particularly suitable for sparse datasets such as those encountered in natural language processing and recommendation systems (Nacson et al., 2019). The parameter update rule for Adagrad can be expressed as:

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{G_t + \epsilon}} \nabla J(\theta_t)$$

where $G_t$denotes the accumulated sum of squared gradients up to iteration t, and $\epsilon$is a small constant introduced to ensure numerical stability. This adaptive scaling mechanism allows Adagrad to automatically reduce step sizes for frequently updated parameters while preserving learning capacity for rare but informative features.

3.3. Theoretical Convergence in Non-Convex Landscapes

The theoretical understanding of these optimizers in high-dimensional non-convex settings has advanced significantly. Recent research into the Adam optimizer has established that it converges to \epsilon -stationary points with a gradient complexity of O(\epsilon^{-4}) under generalized smoothness conditions, even without the assumption of globally bounded gradients. Furthermore, a new characterization of the "Adam vector field" reveals that the optimizer converges to zeros of a macroscopic vector field that describes its average behavior (He et al., 2023).

To address the saddle point problem directly, "perturbed" gradient descent methods have been developed. These algorithms periodically add small amounts of random noise to the iterates. Because the region of attraction for a saddle point is a thin, high-dimensional band, random perturbations ensure that with high probability, the new state falls into a region where the negative curvature of the Hessian allows for rapid escape (Jin et al., 2017).

4. Genetic Algorithms: Scaling Global Search via Evolution

While gradient-based methods are efficient local optimizers, they are fundamentally limited by their reliance on local derivative information. Genetic Algorithms (GAs) provide a global search alternative by simulating natural evolution through selection, crossover, and mutation within a population of candidate solutions (Ovbije et al., 2025).

4.1. The Architecture of Genetic Operators in High Dimensions

A standard GA begins by encoding potential solutions as "chromosomes" typically vectors of real numbers or bitstrings in the d-dimensional space. The fitness of each individual is evaluated via the objective function, and those with higher fitness are selected for reproduction (Li et al., 2023)

- **Crossover (Recombination):** This operator combines parts of two parent chromosomes to create offspring. In high dimensions, operators like blended crossover are used to explore the mid-space between parents (Ma et al., 2022).
- **Mutation:** Mutation introduces random changes to a chromosome, ensuring population diversity and preventing premature convergence (Abulail, 2025).

4.2. The LSGO Decomposition Strategy

The primary challenge for GAs in high dimensions is the exponential expansion of the search space. Large-Scale Global Optimization (LSGO) techniques employ a "divide and conquer" approach. The most popular framework is Cooperative Coevolution (CC), which decomposes the d-dimensional problem into m smaller sub-problems (Fonseca et al., 2022).

The effectiveness of this decomposition depends on variable interactions. A function f(x) is partially separable if it can be decomposed into m independent subgroups, where optimizing each subgroup independently leads to the global optimum (Tian et al., 2024).

Table 2. Interaction Types and Corresponding Optimization Strategies

Interaction Type	Definition	Optimization Strategy
Fully Separable	f(x) = \sum f_i(x_i) for all i.	Direct optimization of individual components (Ma et al., 2022).
Partially	Variables form disjoint,	Grouping-based Cooperative Coevolution

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Separable	independent subgroups.	(Ma et al., 2022).
Fully Non-separable	Every variable interacts with every other variable.	Non-decomposition methods or high-density local search (Ma et al., 2022).

4.3. Advanced Grouping and Similarity Metrics

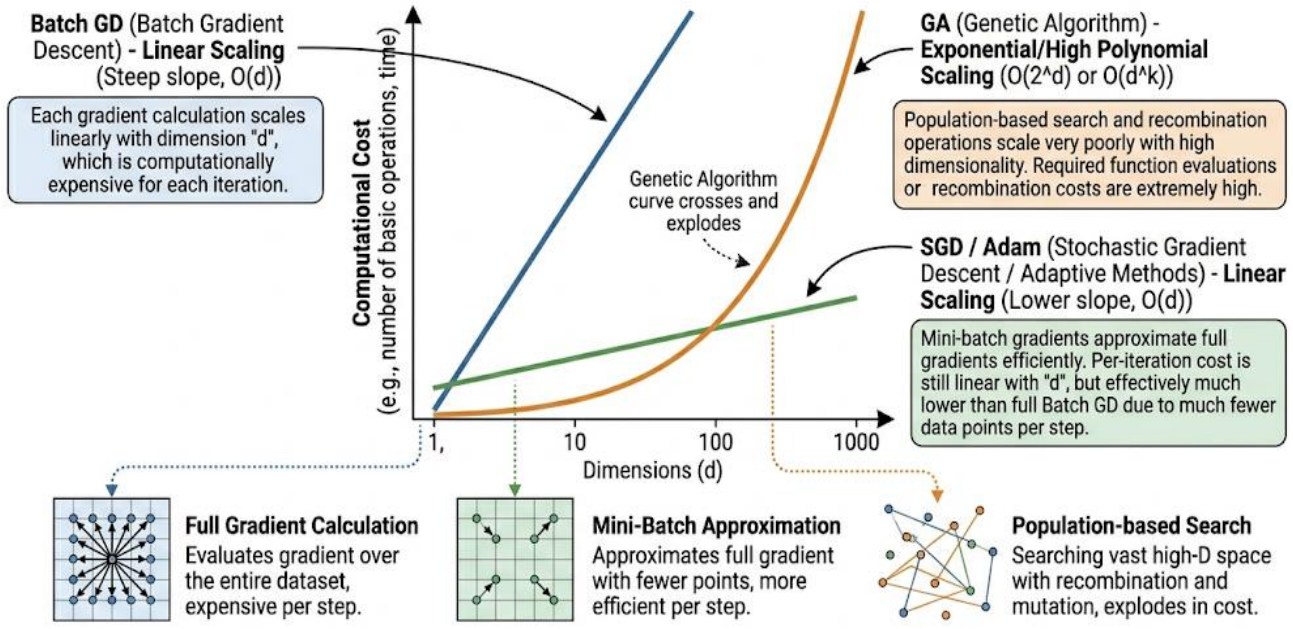
To achieve accurate decomposition, modern LSGO algorithms use interaction detection methods. For example, the Min-Hash Decomposition (MHD) algorithm uses "min-hash" permutations and the Jaccard similarity coefficient to cluster interdependent variables into sub-components with minimal inter-group dependencies (Xie et al., 2025).

5. Computational Complexity and Scaling Analysis

The choice between gradient-based and evolutionary optimization is often a trade-off between solution precision and computational resources. A Big O analysis reveals the scaling laws governing their performance as dimensionality d increases (Chen et al., 2025). Figure 2 compares the scaling behavior of gradient-based and evolutionary algorithms as dimensionality increases.

The results emphasize the efficiency of gradient methods in high-dimensional settings and the higher computational burden of genetic algorithms.

Figure 2. Computational Complexity Scaling of Optimization Algorithms



5.1. Big O Complexity Comparison

The computational cost of gradient-based optimization methods is primarily determined by the evaluation of the gradient vector. For a dataset containing N samples and d features, **Batch Gradient Descent (GD)** has a per-iteration computational complexity of:

$$O(N \cdot d)$$

In contrast, Stochastic Gradient Descent (SGD) significantly reduces the per-iteration cost to:

$$O(d)$$

since it computes gradients using only a single sample (or a small mini-batch). However, in engineering applications where the objective function depends on numerical simulations (e.g., CFD or finite element models), the dominant computational cost arises from expensive function

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

evaluations rather than gradient computation (Anderson & Zingg, 2008).

Table 3. Computational Complexity Comparison and Scaling Constraints

Algorithm	Per-Iteration Time Complexity	Scaling Constraint
Batch GD	$O(N \cdot d)$	Slow for large N; memory bottleneck (Ruder, 2016)
SGD	$O(d)$	High variance; requires many iterations (Nishchal & Bhandari, 2021)
Adam	$O(d)$	Additional moment estimates increase memory usage (Li et al., 2023)
Standard GA	$O(P \cdot d \cdot f)$	Fitness evaluation cost f is computationally expensive (Anderson & Zingg, 2008)
LSGO (CC)	$O(\sum P_i \cdot d_i \cdot f_i)$	Decomposition accuracy is the main bottleneck (Ma et al., 2022)

5.2. Scaling and Efficiency in High-Fidelity Simulations

Empirical studies in aerodynamic shape optimization demonstrate substantial differences in computational efficiency between gradient-based and evolutionary methods. In discrete-adjoint gradient-based optimization, the gradient computation scales approximately linearly with the number of design variables d, making it highly efficient even for large-scale problems.

In contrast, genetic algorithms (GAs) applied to the same high-fidelity design space require significantly more function evaluations. In practice, they may need **5 to 200 times more evaluations** than gradient-based methods to reach comparable optimal solutions (Ruder, 2016). This disparity highlights the computational advantage of gradient-informed methods in simulation-driven optimization problems.

6. Integrative Paradigms: Synthesizing Global and Local Search

The limitations of standalone gradient descent and genetic algorithms have led to the development of hybrid paradigms. These systems capitalize on the exploratory power of evolution while utilizing the exploitative precision of gradients (Nishchal & Bhandari, 2021).

6.1. The Two-Stage Evolutionary Strategy

A prevalent hybrid structure involves using a GA in the initial stage to survey the search space. Once the population has converged toward a likely optimal region, a gradient-based refiner is applied to elite individuals. In heart disease prediction tasks, a hybrid GA-SVM model demonstrated significantly higher accuracy (95%) compared to traditional grid search. Similarly, in complex medical analysis, hybrid AI-driven approaches have been utilized to enhance diagnostic precision (Sindhushree et al., 2025).

6.2. Case Study: Adam RNA-GA

A sophisticated example is the Adam RNA-GA, which incorporates RNA-inspired crossovers and mutations to maintain diversity while escaping local optima. The algorithm applies Adam-based gradient mutation only to high-quality individuals in later stages. To manage computational load, the Adam RNA-GA restricts gradient operations to the optimal individual and employs adaptive descent precision (Wu et al., 2024). In benchmark tests, this hybrid approach outperformed standalone GAs and other advanced meta-heuristics like the Grey Wolf Optimizer (Murad & Tayfor, 2024).

7. Theoretical Benchmarking and Performance Metrics

Evaluating high-dimensional optimizers requires standardized test suites that reflect real-world complexities (Yousef & Allmer, 2023).

7.1. Analysis of CEC 2013 Benchmark Results

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

In 1,000-dimensional tests, the performance of LSGO algorithms is measured across several function categories. The MHD algorithm demonstrated superior accuracy on partially separable functions (f4-f11). This is attributed to its ability to cluster active decision variables into subgroups using Jaccard similarity, which improves search stability. On overlapping functions like f14, while most algorithms fell into premature convergence, MHD maintained a consistent convergence trend, eventually obtaining better solutions (Ma et al., 2022).

7.2. Comparative Performance Metrics in Aerodynamics

In aerodynamic shape optimization, comparisons between discrete-adjoint methods and genetic algorithms indicate that both reliably converge to the same global optimum. However, the genetic algorithm's inability to exploit gradient information result in a computational cost that increases dramatically as tighter convergence requirements are applied, highlighting the necessity of gradient information for high-fidelity tasks (Dauphin et al., 2014).

8. Emerging Paradigms and Future Outlook (2025-2026)

The field of optimization is entering an era of "strategically guided randomness" and AI-driven discovery (Beyer, 2025).

8.1. Strategic Monte Carlo Optimization (SMCO)

Introduced in 2026, Strategic Monte Carlo Optimization (SMCO) represents a paradigm shift toward viewing optimization as a series of strategic "two-armed" choices. Grounded in the Strategic Law of Large Numbers, SMCO requires only the direction of increase/decrease rather than exact gradients, making it robust in highly irregular or non-smooth landscapes (Chen et al., 2026).

8.2. Quality-Diversity (QD) Optimization

The shift from finding a single optimum to generating a diverse set of alternatives is embodied in Quality-Diversity (QD) optimization. Algorithms like MAP-Elites partition behavior space into "niches," maintaining an elite individual in each to provide a "toolbox" of solutions for uncertain environments. Furthermore, recent theoretical advancements in population sizing and restart strategies for evolution strategies emphasize identifying optima in multimodal landscapes (Schönenberger & Beyer, 2025).

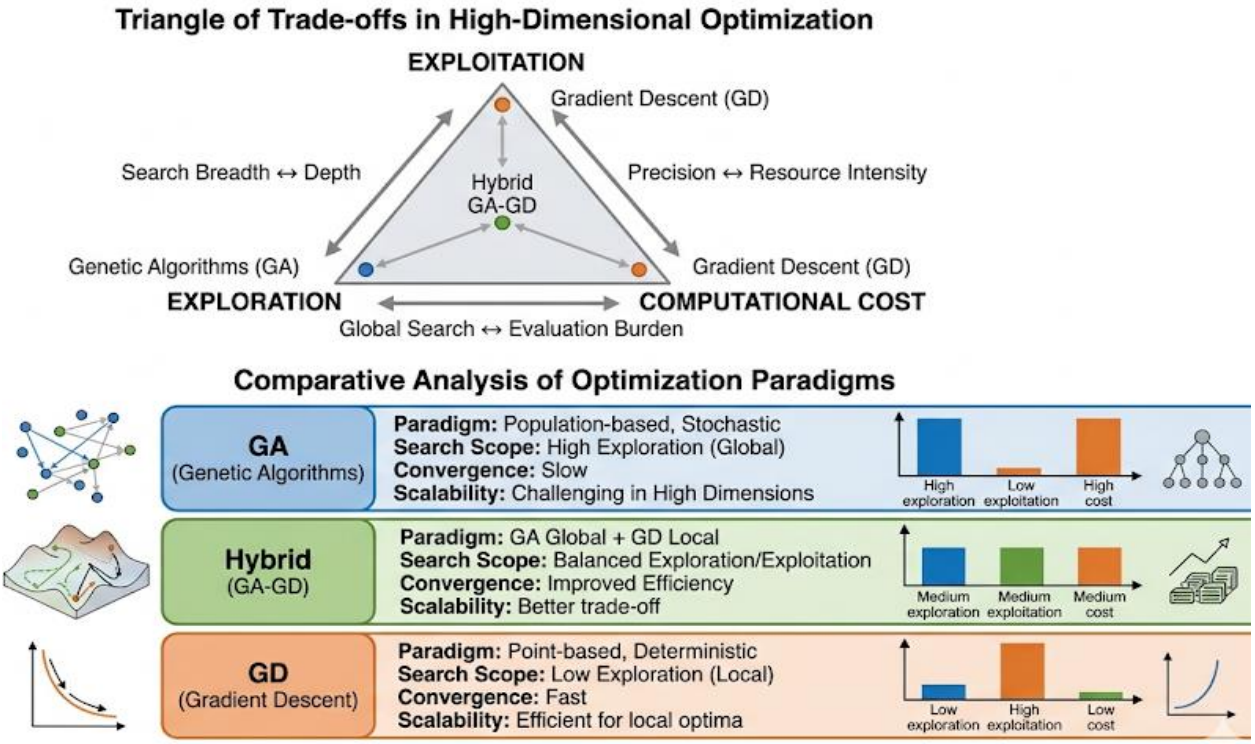
8.3. Ethical and Green Optimization

As optimization scales to influence healthcare allocation and climate modeling, there is a burgeoning movement toward "Ethical Optimization Frameworks" to ensure equitable outcomes. Simultaneously, "Green Optimization" aims to minimize the energy consumption of the optimization process itself, moving from computationally expensive methods toward more efficient surrogate modeling and sparse Gaussian process models (Schönenberger & Beyer, 2025). Figure 3 summarizes the trade-offs between gradient-based, evolutionary, and hybrid optimization approaches. It highlights the necessity of hybrid frameworks for balancing computational efficiency and global search capability.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Figure 3: Comparative Landscape of Optimization Paradigms in High Dimensions



9. Conclusion

This paper has examined the intricate challenges and algorithmic strategies involved in optimizing high-dimensional functions, where the curse of dimensionality, data sparsity, metric concentration, and the overwhelming presence of saddle points fundamentally alter the optimization landscape. Gradient-based methods, particularly advanced variants such as Stochastic Gradient Descent (SGD), Momentum, Adagrad, and the Adam optimizer, excel at local refinement by exploiting derivative information, achieving near-linear scaling with dimensionality d d d and demonstrating strong theoretical convergence guarantees even in non-convex settings. However, their reliance on local gradients makes them susceptible to high-error plateaus and saddle points, often requiring perturbations or adaptive mechanisms to escape these regions effectively. In contrast, genetic algorithms provide robust global search capabilities through population-based evolution, incorporating selection, crossover, and mutation operators. To scale effectively in large-scale global optimization (LSGO), techniques such as Cooperative Coevolution (CC) and advanced decomposition methods like Min-Hash Decomposition (MHD) leverage variable interaction detection to partition the problem into manageable sub-components. While effective for exploring vast search spaces, GAs typically incur higher computational costs due to the large number of fitness evaluations required. A rigorous Big O complexity analysis reveals the fundamental trade-offs: gradient-based approaches offer superior efficiency for differentiable functions and high-fidelity simulations, whereas evolutionary methods provide essential diversity and robustness in multimodal or black-box scenarios. Hybrid

Online ISSN 3007-3197
Print ISSN 3007-3189

paradigms that combine the exploratory strength of genetic algorithms with the precision of gradient descent exemplified by the Adam RNA-GA algorithm emerge as particularly promising, delivering improved accuracy and convergence in complex domains such as medical diagnostics and engineering design. Looking forward, the field is shifting toward integrative strategies that embrace "strategically guided randomness," including Strategic Monte Carlo Optimization (SMCO) and Quality-Diversity (QD) algorithms like MAP-Elites. As optimization increasingly influences critical applications in healthcare, climate modeling, and artificial intelligence, considerations of ethical outcomes and computational sustainability ("green optimization") will become equally vital. Ultimately, effective high-dimensional optimization requires a sophisticated synthesis of local gradient information and global evolutionary search. Future research should focus on developing more adaptive hybrid frameworks, improving decomposition techniques for fully non-separable problems, and advancing energy-efficient surrogate models to ensure scalable, robust, and responsible optimization in an era of ever-increasing dimensionality.

References

Anderson, D. W., & Zingg, D. W. (2008). *Aerodynamic shape optimization: Genetic and gradient-based algorithms*. University of Toronto Institute for Aerospace Studies. http://oddjob.utias.utoronto.ca/dwz/Miscellaneous/remn08.pdf

Beyer, H.-G. (2025). On a population sizing model for evolution strategies in multimodal landscapes. *IEEE Transactions on Evolutionary Computation*. https://doi.org/10.1109/tevc.2024.3419931

Chen, X., Chen, Z., Gao, W. Y., Yan, X., & Zhang, G. (2026, January 19). Optimization via the strategic law of large numbers. *Proceedings of the National Academy of Sciences*. https://cowles.yale.edu/news/260119/scientists-introduce-breakthrough-algorithm-strategy-outperforms-classic-optimization

Dauphin, Y. N., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., & Bengio, Y. (2014). Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *Advances in Neural Information Processing Systems*, *27*, 2933–2941.

He, M., Liang, Y., Liu, J., & Xu, D. (2023). Convergence of Adam for non-convex objectives: Relaxed hyperparameters and non-ergodic case. *arXiv*. https://doi.org/10.48550/arXiv.2307.11782

Jin, C., Ge, R., Netrapalli, P., Kakade, S. M., & Jordan, M. I. (2017). How to escape saddle points efficiently. *Proceedings of the 34th International Conference on Machine Learning*, *70*, 1724–1732.

Li, H., Rakhlin, A., & Jadbabaie, A. (2023). Convergence of Adam under generalized smoothness without boundedness of gradients. *arXiv*. https://doi.org/10.48550/arXiv.2309.07157

Ma, J., Chang, F., & Yu, X. (2022). Large-scale evolutionary optimization approach based on decision space decomposition. *Frontiers in Energy Research*, *10*, Article 926161. https://doi.org/10.3389/fenrg.2022.926161

Murad, S. H., & Tayfor, N. B. (2024). Hybrid genetic algorithms-driven optimization of machine learning. *Health and Technology*. https://pmc.ncbi.nlm.nih.gov/articles/PMC12305601/

Nishchal, J., & Bhandari, N. (2021). Computational complexity of gradient descent algorithm. *TechRxiv*. https://doi.org/10.36227/techrxiv.14544000.v1

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Ovbije, O. G., Joseph, O. A., & Joshua, A. S. (2025). The evolving landscape of optimization: Current trends and future directions. *International Journal of Modern Science and Research Technology*, *3*(8), 108–115. https://doi.org/10.5281/zenodo.16894145

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv*. https://doi.org/10.48550/arXiv.1609.04747

Schönenberger, L., & Beyer, H.-G. (2025). Optimal restart strategies for parameter-dependent optimization algorithms. In *Foundations of Genetic Algorithms XVIII (FOGA '25)*. ACM. https://doi.org/10.1145/3594805.3607126

Sindhushree, S. R., Himel, S. K., Okobi, O. E., & Kushimo, A. (2025). Artificial intelligence based early detection of skin cancer using dermoscopic image analysis. *ResearchGate*. https://www.researchgate.net/publication/403492735

Tailor, G. (2024, July 31). The curse of dimensionality: Challenges and solutions. *Medium*. https://medium.com/@gauravtailor43/the-curse-of-dimensionality-challenges-and-solutions-285f4f98e6a3

Wu, Q., Zhao, Z., Chen, M., Chi, X., Zhang, B., Wang, J., Zhilenkov, A. A., & Chepinskiy, S. A. (2024). An RNA evolutionary algorithm based on gradient descent for function optimization. *Journal of Computational Design and Engineering*, *11*(4), 332–357. https://doi.org/10.1093/jcde/qwae054

Meselhi, M., Sarker, R., Essam, D., & Elsayed, S. (2018). Enhanced differential grouping for large scale optimization. *Proceedings of the 7th International Conference on Operations Research and Enterprise Systems*, 217–224. https://doi.org/10.5220/0006938902170224

Schumann, J. F., & Aragón, A. M. (2021). A machine learning approach for fighting the curse of dimensionality in global optimization. *arXiv*. https://doi.org/10.48550/arxiv.2110.14985

Yousef, M., & Allmer, J. (2023). Deep learning in bioinformatics. *Turkish Journal of Biology*, *47*, 366–382. https://doi.org/10.55730/1300-0152.2671

Xu, J., Bai, J., Zhang, X., Lv, H., Ji, J., Liu, P., Chen, Y., & Li, X. (2016). A comprehensive overview of lncRNA annotation resources. *Briefings in Bioinformatics*, *17*(6), 1047–1057. https://doi.org/10.1093/bib/bbw015

Devi, G., Chauhan, C., & Chakraborti, S. (2016). Conceptualizing curse of dimensionality with parallel coordinates. *Proceedings of the AAAI Conference on Artificial Intelligence*, *30*(1). https://doi.org/10.1609/aaai.v30i1.9855

Radovanović, M., Nanopoulos, A., & Ivanović, M. (2009). Nearest neighbors in high-dimensional data. *Proceedings of the 26th Annual International Conference on Machine Learning*, 865–872. https://doi.org/10.1145/1553374.1553485

Verleysen, M., Francois, D., Simon, G., & Wertz, V. (2003). On the effects of dimensionality on data analysis with neural networks. *Lecture Notes in Computer Science*, 105–112. https://doi.org/10.1007/3-540-44869-1_14

Nacson, M. S., Srebro, N., & Soudry, D. (2019, April). Stochastic gradient descent on separable data: Exact convergence with a fixed learning rate. In *The 22nd International Conference on Artificial Intelligence and Statistics* (pp. 3051-3059). PMLR.

Abulail, R. (2025). Premature Avoidance in Genetic Algorithm using Dynamic Mutation Probability. *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl*, *16*, 528-542.

Fonseca, T. H. L., Nassar, S. M., de Oliveira, A. C. M., & Agard, B. (2022). Low-dimensional space modeling-based differential evolution for large-scale global optimization problems. *IEEE Transactions on Evolutionary Computation*, *27*(5), 1529-1543.

Online ISSN 3007-3197
Print ISSN 3007-3189

Tian, M., Chen, M., Du, W., Tang, Y., Jin, Y., & Yen, G. G. (2024). A composite decomposition method for large-scale global optimization. *IEEE Transactions on Artificial Intelligence*, *5*(12), 6304-6319.

Xie, Z., Tian, W., Li, R., Lu, J., Xiao, W., & Xu, Z. (2025). Horse-minhash: high-performance and secure jaccard similarity estimation for cloud storage. *IEEE Transactions on Dependable and Secure Computing*.

Chen, W., Lin, B., Zhang, X., Lin, X., Zhao, H., Zhang, Q., & Kwok, J. T. (2025). Gradient-based multi-objective deep learning: Algorithms, theories, applications, and beyond. *arXiv preprint arXiv:2501.10945*.