OBJECTIVE-PRESCRIBED ESCAPE FOR GRADIENT DESCENT IN NON-CONVEX OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Gradient-based optimization methods may stagnate in local attraction basins when applied to non-convex objective functions. This paper proposes a deterministic, first-order escape mechanism that preserves standard gradient descent as the default optimization procedure. An objective-prescription mechanism is activated only when stagnation is detected. Rather than modifying the gradient direction or the objective landscape, the proposed approach prescribes a strictly decreasing evolution of the objective function and determines the step size required to realize the prescribed objective value along the current gradient direction. The resulting escape step is obtained by solving a scalar nonlinear equation, which may admit multiple solutions. Small displacements are discarded, and deterministic root-selection strategies are considered to identify an effective escape step. After this temporary intervention, the algorithm immediately returns to standard gradient descent with its predefined step size. Numerical experiments on one-dimensional non-convex benchmark functions demonstrate that the proposed mechanism can escape local attraction basins and reach the global minimum from a broad range of initial conditions. The results provide an initial numerical validation of the proposed deterministic, objective-prescribed escape principle.

1 INTRODUCTION

Local minima and stagnation can constitute a significant limitation of gradient-based optimization in non-convex landscapes. When an iterative trajectory approaches a local minimum or enters a region of very small gradient, conventional gradient descent may become unable to make significant progress or explore other regions of the objective landscape. Existing approaches address this issue through stochastic perturbations, restart strategies, adaptive mechanisms, or additional curvature information. In this work, we investigate a different approach that remains fundamentally gradient-based. The proposed method does not replace gradient descent. Instead, gradient descent constitutes the default optimization mechanism, while an auxiliary objective-prescription mechanism is activated only when stagnation is detected. Its purpose is to generate a sufficiently large displacement capable of escaping the current attraction basin. Once the escape step has been performed, the algorithm immediately returns to standard gradient descent with its predefined step size. The central idea is to use the gradient direction not only for local descent, but also as a search direction along which a temporary evolution of the objective function is prescribed. The prescribed evolution is chosen to be strictly decreasing, so that each prescribed step targets a lower objective value than the current one. Rather than selecting the escape step directly, the method determines the step size required to realize the prescribed objective value along the current gradient direction:

$$g(x_k - \eta_k g'(x_k)) = g(x_{k+1})^{\text{target}}$$

This transforms the escape problem into an inverse step-size determination problem. Depending on the geometry of the objective function, the resulting equation may admit no solution, a single solution, or multiple admissible solutions. When several solutions exist, a deterministic selection rule is used to identify a sufficiently large displacement capable of leaving the current attraction basin. Importantly, the prescription mechanism is only a temporary intervention. It does not replace the underlying gradient descent dynamics: after the escape step, the algorithm immediately returns to the standard gradient update with the predefined step size. The resulting method therefore combines

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the simplicity of gradient descent with a deterministic mechanism designed specifically to overcome stagnation and local attraction.

2 Related works

2.1 Stochastic escape mechanisms

Non-convex optimization landscapes may contain local minima and saddle points that can cause gradient-based methods to stagnate or converge to suboptimal stationary points. Stochastic approaches address this issue by introducing randomness into the optimization process. Stochastic gradient methods have been shown to escape strict saddle points under suitable assumptions (Lee et al., 2016; Ge et al., 2015). Perturbed gradient descent further exploits explicit random perturbations to efficiently escape saddle regions (Jin et al., 2017). Other stochastic global optimization strategies, such as simulated annealing, use probabilistic transitions to explore regions separated by objective barriers (Kirkpatrick et al., 1983). Although effective for exploration, these approaches rely on stochastic perturbations, making the escape trajectory inherently non-deterministic.

2.2 Deterministic escape mechanisms

Deterministic approaches seek to overcome problematic stationary regions without relying on random perturbations.However, standard deterministic gradient descent can take exponential time to escape from specific saddle point geometries (Du et al., 2017). To accelerate this process, curvature-based methods, including trust-region and cubic-regularization approaches, exploit second-order information to identify suitable search directions (Conn et al., 2000; Nesterov & Polyak, 2006). Other deterministic global optimization approaches explicitly target local minima. Filled-function methods construct auxiliary functions designed to reduce the attraction of previously identified local minima (Ge, 1987), while tunneling methods introduce a dedicated phase to move from one local minimum toward another potentially better basin (Levy & Montalvo, 1985).

2.3 Positioning of the proposed approach

The proposed method belongs to the deterministic class but differs from these approaches in several respects. It is first-order, does not require explicit curvature information, and does not modify the objective function. Standard gradient descent remains the default optimization mechanism, while an escape intervention is activated only when stagnation is detected. The intervention temporarily prescribes a target value for the original objective and determines the corresponding step size along the current gradient direction. After this escape step, the algorithm immediately returns to standard gradient descent. The proposed method can therefore be characterized as a deterministic, first-order, objective-prescribed escape mechanism for overcoming local stagnation.

3 Proposed Method

3.1 Gradient descent and stagnation detection

Consider a non-convex scalar function $g(x)$. The classical gradient descent method is defined by

$$x_{k+1} = x_k - \eta_0 g'(x_k)$$

where $\eta_0 > 0$ denotes the fixed step size. When the trajectory enters a region where the displacement becomes very small, the algorithm may remain within the basin of attraction of a stationary point. We therefore consider that a stagnation condition is detected when

$$|x_k - x_{k-1}| < \epsilon_{\text{stag}}$$

Upon detecting such stagnation, the algorithm temporarily activates the objective prescription mechanism described in the following section. In the absence of stagnation, the update remains that of standard gradient descent.

3.2 OBJECTIVE-PRESCRIBED ESCAPE MECHANISM

We prescribe a desired evolution of the objective function according to

$$g(x_k) = g(x_0)f(k) \tag{1}$$

where $f(k)$ is a positive, strictly decreasing function. Consequently,

$$\frac{g(x_{k+1})}{g(x_k)} = \frac{f(k+1)}{f(k)} = r_k \tag{2}$$

with $0 < r_k < 1$. Thus, the prescribed evolution satisfies

$$g(x_{k+1}) = r_k g(x_k) \tag{3}$$

and therefore imposes a strictly decreasing sequence of objective values,

$$g(x_{k+1}) < g(x_k) \tag{4}$$

The prescribed objective evolution must simultaneously be compatible with a gradient-based update,

$$x_{k+1} = x_k - \eta_k g'(x_k) \tag{5}$$

The escape step is therefore obtained by solving for η_k the nonlinear scalar equation

$$g(x_k - \eta_k g'(x_k)) = r_k g(x_k) \tag{6}$$

This formulation transforms the escape problem into an inverse step-size determination problem: rather than prescribing the step size directly, the desired evolution of the objective function determines the admissible step sizes.

Depending on the geometry of g, the equation may have no solution, a single solution, or multiple solutions. When several solutions exist, they may correspond to substantially different displacements and may lead to different attraction basins. A deterministic selection criterion is therefore required to identify an appropriate escape step.

3.3 ESCAPE-STEP SELECTION AND COMPLETE ALGORITHM

For the prescribed dynamics, the escape step is obtained by solving

$$g(x_k - \eta_k g'(x_k)) - r_k g(x_k) = 0 \tag{7}$$

When several solutions are available, small displacements are first discarded according to

$$|\eta_j| > \eta_{\min} \tag{8}$$

Two deterministic selection strategies are then considered. Strategy 1 selects the largest admissible step,

$$\eta_k = \max_j \eta_j \tag{9}$$

while Strategy 2 evaluates the objective value associated with each admissible root and selects the one producing the lowest objective value,

$$\eta_k = \arg\min_{\eta_j} g(x_k - \eta_j g'(x_k)) \tag{10}$$

If no admissible root is found, the algorithm falls back to the standard step η_0. This fallback prevents the escape mechanism from interrupting the optimization process.

Once an escape step has been selected, the new iterate is computed as

$$x_{k+1} = x_k - \eta_k g'(x_k) \tag{11}$$

The algorithm then immediately returns to the standard gradient-descent mode, using the predefined step η_0. Consequently, the prescription mechanism acts only as a temporary escape intervention and does not replace the underlying gradient-descent dynamics.

3.4 ALTERNATIVE FORMULATION OF THE PRESCRIBED DYNAMICS

Taking the ratio of two consecutive objective increments gives:

$$q_k = \frac{f(k+1) - f(k)}{f(k) - f(k-1)} \tag{12}$$

which leads to

$$g(x_{k+1}) = g(x_k) + q_k(g(x_k) - g(x_{k-1})) \tag{13}$$

Consequently, the escape step can alternatively be determined by solving

$$g(x_k - \eta_k g'(x_k)) = g(x_k) + q_k(g(x_k) - g(x_{k-1})) \tag{14}$$

This formulation uses only the current and previous objective values and provides an alternative numerical realization of the prescribed dynamics.

4 EXPERIMENTAL RESULTS

4.1 EXPERIMENTAL SETUP

The proposed method is evaluated first on one-dimensional non-convex objective functions with multiple local minima and attraction basins. The standard gradient-descent step is fixed to $\eta_0 = 0.01$, while stagnation is detected when the displacement satisfies $|x_k - x_{k-1}| < \epsilon_{\text{stag}}$. When stagnation is detected, the prescribed escape step is obtained by solving the nonlinear equation

$$g(x_k - \eta_k g'(x_k)) - r_k g(x_k) = 0 \tag{15}$$

The resulting scalar equation is solved numerically using a bisection-based root-search procedure. The interval of admissible step sizes is initially defined over a prescribed range and is progressively subdivided to identify intervals in which a sign change occurs. Each detected sign-changing interval is then refined by bisection until the desired numerical precision is reached.

Since the prescription equation may admit multiple roots, the root-search procedure is applied over the complete prescribed interval rather than stopping after finding the first solution. The resulting set of admissible roots is subsequently filtered according to the minimum escape-step criterion and passed to the deterministic selection strategies described in Section 3.

If no admissible root is detected, the algorithm falls back to the standard step η_0, allowing the gradient-descent trajectory to continue.

Parameter	**Value**
Base step η_0	0.01
Stagnation threshold ϵ_{stag}	10^{-3}
Minimum escape step $\eta_{\min}$	1
Maximum iterations	300
Root search interval	$[-50, 50]$
Root search resolution	10000
Prescription function	$f(k) = e^{-\lambda k^p}, \lambda = 10^{-3}, p = 0.5$

Table 1: Experimental parameters.

4.2 ESCAPE FROM LOCAL MINIMA

Figure 1 illustrates the behavior of the proposed method on the non-convex function $g(x) = x^4 - 2x^2 - 0.5x$, which has two local minima. Starting from $x_0 = -1.57$, the standard gradient descent initially moves toward the local minimum located in the left basin. As the trajectory approaches this minimum, the displacement between consecutive iterates becomes sufficiently small and the stagnation criterion is triggered. The objective-prescription mechanism is then activated and generates an escape step that moves the trajectory out of the current attraction basin. After this

single intervention, the algorithm immediately returns to the standard gradient-descent mode and progressively converges toward the lower, global minimum.

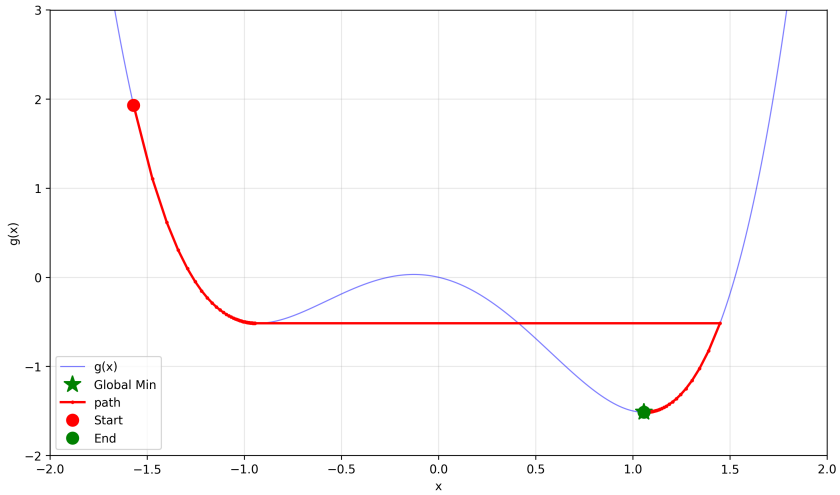

Figure 1: Behavior of the proposed method on the non-convex function $g(x) = x^4 - 2x^2 - 0.5x$.

Figure 2 shows the evolution of the step size η_k along the same optimization trajectory. During the normal optimization phase, the step size remains fixed at the prescribed value $\eta_0 = 0.01$. The isolated peak in η_k corresponds to the activation of the objective-prescription mechanism following the detection of stagnation. This large step produces the displacement required to leave the current attraction basin. Immediately after the escape step, the step size returns to η_0, illustrating the temporary nature of the proposed intervention. Thus, the figure clearly shows the hybrid behavior of the method: standard gradient descent is maintained throughout the trajectory, while a single large prescribed step is introduced only when stagnation occurs.

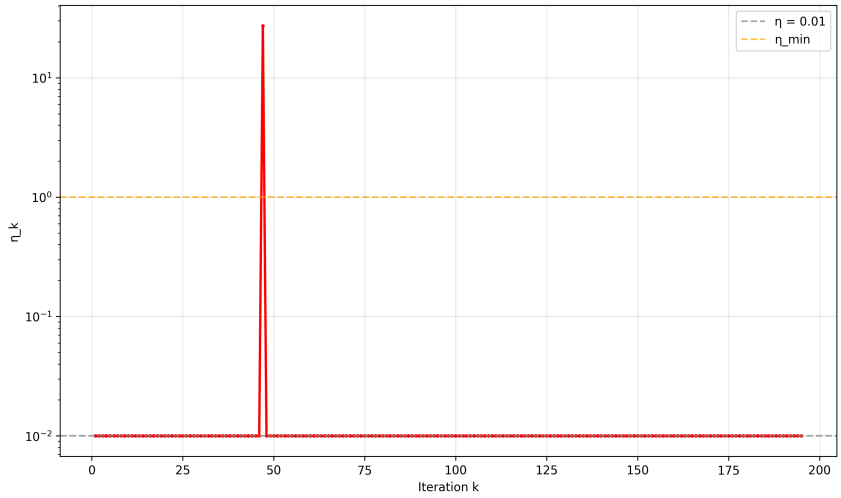

Figure 2: Illustration of the step evolution.

Together, Figures 1 and 2 provide a direct visualization of the proposed escape mechanism: the large displacement observed in the trajectory coincides with the isolated peak in the step-size evolution, confirming that the basin transition is produced by a temporary prescription step rather than by a permanent modification of the gradient-descent dynamics.

4.3 ROBUSTNESS WITH RESPECT TO THE INITIAL POINT

Table 2 summarizes the performance of the proposed method for a range of initial points $x_0 \in [-4, 4]$. The results show that the algorithm converges successfully for all tested initializations. In particular, all trajectories reach the same solution, $x^* = -1.514754$, with a final distance of zero from the reference minimum.

An interesting behavior can be observed in the number of prescription interventions. For initial points located in the left part of the domain, the algorithm requires a single escape intervention to reach the basin of the global minimum. In contrast, for initial points located closer to or inside the attraction basin of the global minimum, standard gradient descent is sufficient and no prescription step is required.

These results illustrate an important property of the proposed approach: the prescription mechanism is activated only when necessary. It does not systematically perturb the gradient-descent trajectory, but acts as a temporary escape mechanism when the current trajectory is trapped in an unfavorable attraction basin. Once the appropriate basin is reached, the algorithm proceeds exclusively with the standard gradient-descent step.

Overall, the experiments demonstrate a 100% convergence rate over the tested initial conditions, suggesting that the proposed escape mechanism is robust with respect to the initialization for this objective function.

x_0	x_{final}	$g(x_{\text{final}})$	**Distance**	**Prescriptions**	**Converge**
-4.0	1.057454	-1.514754	0.000000	1	True
-3.5	1.057454	-1.514754	0.000000	1	True
-3.0	1.057454	-1.514754	0.000000	1	True
-2.5	1.057454	-1.514754	0.000000	1	True
-2.0	1.057454	-1.514754	0.000000	1	True
-1.5	1.057454	-1.514754	0.000000	1	True
-1.0	1.057454	-1.514754	0.000000	1	True
-0.5	1.057454	-1.514754	0.000000	1	True
0.0	1.057454	-1.514754	0.000000	0	True
0.5	1.057454	-1.514754	0.000000	0	True
1.0	1.057454	-1.514754	0.000000	0	True
1.5	1.057454	-1.514754	0.000000	0	True
2.0	1.057454	-1.514754	0.000000	0	True
2.5	1.057454	-1.514754	0.000000	0	True
3.0	1.057454	-1.514754	0.000000	0	True
3.5	1.057454	-1.514754	0.000000	0	True
4.0	1.057454	-1.514754	0.000000	0	True

Table 2: Convergence and performance summary across a range of initial conditions.

4.4 COMPARISON WITH CLASSICAL GD

To evaluate the effectiveness of the proposed hybrid method, we compared it with classical gradient descent (GD) on the multimodal non-convex Rastrigin function:

$$g(x) = 10 + x^2 - 10\cos(2\pi x) \tag{16}$$

This function presents several local minima and a unique global minimum at $x^* = 0$, $g(x^*) = 0$. For both methods, the initial step size was fixed to $\eta_0 = 0.001$. The initial point x_0 was varied from -4 to 5 with a step of 0.01, resulting in 950 different initializations. Both algorithms were evaluated using the same convergence criteria and the same initial points.

The results are summarized in Table 3.

Method	Success	Success Rate	Mean Prescriptions
Classic GD	$101/950$	10.6%	NA
Proposed Method: OPE-GD	$950/950$	100.0%	1.67

Table 3: Performance comparison between classical GD and the proposed method on the Rastrigin function.

Classical gradient descent reaches the global minimum in only 101 out of the 950 tested initializations, corresponding to a success rate of 10.6%. For the remaining initial points, GD converges to one of the local minima surrounding the global minimum. This behavior illustrates the difficulty of standard gradient descent on highly multimodal non-convex objectives.

In contrast, the proposed hybrid method successfully reaches the global minimum for all 950 initializations, achieving a success rate of 100%. The escape mechanism is activated only when stagnation is detected and requires an average of 1.67 prescription interventions per trajectory. This confirms that the proposed mechanism does not continuously modify the standard gradient-descent dynamics, but instead provides temporary interventions to escape unfavorable attraction basins.

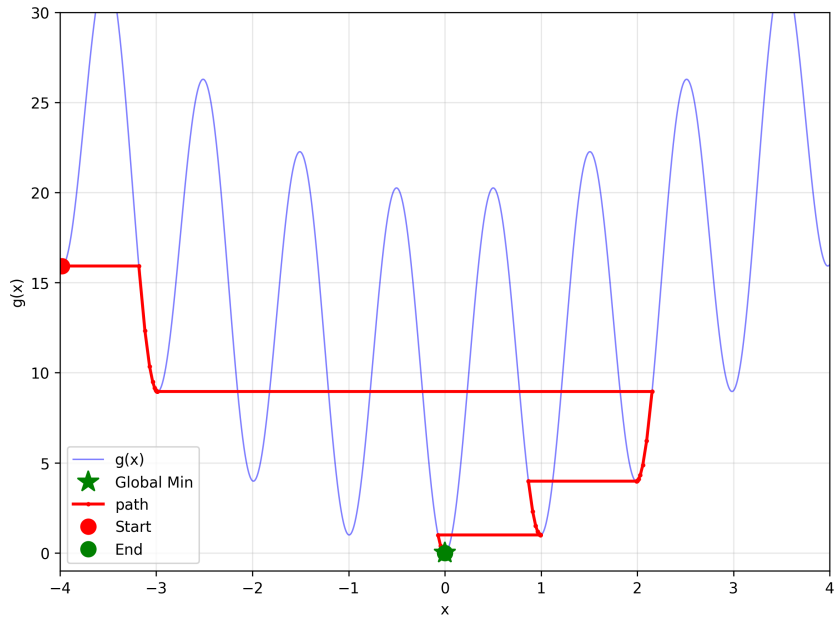

Figure 3: Illustration of the escape mechanism for Rastrigin function .

Figure 3 illustrates the behavior of the proposed method on the non-convex Rastrigin function g, which has multiple local minima. Starting from $x_0 = -3.98$, the algorithm activates the prescription mechanism 4 times to successfully reach the global minimum.

5 Conclusion

This work introduced a deterministic, first-order mechanism for escaping local attraction in non-convex optimization. The proposed approach preserves standard gradient descent as the default optimization mechanism and activates an objective-prescription mechanism only when stagnation is detected. The escape step is determined by prescribing a decreasing evolution of the objective function and solving the resulting nonlinear equation along the current gradient direction. When several admissible roots exist, deterministic selection strategies can be used to identify an effective escape step. After this temporary intervention, the algorithm immediately returns to standard gradient descent. Numerical experiments on a one-dimensional non-convex function with two local minima

show that the proposed mechanism can escape local attraction and reach the global minimum from a broad range of initial conditions. These results provide a first validation of the proposed approach. A natural extension of this work is to investigate the method in higher-dimensional non-convex optimization, and particularly in neural-network training, where the objective landscape presents substantially more complex geometries.

AI USE STATEMENT

In this work, we used generative AI tools for generating the core simulation and optimization code used in our experiments, as well as for formatting the manuscript text and mathematical formulas into LaTeX, converting raw experimental data into structured tables, and converting flat reference strings into BibTeX entries. We have not used generative AI tools for the original conceptual research ideas or the underlying theoretical development of the algorithm. Additionally, we used generative AI tools for academic translation and grammatical refinement of the manuscript to ensure professional English clarity. We have reviewed all AI-assisted work. Specifically, the AI-generated simulation code was fully verified, debugged, and tested for correctness by the author, and all typeset LaTeX components and equations were manually verified against the original formulations. We take responsibility for the final content of this work, including text, claims, code, or artifacts produced with the aid of generative AI.

REFERENCES

Andrew R Conn, Nicholas IM Gould, and Philippe L Toint. *Trust Region Methods*. SIAM, Philadelphia, 2000.

Simon S Du, Jason D Lee, Haochuan Li, Liping Wang, and Aarti Singh. Gradient descent can take exponential time to escape saddle points. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Renpu Ge. The theory of filled function methods for finding global minimizers of nonlinearly constrained minimization problems. *Journal of Computational Mathematics*, 5:1–9, 1987.

Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points—online stochastic gradient for tensor decomposition. In *Proceedings of the 28th Conference on Learning Theory (COLT)*, volume 40, pp. 797–842. PMLR, 2015.

Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M Kakade, and Michael I Jordan. How to escape saddle points efficiently. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70, pp. 1724–1732. PMLR, 2017.

Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

Jason D Lee, Max Simchowitz, Michael I Jordan, and Benjamin Recht. Gradient descent only converges to minimizers. In *Proceedings of the 29th Conference on Learning Theory (COLT)*, volume 49, pp. 1246–1257. PMLR, 2016.

A V Levy and A Montalvo. The tunneling algorithm for the global minimization of functions. *SIAM Journal on Scientific and Statistical Computing*, 6(1):15–29, 1985.

Yurii Nesterov and Boris T Polyak. Cubic regularization of newton method and its global performance. *Mathematical Programming*, 108(1):177–205, 2006.