OptiBridge: Multi-Scale Multi-Shift Bridging for Conditioning Optimization Landscapes

Farnaz Salehi Sadaghiani FSALEHI@UIC.EDU
Mojtaba Soltanalian MSOL@UIC.EDU
University of Illinois Chicago, USA

Abstract

This paper introduces OptiBridge, a novel optimization framework designed to tackle the challenges of complex, nonconvex landscapes that contain numerous local optima, as often encountered in NP-hard problems. OptiBridge employs a multi-scale, multi-shift strategy to enlarge the attraction basins of high-quality local minima while simultaneously reducing the influence of their low-quality counterparts. In doing so, it enhances the effectiveness of standard local optimization methods, such as gradient-based first-order algorithms. Through a series of experiments on benchmark nonconvex test functions, we demonstrate that OptiBridge consistently improves the performance of local optimizers.

1. Introduction

Non-convex optimization problems, particularly those classified as NP-hard, are typically difficult to solve due to their complex landscapes that contain numerous local optima. Traditional optimization methods often become trapped in these suboptimal points, struggling to reach the global optimum; thus yielding solutions of varying quality. In this paper, we introduce OptiBridge, a novel approach that systematically reshapes the optimization landscape by creating strategic “bridges” among distinct regions of the search space. This mechanism enhances the accessibility of high-quality solutions while reducing the influence of poor local optima, thereby improving the effectiveness of standard local optimization methods.

2. Background and Related Work

Local optimization methods converge to local optima of the objective function, which may or may not include the global optimum. In practice, the quality of these solutions can vary significantly. To evaluate the effectiveness of a local minimizer $\ell(\cdot)$ applied to a positive function $f(x) > 0$, we define the following figure of merit:

$$\mathcal{M}_\ell = \frac{1}{K} \sum_{k=1}^{K} \frac{f(x_k^0)}{f(\hat{x}_k^\ell)}, \tag{1}$$

where x_k^0 is the k-th initial point, $\hat{x}_k^\ell$ is the corresponding local minimum obtained by $\ell(\cdot)$, and K is the total number of trials. By construction, $\mathcal{M}_\ell \geq 1$. A larger value of $\mathcal{M}_\ell$ indicates a more effective local minimizer. This metric provides a quantitative measure of how well a local optimizer identifies high-quality regions of the search space.

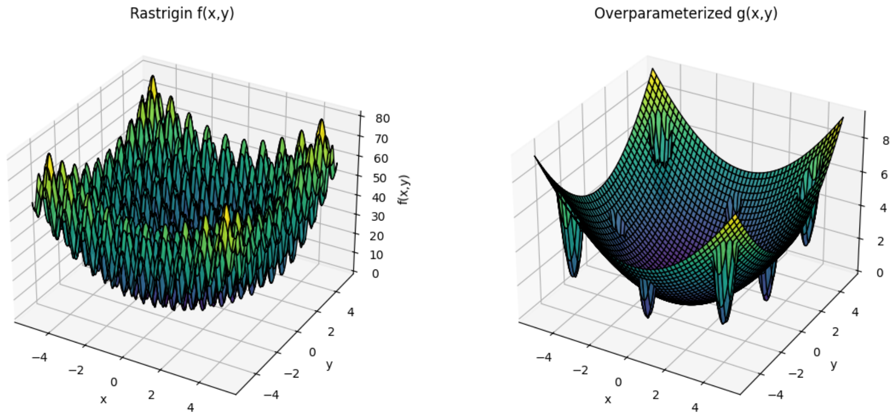

Figure 1: OptiBridge reduces poor local minima and enlarges the attraction domain of the global minimum.

It has long been recognized in mathematical optimization that introducing additional variables—i.e., judicious over-parametrization—can reduce the number of poor local optima and improve the navigability of the landscape. By expanding the parameter space, one can effectively "smooth out" certain traps in the optimization landscape, making high-quality optima more accessible [2, 7, 9, 15, 16, 18].

In practical applications, some global optima naturally possess large attraction domains. For example, in computational imaging and signal processing, applying first-order methods with relatively large initial step sizes can help algorithms land within these desirable regions [4]. Such observations highlight the potential advantage of intentionally reshaping or conditioning optimization landscapes so that better local optima dominate larger regions of the search space.

This motivates the development of methods that can systematically enlarge the attraction domains of high-quality optima while reducing the prevalence of poor ones. In this work, we introduce OptiBridge, a novel framework that leverages over-parametrization and multi-scale multi-shift transformations to achieve this goal in a principled manner, thereby enhancing the performance of conventional local optimizers. Figure 1 illustrates the impact of OptiBridge's over-parameterized objective $g(x, y)$ (defined in Section 3.1), when applied to the Rastrigin function [12], mitigating poor local minima and expanding the attraction domain of the global minimum at $(x, y) = (0, 0)$.

2.1. Relevant Ideas from the Past and Points of Tangency

OptiBridge shares conceptual similarities with tunneling and homotopy methods in optimization, as all aim to overcome the limitations of conventional techniques, particularly when navigating complex landscapes with many local optima.

Tunneling Methods provide a mechanism to escape local optima by creating a "tunnel" through the optimization landscape. Once a local optimum is reached, these methods modify the objective function or introduce auxiliary functions to enable the search to explore other promising regions, effectively bypassing traps in the landscape [1, 5, 14, 19].

Homotopy Optimization constructs a continuous transformation (homotopy) between an easy-to-solve problem and the original, more difficult problem. By gradually morphing the simpler problem into the target problem, the solution can be traced from the former to the latter. Notable work in this area includes [17], which applies homotopy methods to solve systems of nonlinear equations—a core challenge in optimization. Additional investigations of this concept can be found in [3, 8, 10, 13].

Novel Contributions of OptiBridge: While tunneling and homotopy methods have advanced the navigation of complex landscapes, OptiBridge introduces a novel multi-dimensional bridging framework that facilitates both escaping from suboptimal local optima and systematically enlarging the attraction basins of superior optima. Unlike tunneling, which modifies the landscape locally, OptiBridge adjusts the landscape globally to enhance overall navigability. In contrast to homotopy methods, which rely on tracing continuous transformation paths, OptiBridge employs a multi-scale, multi-shift strategy that dynamically reshapes the landscape and creates an equivalent optimization problem rather than an approximation. This unique mechanism directly conditions the landscape for more efficient exploration of high-quality local optima.

3. Theoretical Framework and Methodology

OptiBridge introduces a novel conceptual framework for optimization that strategically leverages over-parameterization. In this context, over-parameterization refers to expanding the parameter space beyond the minimal requirements to define the optimization problem. This expansion creates a richer, more navigable landscape, inherently capable of bypassing suboptimal local optima that often trap conventional optimization methods.

To illustrate the OptiBridge objective, consider minimizing a function $f(x)$ characterized by numerous local optima, many of which are poor in quality. We introduce an over-parameterization strategy via a new function:

$$g(x,t) = tf(x) + (1-t)f(-x), \quad 0 \le t \le 1, \tag{2}$$

which effectively creates a “bridge,” “tunnel,” or “wormhole” in the optimization landscape, connecting the attraction domain around $f(x)$ with that of $f(-x)$ through the additional parameter t.

Let x^* be a local minimum of $f(x)$. If $f(-x^*) < f(x^*)$, traditional optimization methods starting at x^* would converge to x^* due to its local optimality. However, by optimizing $g(x,t)$ using a first-order method, the algorithm is encouraged to explore beyond x^*, leveraging the bridge created by g. This allows the method to navigate toward better solutions, including regions near $-x^*$, thereby reducing the likelihood of getting stuck in a lower-quality local optimum and improving overall solution quality.

3.1. OptiBridge Algorithm

The OptiBridge approach is summarized in Algorithm 1. For clarity of presentation, the shifts are omitted. The explanation of the algorithm is given below.

OptiBridge Objective Function: The essence of OptiBridge lies in its multi-scale, multi-shift optimization objective, defined as

$$g(x) = \sum_{k,l} w_{k,l} f(\eta_k x + \delta_l), \tag{3}$$

where the scales $\{\eta_k\}$ and shifts $\{\delta_l\}$ are embedded in the objective function, and the weights satisfy $\sum_{k,l} w_{k,l} = 1$ with $w_{k,l} \geq 0$ for all k and l. The OptiBridge method employs this objective function and minimizes it using conventional first-order gradient-based methods such as Gradient Descent.

The OptiBridge framework introduces a new way of optimization through bridging attraction domains: instead of applying a first-order method (or other standard optimization techniques) to $f(x)$, the optimization process is implicitly guided to achieve an outcome that improves upon the minimum among $\{f(\eta_k x + \delta_l)\}$. In doing so, OptiBridge not only increases the likelihood of escaping poor local optima—similar in spirit to the benefits of over-parameterization—but also steers the optimization trajectory toward regions of the search space that are more likely to contain the global optimum or high-quality local optima.

Selection of Weights: One useful approach in defining the weights is the adoption of a softmin formulation

$$w_{ij} = \frac{\exp(-\lambda f(\eta_i x + \delta_j))}{\sum_{k,l} \exp(-\lambda f(\eta_k x + \delta_l))}, \quad \lambda > 0, \tag{4}$$

then points with lower function values $f(\eta_i x + \delta_j)$ receive larger weights. By gradually increasing λ over the iterations, the weighting scheme transitions from a diffuse distribution to a near-binary selection. Specifically, when λ is small, the weights are spread across multiple scales and shifts, promoting exploration. As $\lambda \to \infty$, the weights concentrate on the minimizing pair (η_i, δ_j), effectively selecting the scale and shift that yield the smallest value of $f(\eta_i x + \delta_j)$. Although the softmin weights are not theoretically optimal, we observed that they perform well in practice.

Selection of Shifts and Scales: The choice of shifts $\{\delta_l\}$ and scales $\{\eta_k\}$ plays a pivotal role in defining both the granularity and the extent of search space exploration. For a general function $f(x)$ with no prior structural knowledge, a principled strategy is to draw inspiration from multi-resolution analysis and minimum-redundancy sampling. To this end, we introduce chirp-inspired shifts and scales, employing exponential spacing to balance coverage and efficiency. Specifically, one can define

$$\eta_k = 2^{k-1}\eta_1, \quad \delta_l = 2^{l-1}\delta_1. \tag{5}$$

This exponential scaling equips the optimizer with the ability to probe the landscape at multiple resolutions, thereby capturing both fine and coarse structural features.

Algorithm 1: OptiBridge Algorithm

Data: Objective function f, gradient ∇f, initial point x_0, scales $\{\eta_k\}$, step size α, maximum iterations K, initial and final softmin parameters $\lambda_{\text{start}}, \lambda_{\text{end}}$, tolerance tol

Result: Optimal point x and function value $f(x)$

$x \leftarrow x_0$;
for $i = 1$ **to** K **do**
  $\lambda \leftarrow \lambda_{\text{start}} + i \cdot (\lambda_{\text{end}} - \lambda_{\text{start}})/K$;
  $w_j \leftarrow \exp(-\lambda f(\eta_j \cdot x)) / \sum_{k=1}^{m} \exp(-\lambda f(\eta_k \cdot x))$
  $grad \leftarrow \sum_{j=1}^{m} w_j \cdot \eta_j \cdot \nabla f(\eta_j \cdot x)$;
  if $\|grad\| < tol$ **then**
    break;
  end
  $x \leftarrow x - \alpha \cdot grad$;
end
return $x, f(x)$;

4. Illustrative and Experimental Results

This section presents a comparison of the performance of OptiBridge with the standard Gradient Descent method on several benchmark nonconvex problems. The results highlight the superior ability of OptiBridge to identify high-quality solutions. We used scales $= \{0.25, 0.5, 1, 2\}$, shifts $= \{0\}$, step size $\alpha = 0.001$, maximum iterations $K = 5000$, $\lambda_{\text{start}} = 1$, $\lambda_{\text{end}} = 20$, and tolerance tol $= 10^{-8}$. We generated random initial points x_0 using a fixed random seed of 42. The results are summarized in Table 1. In the table, n denotes the number of variables, and $f(x^*)$ represents the global optimal value. **OB Value** is the optimal value obtained by the OptiBridge method, and **OB Time** is its computation time in seconds. **GD Value** refers to the optimal value obtained by the Gradient Descent method, with **GD Time** denoting its computation time in seconds. **Best GD Value** is the smallest value achieved by Gradient Descent over 100 different initial points x_0, while **100 GD Time** corresponds to the total computation time for these 100 runs. The results show that the OptiBridge method consistently reached the global optimum or a value very close to it, whereas in most cases Gradient Descent failed to attain the global optimal value, even after 100 runs.

Table 1: Comparison of OptiBridge and Gradient Descent on benchmark functions

Problem	n	$f(x^*)$	OB Value	OB Time (s)	GD Value	GD Time (s)	Best GD Value	100 GD Time (s)
Rastrigin [12]	30	0	0	0.006	268.637	0.002	177.102	0.148
Griewank [12]	30	0	0	0.360	869.995	1.443	600.971	100.596
Bohachevsky [6]	10	0	0	0.417	11.917	0.467	10.512	33.494
Mccormick [6]	2	-1.913	-1.913	0.804	1.228	0.114	-1.913	14.020
Ackley [12]	30	0	5.927e-05	2.194	19.583	0.297	18.955	30.980
Drop Wave [12]	30	-1	-0.880	17.365	-0.016	3.913	-0.023	186.618
Salomon [6]	30	0	5.022e-06	1.019	1.600	0.013	1.300	1.463
Alpine [6]	30	0	0	14.754	0.388	2.561	0.228	124.367
Three Hump Camel [6]	2	0	1.763e-17	0.076	0.299	0.182	2.960e-09	14.875
Egg Crate [6]	2	0	4.261e-19	0.062	9.488	0.026	8.666e-19	2.334
Goldstein Price [6]	2	3	3	0.045	956600	0.626	3	45.596
Cross in Tray [6]	2	-2.063	-2.062	2.952	-1.715	0.709	-2.063	42.697
Levi N.13 [11]	2	0	2.356e-15	2.231	8.187	0.361	2.516	38.055

5. Conclusion

OptiBridge introduces a transformative approach to optimization, substantially enlarging the attraction basins of desirable local optima and diminishing the effect of suboptimal solutions. Through its multi-scale, multi-shift methodology, it provides a robust and flexible framework for efficiently navigating the complex landscapes of non-convex problems, offering promising avenues for both theoretical advancement and practical applications.

References

[1] Bikram Adhikary and Md Abu Talhamainuddin Ansary. A tunneling method for nonlinear multi-objective optimization problems. *arXiv preprint arXiv:2407.04436*, 2024.

[2] Simon Du and Jason Lee. On the power of over-parametrization in neural networks with quadratic activation. In *International conference on machine learning*, pages 1329–1338. PMLR, 2018.

[3] Daniel M Dunlavy and Dianne P O'Leary. Homotopy optimization methods for global optimization. Technical report, Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA, 2005.

[4] Sara Fridovich-Keil, Fabrizio Valdivia, Gordon Wetzstein, Benjamin Recht, and Mahdi Soltanolkotabi. Gradient descent provably solves nonlinear tomographic reconstruction. *ArXiv*, pages arXiv–2310, 2023.

[5] Patrick JF Groenen and Willem J Heiser. The tunneling method for global optimization in multidimensional scaling. *Psychometrika*, 61(3):529–550, 1996.

[6] Momin Jamil and Xin-She Yang. A literature survey of benchmark functions for global optimisation problems. *International Journal of Mathematical Modelling and Numerical Optimisation*, 4(2):150–194, 2013.

[7] Kenji Kawaguchi. Deep learning without poor local minima. *Advances in neural information processing systems*, 29, 2016.

[8] Andreas Klingler and Tim Netzer. Homotopy methods for convex optimization. *SIAM Journal on Optimization*, 35(3):1498–1523, 2025.

[9] Dawei Li, Tian Ding, and Ruoyu Sun. Over-parameterized deep neural networks have no strict local minima for any continuous activations. *arXiv preprint arXiv:1812.11039*, 2018.

[10] Xi Lin, Zhiyuan Yang, Xiaoyuan Zhang, and Qingfu Zhang. Continuation path learning for homotopy optimization. In *International Conference on Machine Learning*, pages 21288–21311. PMLR, 2023.

[11] Sudhanshu Mishra. Some new test functions for global optimization and performance of repulsive particle swarm method. 2006.

[12] Marcin Molga and Czesław Smutnicki. Test functions for optimization needs. *Test functions for optimization needs*, 101(48):32, 2005.

[13] Shayan Pardis, Matthew Chignoli, and Sangbae Kim. Probabilistic homotopy optimization for dynamic motion planning. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 12856–12863. IEEE, 2024.

[14] Seung Park, Kyunghyun Baek, Seungjin Lee, and Mahn-Soo Choi. Global optimization in variational quantum algorithms via dynamic tunneling method. *New Journal of Physics*, 26 (7):073053, 2024.

[15] Itay Safran and Ohad Shamir. Spurious local minima are common in two-layer relu neural networks. In *International conference on machine learning*, pages 4433–4441. PMLR, 2018.

[16] Itay M Safran, Gilad Yehudai, and Ohad Shamir. The effects of mild over-parameterization on the optimization landscape of shallow relu neural networks. In *Conference on Learning Theory*, pages 3889–3934. PMLR, 2021.

[17] Layne T Watson. Globally convergent homotopy algorithms for nonlinear systems of equations. *Nonlinear Dynamics*, 1(2):143–191, 1990.

[18] Ji Xu, Daniel J Hsu, and Arian Maleki. Benefits of over-parameterization with EM. *Advances in Neural Information Processing Systems*, 31, 2018.

[19] Yong Yao. Dynamic tunneling algorithm for global optimization. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5):1222–1230, 1989.