Towards an Algorithm-Agnostic Framework for Mitigating Ill-Conditioning in Black-Box Optimization via PCA Whitening

Chengjie Zhang[1,3], Bo Zhang[2], Lin Wang[1,3†], Bo Yang[3,1]
[1]Shandong Key Laboratory of Ubiquitous Intelligent Computing, University of Jinan, Jinan 250022, China
[2]Shandong Provincial Key Laboratory of Green and Intelligent Building Materials, University of Jinan, Jinan 250022, China
[3]Quan Cheng Laboratory, Jinan 250100, China
[†]Corresponding author: wangplanet@gmail.com

***Abstract*—Ill-conditioned problems present formidable challenges to search efficiency due to severe scale imbalances. Most existing mitigation strategies rely on embedding adaptive mechanisms within specific algorithms, which often leads to tight algorithmic coupling and restricted generalizability across diverse optimization paradigms. To overcome these limitations, this paper introduces a general, algorithm-agnostic geometric correction framework. By integrating Random Perturbation–Based Sample Augmentation with PCA whitening, the framework fundamentally reparameterizes narrow, steep-walled "valleys" into isotropic, spherical-like latent spaces that are significantly more amenable to optimization. Designed as a "plug-and-play" external module, the proposed framework enables seamless integration without modifying the core search strategy of the underlying optimizers. Extensive evaluations on ten high-condition-number BBOB benchmark functions demonstrate that the framework consistently and significantly bolsters the performance of multiple representative optimization algorithms on ill-conditioned landscapes, markedly accelerating convergence while enhancing the quality of final solutions.**

***Index Terms*—Black-Box Optimization, Ill-Conditioned Problems, Geometric Correction, PCA Whitening**

I. Introduction

Optimization aims to find the global optimum for a given problem. In many real-world applications—such as aerodynamic shape design [1], [2], or complex engineering simulations [3]–[5]—the problem f often exhibits black-box characteristics, where the analytical form of f is unknown and only input-output evaluations are available [6], [7]. Mathematically, a continuous black-box optimization (BBO) problem is described as:

$$\min_{x \in \Omega} f(x), \quad \Omega \subseteq \mathbb{R}^d, \tag{1}$$

where d denotes the dimensionality of the decision space. Lacking explicit gradient or structural information, BBO efficiency hinges on using zero-order sampling to adapt to the latent geometric structure—a task that becomes increasingly difficult in complex landscapes.

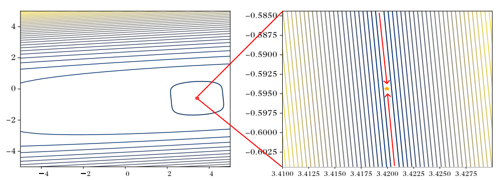

Fig. 1. Visualization of an ill-conditioned objective landscape at global (left) and local (right) scales. To reach the optimum ($\star$), the optimizer must precisely navigate along the longitudinal axis of the steep, narrow valley.

As a class of difficult BBO problems, ill-conditioned problems [8] exhibit an extreme sensitivity to specific directions in the search space, where minute perturbations in certain decision variables can lead to disproportionately large fluctuations in the objective value $f(x)$. Mathematically, the conditioning of a problem is characterized by the Hessian matrix $H(x) = \nabla^2 f(x)$, which captures the local curvature of the objective landscape. A problem is termed "ill-conditioned" when the condition number—the ratio of the largest to the smallest eigenvalue of the Hessian ($\kappa = \lambda_{max}/\lambda_{min}$)—is extremely large. Geometrically, such a high condition number manifests as a landscape dominated by extremely narrow and steep "valleys" (as illustrated in Figure 1). These features are often not aligned with the coordinate axes, implying a high degree of inter-dependency between decision variables.

In such anisotropic terrains, conventional optimizers face a severe dilemma regarding step-size control. In high-curvature (sensitive) directions, large step sizes are catastrophic; they cause the optimizer to overshoot the narrow valley's floor,

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61872419, 62072213, and 62403209; in part by the Shandong Provincial Natural Science Foundation under Grant Nos. ZR2022JQ30, ZR2022ZD01, ZR2023LZH015, and ZR2024QF021; in part by the Shandong Provincial Key R&D Program Project under Grant Nos. 2024CXPT084 and 2025CXPT105; in part by the Key Research Project of Quan Cheng Laboratory, China under Grant Nos. QCL20250105 and QCL20250304; in part by the Open Fund of the State Key Laboratory of Silicate Materials for Architectures under Grant No. SYSJJ2025-02; in part by the New 20 Measures for Colleges and Universities of Jinan under Grant No. 202534127; in part by the Youth Innovation Team Program of Shandong Higher Education Institutions, China under Grant No. 2024KJH104; and in part by the University of Jinan Young Faculty Interdisciplinary Convergence Development Project 2025 under Grant No. XKJC-202507.

leading to violent oscillations or even total divergence. While a small step size ensures stability in sensitive regions, it significantly restricts the exploration rate along insensitive directions, failing to traverse the long axis of the landscape efficiently and resulting in glacially slow convergence [9], [10]. Therefore, to reach the global optimum, the optimizer must prudently adjust its search steps to accurately navigate a narrow, curved parabolic valley.

To mitigate the detrimental effects of ill-conditioning, several strategies have been proposed. One prominent approach is internal adaptive mechanisms within the optimizer to counteract scale imbalance. A quintessential example is Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [10], [11], which iteratively learns the structure of the problem by updating a covariance matrix. By aligning its search distribution with the landscape's local curvature, CMA-ES effectively "decorrelates" the decision variables. However, such adaptive mechanisms are typically tightly coupled with the optimizer's internal logic, making them difficult to migrate or "plug-and-play" across diverse optimization frameworks.

Recently, the landscape transformation methods seek to "rectify" the problem at its source by fundamentally reshaping the objective landscape. Through space mapping or warping, these techniques project narrow, steep-walled "valleys" into an isotropic, spherical-like latent space that is significantly easier to navigate. This problem-centric paradigm addresses the intrinsic ill-conditioning at its root, offering a versatile and algorithm-agnostic pre-conditioning technique. Unfortunately, current research remains largely confined to Surrogate-Assisted Optimization (SAO), where it is typically inextricably linked to specific models [12], [13]. This strict coupling hinders its generalizability and prevents its adoption as a standalone, "plug-and-play" component for arbitrary black-box optimizers.

In this paper, we propose a general framework for ill-conditioned optimization. By leveraging random perturbation and PCA whitening techniques, the framework effectively transforms narrow, ill-conditioned valleys into isotropic, spherical-like landscapes that are significantly more amenable to optimization. As a model-agnostic and plug-and-play solution, our approach is compatible with arbitrary optimizers without altering their core search logic. We validate the framework across several ill-conditioned benchmark problems using diverse optimizers. Experimental results demonstrate that our method consistently enhances the performance of various optimizers, achieving both accelerated convergence and superior final solution quality.

II. Related Work

A. Algorithm-Internal Adaptation

Existing studies on ill-conditioned black-box optimization have mainly focused on enhancing the adaptability of algorithms to high condition numbers and rotated landscapes by adjusting internal parameters, adaptive mechanisms, or update dynamics. In differential evolution (DE), JADE [14] improves stability and convergence on functions with ill-conditioned and narrow-valley structures via success-history based parameter adaptation and direction-guided mutation. Further, Li et al. [15] proposed a PCA-based crossover that periodically estimates principal directions from the population covariance and performs recombination in the PCA coordinate system, thereby alleviating the directional mismatch of coordinate-wise operators in elongated valleys.

For particle swarm optimization (PSO), a large body of work enhances the expressiveness of direction and correlation by modifying the velocity update dynamics, such as introducing covariance or correlation accumulation mechanisms [16], or constructing and selecting candidate directions in the velocity space to mitigate direction collapse [17].

For estimation of distribution algorithms (EDA), extensive studies have investigated adaptive variance scaling and mean-shift correction from the perspective of matching the search distribution to the objective geometry [18]–[23]. CMA-ES [10] learns a full covariance matrix to gradually align the sampling distribution with the level sets of the objective function, which is a classical mechanism for coping with anisotropy and rotation. Subsequent studies further improve its reliability under strong ill-conditioning and high dimensionality by stabilizing covariance learning or introducing subspace and coordinate-wise updates [24]–[26].

B. Landscape Transformation

Different from the above algorithm-internal adaptation paradigm, a few works have started to explore explicit transformation of the search space to mitigate ill-conditioning. LABCAT [12] combines rescaling and PCA alignment within a trust-region Bayesian optimization framework to make local search spaces more isotropic, but the transformation is tightly coupled with surrogate modeling and trust-region management, limiting its reusability as an independent module. SACOBRA [13] constructs an online whitening mapping in the neighborhood of the current best solution to transform high-condition local landscapes into approximately isotropic spaces before surrogate modeling and search. Although conceptually aligned with the idea of reshaping narrow valleys by transformation, this approach remains strongly bound to the surrogate framework, often requires additional function evaluations, and may suffer from a mismatch between the transformation region and the dynamically shifting search focus.

Motivation Mainstream strategies typically mitigate ill-conditioning through internal adaptation, such as distribution learning or directional enhancement. While explicit geometric transformation offers a theoretically more direct remedy, existing methods are often hindered by rigid coupling with surrogate frameworks and high budgets. These limitations motivate the present work to develop a simple, algorithm-agnostic geometric correction framework. Crucially, our goal is not to replace specialized optimizer (e.g., CMA-ES), but to provide a portable, modular 'upgrade' that empowers a wide variety of general-purpose optimizers to robustly handle ill-conditioned landscapes without altering their core search logic.

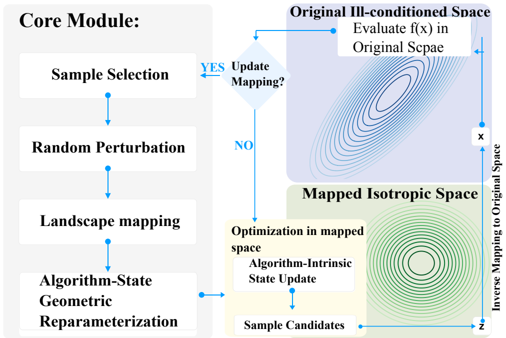

Fig. 2. The proposed landscape transformation framework.

III. Methodology

This section details the proposed algorithm-agnostic geometric correction framework, which comprises three primary components: data preparation, PCA whitening, and geometry-consistent state reparameterization. The data preparation module is responsible for sample selection and random-perturbation-based augmentation, establishing a robust statistical foundation for the subsequent mapping. Building upon this, the PCA whitening component executes the core geometric reparameterization, transforming the ill-conditioned search landscape into an approximately isotropic latent space $\mathcal{Z}$. Finally, the framework transitions the optimization process to this transformed space by reparameterizing the internal states of the optimizer in a geometry-consistent manner, thereby synchronizing its search history with the rectified landscape to ensure continuity. The overall procedure is illustrated in Fig. 2.

A. Preparation for Affine Mapping Estimation

To accurately capture the local geometric structure of the current search region, we employ an archive-based sample selection scheme. At each mapping update, the $K \cdot d$ most recently evaluated samples are extracted from the solution archive to serve as training data, where d represents the problem dimensionality and K is a tunable hyperparameter. This sliding-window approach ensures that the learned transformation remains synchronized with the evolving search dynamics while maintaining a critical balance between temporal locality and sample sufficiency.

Considering numerical instability in covariance estimation—particularly under finite sample sizes or extreme anisotropy-a random perturbation-based sample augmentation strategy is introduced. Specifically, for each selected sample $\tilde{x}_i \in \mathcal{X}$ from the archive, an augmented counterpart x_i is generated by applying a stochastic additive perturbation:

$$x_i = \tilde{x}_i + \sigma(t) \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \tag{2}$$

where ϵ is a d-dimensional noise vector drawn from a standard multivariate normal distribution, representing the direction and relative scale of the random walk.

To ensure that the geometric correction remains adaptive to the shifting search focus while maintaining high precision near the optimum, the perturbation magnitude follows an exponential decay schedule relative to the search progress, formulated as:

$$\sigma(t) = \sigma_0 \exp(-\kappa t), \quad t = \frac{\text{FEs}}{\text{MaxFEs}} \in [0, 1], \tag{3}$$

where σ_0 and κ denote the initial perturbation strength and the decay rate, respectively.

By enriching the sample distribution with this additive noise, the framework effectively regularizes the estimated covariance structure and reduces the risk of mapping degeneration. Crucially, this augmentation is performed solely within the mapping learning module; as these perturbed samples are not evaluated via the objective function, the strategy incurs no additional function evaluation cost.

B. Construction of the PCA Whitening Mapping

Following sample selection and stochastic augmentation, a PCA whitening mapping is employed for local geometric reparameterization, owing to its capacity to simultaneously decorrelate variables and normalize directional scales, effectively counteracting the anisotropy inherent in high-condition-number landscapes.

Let $\mathcal{X} = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ represent the augmented training set. The sample mean μ and empirical covariance Σ are calculated as:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^\top. \tag{4}$$

An eigen-decomposition $\Sigma = U \Lambda U^\top$ is performed , where U denotes the matrix of orthogonal eigenvectors and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$ is the diagonal matrix of eigenvalues sorted in descending order. Based on this decomposition, the PCA whitening mapping is defined as the affine transformation:

$$z = Tx + b, \tag{5}$$

where $T = \Lambda^{-\frac{1}{2}} U^\top$ and $b = -T\mu$.

Geometrically, this transformation corresponds to rotating the original coordinate system to align with the principal curvature directions of the local landscape, followed by normalizing the scales along each axis using the reciprocal square root of the eigenvalues. Consequently, the highly anisotropic, narrow elliptical "valleys" characteristic of ill-conditioned problems are effectively reshaped into isotropic, spherical-like latent spaces. This reparameterization mechanism eliminates the linear coupling between decision variables, enabling the optimizer to explore the search space efficiently with a unified step size. By addressing the underlying geometry, this framework mitigates the inherent trade-off between search precision and convergence speed at its source.

TABLE I

MAPPING COMPARISON ON $D = 10$ (MEAN OVER 15 RUNS, $\alpha = 0.05$). FOR EACH ALGORITHM, THE BETTER MEAN RESULT BETWEEN *Base* AND *Ours* IS HIGHLIGHTED IN BOLD. THE SYMBOLS $(+)$, $(\approx)$, AND $(-)$ INDICATE THAT *Ours* IS STATISTICALLY SIGNIFICANTLY BETTER, NOT SIGNIFICANTLY DIFFERENT, OR SIGNIFICANTLY WORSE THAN *Base*, RESPECTIVELY, ACCORDING TO THE MANN–WHITNEY U TEST.

Func	(μ, λ)-ES		DE		PSO		OPT-GAN		UMDA	
	Base	Ours	Base	Ours	Base	Ours	Base	Ours	Base	Ours
f6	4.41e-14	**3.43e-14** ($\approx$)	1.28e-14	**1.02e-14** ($\approx$)	2.99e+00	**2.11e-01 (+)**	5.24e-02	**1.42e-04 (+)**	2.70e+01	**6.93e+00 (+)**
f7	4.13e+00	**1.16e+00 (+)**	0.00e+00	0.00e+00 ($\approx$)	4.62e-01	**8.59e-02 (+)**	**4.05e+00**	1.38e+01 (-)	5.96e+00	**1.38e+00 (+)**
f8	2.17e-03	**7.15e-07 (+)**	**1.41e-02**	1.07e+00 (-)	**6.46e+00**	6.78e+00 ($\approx$)	4.02e+00	**5.32e-01 (+)**	1.10e+01	**8.46e+00** ($\approx$)
f9	2.68e-01	**7.35e-07 (+)**	9.14e-01	**7.25e-01** ($\approx$)	8.93e+00	**6.08e+00 (+)**	5.84e-01	**3.32e-06 (+)**	1.00e+01	**8.44e+00 (+)**
f10	1.40e+02	**6.24e-01 (+)**	0.00e+00	0.00e+00 ($\approx$)	1.46e+03	**2.25e+01 (+)**	1.67e+02	**2.99e+00 (+)**	2.29e+04	**2.01e+03 (+)**
f11	6.79e+01	**4.94e-02 (+)**	1.43e-14	**0.00e+00** ($\approx$)	1.32e+00	**3.63e-02 (+)**	4.11e-01	**8.60e-02 (+)**	5.43e+01	**4.55e+01** ($\approx$)
f12	1.15e+01	**8.44e+00** ($\approx$)	3.10e-02	**2.40e-03 (+)**	6.29e+03	**3.45e+03** ($\approx$)	1.18e+03	**2.07e+00 (+)**	3.35e+04	**2.03e+00** ($\approx$)
f13	**3.04e+00**	3.70e+00 ($\approx$)	2.36e-10	**1.86e-10 (+)**	3.26e+01	**1.06e+01 (+)**	1.36e+01	**1.55e+00 (+)**	4.82e+01	**3.84e+00 (+)**
f14	**3.36e-05**	3.42e-05 ($\approx$)	3.65e-09	**0.00e+00** ($\approx$)	1.27e-02	**1.05e-02 (+)**	3.37e-03	**4.46e-06 (+)**	1.55e-01	**2.43e-03 (+)**
f18	3.65e-01	**1.82e-01** ($\approx$)	3.65e-07	**3.70e-08 (+)**	1.47e+00	**9.93e-01** ($\approx$)	**2.48e+00**	2.65e+00 ($\approx$)	4.59e-01	**6.03e-02 (+)**
+ / $\approx$ / -		**5/5/0**		**3/6/1**		**7/3/0**		**8/1/1**		**7/3/0**

C. Optimization in Mapping Space

Upon establishing the affine mapping $z = Tx + b$, the optimization process is effectively transitioned from the original decision space $\mathcal{X}$ to the reparameterized latent space $\mathcal{Z}$. In this transformed domain, the optimizer navigates a rectified, isotropic version of the objective landscape, which fundamentally simplifies the search dynamics. Consequently, to ensure the optimizer operates correctly within this new coordinate system, its internal state variables must undergo a geometry-consistent reparameterization. This procedure is essential to synchronize the algorithm's historical search information and current population with the updated mapping, thereby maintaining search continuity across periodic updates. To evaluate the generalizability of the proposed framework across diverse optimization paradigms, this study considers five representative baseline optimizers: DE [27], PSO [28], (μ, λ)-ES [10], [11], UMDA [29], and the GAN-based estimation of distribution algorithm-OPT-GAN [30].

To ensure seamless operation within the transformed latent space, the internal states of each optimizer are reparameterized derived from Eq. (5). Essentially, state variables denoting absolute positions undergo the full affine transformation, while those representing relative directions or velocities are mapped via the linear operator T. The specific implementations for diverse optimization paradigms are as follows:

- **DE** [27]: all individuals in the population are transformed by $z = Tx + b$, after which the original mutation, crossover, and selection operators are executed in the mapped space.
- **PSO** [28]: particle positions are transformed by $z = Tx + b$, and velocity vectors are linearly mapped by $v_{\text{mapped}} = Tv$, while all velocity update equations remain unchanged.
- (μ, λ)**-ES** [10], [11]: the distribution mean is mapped by $m_{\text{mapped}} = Tm + b$, the evolution paths and accumulated direction vectors are mapped by $v_{\text{mapped}} = Tv$, and the global step size is rescaled as $\sigma_{\text{mapped}} = \sigma \cdot \lambda_{\max}^{-1/2}$.
- **OPT-GAN** [30]: both maintained samples and newly generated candidates are transformed by $z = Tx + b$, and generative model training and sampling are continued in the mapped space.
- **UMDA** [29]: individual positions are transformed by $z = Tx + b$, and the diagonal variance matrix $D = \text{diag}(s_1^2, \ldots, s_d^2)$ is reparameterized as $D_{\text{mapped}} = \text{diag}(T^\top DT)$.

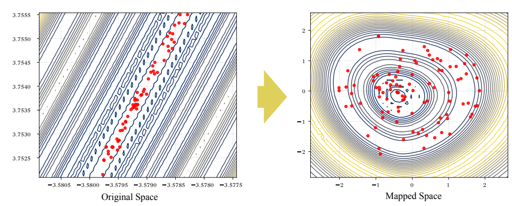

Fig. 3. Illustration of landscape geometry before and after local geometric reparameterization.

These implementations serve as illustrative examples of the general state synchronization process, which can be readily extended to other optimization paradigms. Note that the framework does not interfere with the optimizer's core search strategy or update equations. While specific state variables (e.g., velocity in PSO, covariance in UMDA) require mathematical transformation to align with the whitened space, the generative logic of the underlying algorithm remains completely independent and unmodified.

IV. EXPERIMENTS

A. Experimental Setup

To evaluate the effectiveness of framework, we select ten high-condition-number benchmark functions from the BBOB test suite [31] and conduct experiments under 10, 20, and 40 dimensional settings. To mitigate computational overhead and maintain continuity, the PCA mapping is updated every 5 generations. The fitness evaluations (FEs) are set to $D*10000$ for all algorithms. All remaining algorithmic parameters are set to their default or commonly recommended values.

TABLE II

MAPPING COMPARISON ON $D = 20$ (MEAN OVER 15 RUNS, $\alpha = 0.05$). SEE TABLE I FOR NOTATION.

Func	(μ,λ)-ES		DE		PSO		OPT-GAN		UMDA	
	Base	Ours	Base	Ours	Base	Ours	Base	Ours	Base	Ours
f6	**9.28e-14**	9.57e-14 (≈)	9.84e-03	**1.02e-03 (+)**	3.12e+01	**5.58e-01 (+)**	3.19e-01	**4.76e-05 (+)**	3.42e+01	**8.98e+00 (+)**
f7	6.51e+00	**5.45e+00 (≈)**	1.65e-02	**0.00e+00 (+)**	4.52e+00	**2.36e+00 (+)**	**1.78e+01**	3.19e+01 (-)	1.22e+01	**2.58e+00 (+)**
f8	**9.10e-03**	2.66e-01 (-)	2.94e+00	**1.48e+00 (+)**	6.01e+01	**3.10e+01 (≈)**	3.35e+01	**5.32e+00 (+)**	4.76e+01	**2.20e+01 (+)**
f9	5.41e-01	**5.32e-01 (+)**	1.24e+01	**1.82e+00 (+)**	2.76e+01	**1.97e+01 (≈)**	1.12e+01	**1.29e-05 (+)**	**1.85e+01**	1.86e+01 (≈)
f10	2.06e+02	**1.89e+00 (+)**	2.89e+02	**1.03e-05 (+)**	8.20e+03	**5.26e+02 (+)**	1.77e+03	**4.04e+01 (+)**	3.78e+04	**2.62e+03 (+)**
f11	1.47e+02	**8.90e-02 (+)**	4.55e+00	**5.60e-07 (+)**	1.91e+00	**6.01e-02 (+)**	3.41e+00	**3.09e-01 (+)**	8.85e+01	**5.03e+01 (+)**
f12	**3.90e+00**	4.76e+00 (≈)	4.87e+00	**7.99e-04 (+)**	2.25e+05	**2.07e+04 (+)**	1.26e+04	**5.35e+00 (+)**	1.68e+02	**1.00e+00 (≈)**
f13	3.32e+00	**3.09e+00 (≈)**	3.92e-01	**3.87e-04 (+)**	7.12e+01	**3.14e+01 (+)**	2.82e+01	**4.29e+00 (+)**	4.88e+01	**1.87e+00 (+)**
f14	4.14e-05	**4.12e-05 (≈)**	2.25e-04	**3.25e-07 (+)**	1.98e-01	**1.86e-02 (+)**	1.54e-02	**2.77e-05 (+)**	1.66e-02	**1.58e-03 (+)**
f18	**2.17e-01**	3.05e-01 (≈)	2.07e-01	**3.21e-03 (+)**	**6.22e+00**	6.63e+00 (≈)	**7.41e+00**	9.77e+00 (≈)	5.12e-01	**3.15e-02 (+)**
+ / ≈ / -		**3/6/1**		**10/0/0**		**9/1/0**		**8/2/0**		**7/3/0**

TABLE III

MAPPING COMPARISON ON $D = 40$ (MEAN OVER 15 RUNS, $\alpha = 0.05$). SEE TABLE I FOR NOTATION.

Func	(μ,λ)-ES		DE		PSO		OPT-GAN		UMDA	
	Base	Ours	Base	Ours	Base	Ours	Base	Ours	Base	Ours
f6	3.18e-13	**3.03e-13 (≈)**	8.29e+01	**2.11e+01 (+)**	1.20e+02	**6.72e-01 (+)**	1.57e+01	**9.11e-04 (+)**	5.96e+01	**3.72e+01 (+)**
f7	1.65e+01	**1.60e+01 (≈)**	2.42e+01	**1.47e-01 (+)**	2.36e+01	**1.69e+01 (≈)**	8.78e+01	**5.11e+01 (+)**	2.54e+01	**5.13e+00 (+)**
f8	2.05e-01	**3.86e-06 (+)**	4.09e+01	**3.50e+01 (+)**	1.32e+02	**4.42e+01 (+)**	1.49e+02	**1.65e+01 (+)**	7.76e+01	**3.81e+01 (+)**
f9	7.38e-01	**2.66e-01 (+)**	1.23e+02	**3.49e+01 (+)**	6.95e+01	**5.14e+01 (+)**	4.42e+01	**1.21e+01 (+)**	3.83e+01	**3.82e+01 (≈)**
f10	3.24e+02	**5.94e+00 (+)**	2.01e+05	**4.22e+02 (+)**	4.19e+04	**3.88e+03 (+)**	1.40e+04	**1.22e+02 (+)**	6.92e+04	**2.90e+03 (+)**
f11	1.85e+02	**8.31e-02 (+)**	2.23e+02	**4.29e+01 (+)**	1.90e+00	**2.26e-01 (+)**	3.44e+01	**7.41e-01 (+)**	2.01e+02	**5.47e-02 (+)**
f12	4.69e+00	**3.30e+00 (≈)**	1.89e+04	**4.00e+02 (+)**	2.43e+06	**1.93e+04 (+)**	3.98e+04	**1.57e+01 (+)**	2.50e+02	**3.54e+00 (≈)**
f13	**2.04e+00**	2.18e+00 (≈)	4.12e+01	**3.11e+00 (+)**	2.72e+02	**3.20e+01 (+)**	5.11e+01	**7.45e+00 (+)**	4.23e+01	**1.27e+00 (+)**
f14	**5.77e-05**	5.80e-05 (≈)	5.41e-02	**1.33e-03 (+)**	7.55e-01	**2.02e-02 (+)**	5.11e-02	**9.30e-05 (+)**	1.49e-02	**1.56e-03 (+)**
f18	**3.72e-01**	5.17e-01 (≈)	4.43e+00	**8.11e-01 (+)**	1.64e+01	**1.56e+01 (≈)**	1.82e+01	**1.57e+01 (≈)**	3.67e-01	**3.64e-02 (+)**
+ / ≈ / -		**5/5/0**		**10/0/0**		**8/2/0**		**6/4/0**		**8/1/1**

B. Landscape Visualization Analysis

To provide an intuitive understanding of how the proposed mapping reshapes ill-conditioned landscapes, Fig. 3 visualizes a local region of the F18 function in 2-D for a representative instance. The left subplot depicts the original search space, where the background contours illustrate the local objective landscape around the most recent generations produced by DE, and the scatter points represent the distribution of 100 candidate solutions sampled from the current population.

A typical narrow valley structure can be clearly observed in original space: the level sets are highly elongated, and the sampled solutions are tightly confined along the valley floor, indicating a strongly anisotropic local geometry. The right subplot shows the same region after applying the proposed PCA whitening mapping. In contrast to the original space, the transformed landscape becomes significantly more isotropic, and the previously elongated valley is reshaped into a near spherical geometry. Correspondingly, the mapped solutions are more evenly distributed in the transformed space, exhibiting a substantially reduced directional bias.

This visualization highlights the essential role of landscape transformation in ill-conditioned optimization. Rather than forcing the optimizer to adapt its search dynamics to a severely anisotropic geometry, the proposed mapping directly reshapes the local landscape into a form that is more amenable to efficient exploration, thereby facilitating balanced progress along all principal directions.

C. Overall Performance on Ill-Conditioned Problems

Optimization Result Analysis Tables I-III summarize the final objective function values achieved by the five representative optimizers across 10D, 20D, and 40D settings. For each benchmark, the performance of the original algorithm (Base) is compared against that equipped with the proposed framework (Ours). Statistical significance is assessed via the Mann-Whitney U test at $\alpha = 0.05$, with the accumulated $+/\approx/-$ counts (indicating "Ours" being significantly better, equivalent, or worse) provided at the bottom of each table to quantify the overall impact of the geometric correction.

From a numerical standpoint, the mean result reveal that on several strongly ill-conditioned functions, the integration of the proposed framework yields significantly improved fitness, indicating substantial gains in absolute solution accuracy. This performance enhancement is primarily attributed to the framework's capacity to fundamentally rectify the objective landscape by reparameterizing narrow, anisotropic valleys into isotropic latent spaces. By effectively decoupling decision variables and homogenizing the search scale across all principal directions, the framework enables the underlying optimizers to circumvent the classic step-size dilemma—avoiding

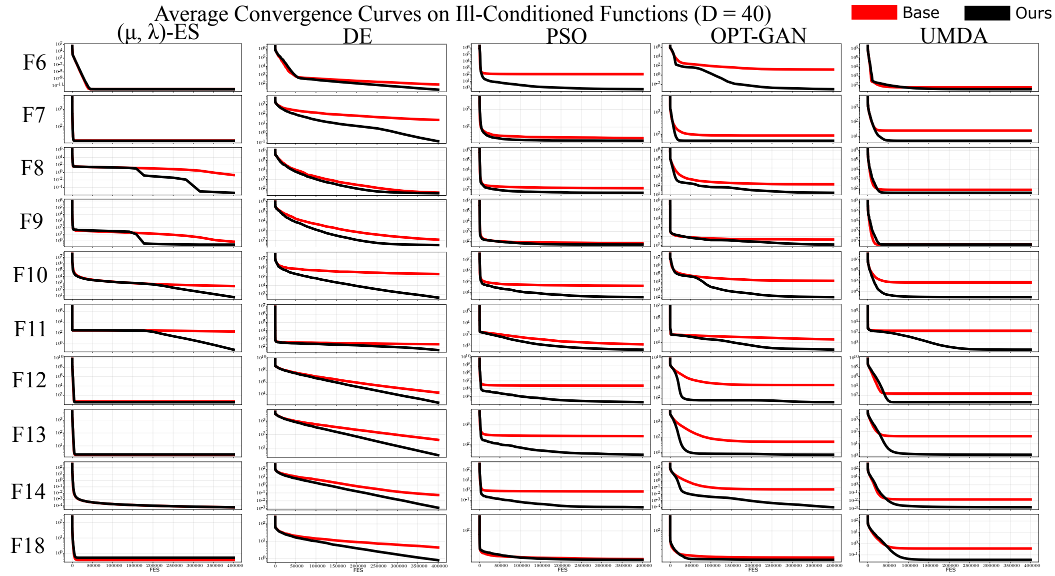

Fig. 4. Average convergence curves on ill-conditioned functions ($D = 40$). Each row corresponds to a fixed algorithm and different test functions, while each column corresponds to a fixed test function and different algorithms. The horizontal axis denotes the number of function evaluations (FEs), and the vertical axis denotes the current best objective value $f(x)$. All curves are averaged over 15 independent runs.

violent oscillations in high-sensitivity directions while maintaining steady exploration along insensitive ones. Furthermore, the consistent performance gains observed in both 20D and 40D scenarios underscore the framework's robustness and scalability, proving its efficacy across varied dimensionalities and diverse optimization paradigms.

Statistical analysis of the experimental results confirms that the integration of the PCA whitening framework yields consistent and robust performance enhancements across most optimizers within the ill-conditioned test suite. Under the $D = 20$ setting, the framework demonstrates dominant performance, with DE, PSO, OPT-GAN, and UMDA achieving significantly improved results in $10/0/0$, $9/1/0$, $8/2/0$, and $7/3/0$ cases, respectively. This superiority persists even as the problem dimensionality increases to $D = 40$; notably, DE maintains a perfect success rate $(10/0/0)$, while PSO, OPT-GAN, and UMDA continue to exhibit strong performance gains with win-draw-loss counts of $8/2/0$, $6/4/0$, and $8/1/1$, respectively. Although the performance gains for (μ, λ)-ES are less pronounced than those for DE, the framework still secures advantages across the majority of problems without compromising baseline performance.

Convergence Analysis To further analyze the convergence behavior in high-dimensional ill-conditioned scenarios, Fig. 4 presents the average convergence curves of all algorithms on several representative functions under $D = 40$.

Specifically for DE, the proposed framework typically exhibits comparable initial search dynamics to the baseline, followed by a steady divergence toward superior solutions in the later stages. This suggests that the geometric correction becomes increasingly pivotal as the search converges and focuses on local regions where anisotropy is more pronounced.For PSO and OPT-GAN, the integrated version

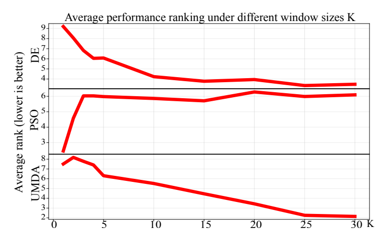

Fig. 5. Effect of window size K on average performance ranking for different optimizers.

demonstrates significantly accelerated convergence trajectories across most representative functions. This indicates that geometric reparameterization effectively enhances the exploitation of search directions within narrow valleys and highly anisotropic landscapes, allowing for more efficient navigation.Regarding UMDA, the framework markedly mitigates the late-stage stagnation prevalent in the baseline method, thereby extending the effective learning phase and facilitating deeper exploration of the objective landscape. Finally, for (μ, λ)-ES, the proposed framework demonstrates a trend similar to that of UMDA, effectively alleviating late-stage stagnation and providing an additional layer of search refinement for the baseline method.

Collectively, the statistical evidence and convergence dynamics demonstrate that the proposed framework significantly enhances both solution accuracy and search efficiency while remaining non-intrusive to the base algorithms' core mechanisms. By rectifying the severe anisotropy and scale imbalance inherent in ill-conditioned landscapes, the framework delivers robust performance gains across diverse algorithmic paradigms and problem dimensionalities.

D. Effect of the Training Window Size K

We further investigate the influence of the training window size K on the performance of the PCA whitening mapping, using the average ranking as a unified performance indicator (see Fig. 5).

The results reveal a pronounced algorithm-dependent effect of K. For DE and UMDA, the average ranking exhibits a clear decreasing trend as K increases, indicating that a larger historical sample window stabilizes the covariance estimation and improves the characterization of the principal directions, with the improvement being most pronounced for UMDA. In contrast, PSO achieves its best or near-best average rankings at small values of K, whereas larger windows tend to degrade performance, suggesting that an overly long memory may weaken the responsiveness of the mapping to the current search dynamics.

V. Conclusion

This paper proposed an algorithm-agnostic geometric correction framework for ill-conditioned black-box optimization based on PCA whitening. By learning local affine reparameterizations from recent search samples and embedding them as a plug-and-play external module, the framework reshapes unfavorable local geometries without modifying the core update mechanisms of base optimizers. Extensive experiments on high-condition-number BBOB benchmarks demonstrate that the proposed framework consistently improves final solution quality across diverse optimization paradigms and dimensional settings.

Future research plan to explore cluster-based local transformations to better accommodate the divergent geometric characteristics of multi-modal landscapes, refining the current archive-based synchronization approach.

References

[1] W. Chen, K. Chiu, and M. Fuge, "Aerodynamic design optimization and shape exploration using generative adversarial networks," in *AIAA Scitech 2019 forum*, 2019, p. 2351.

[2] S. N. Skinner and H. Zare-Behtash, "State-of-the-art in aerodynamic shape optimisation methods," *Applied Soft Computing*, vol. 62, pp. 933–962, 2018.

[3] H. F. Maathuis, R. De Breuker, and S. G. Castro, "High-dimensional bayesian optimisation with large-scale constraints-an application to aeroelastic tailoring," in *AIAA SCITECH 2024 Forum*, 2024, p. 2012.

[4] J. R. Martins and A. B. Lambe, "Multidisciplinary design optimization: a survey of architectures," *AIAA journal*, vol. 51, no. 9, pp. 2049–2075, 2013.

[5] E. F. Campana, M. Diez, U. Iemma, G. Liuzzi, S. Lucidi, F. Rinaldi, and A. Serani, "Derivative-free global ship design optimization using global/local hybridization of the direct algorithm," *Optimization and Engineering*, vol. 17, no. 1, pp. 127–156, 2016.

[6] C. Audet and W. Hare, "Introduction: tools and challenges in derivative-free and blackbox optimization," in *Derivative-Free and Blackbox Optimization*. Springer, 2017, pp. 3–14.

[7] O. Kramer, D. E. Ciaurri, and S. Koziel, "Derivative-free optimization," in *Computational optimization, methods and algorithms*. Springer, 2011, pp. 61–83.

[8] J. Demmel, "Nearest defective matrices and the geometry of ill-conditioning," *Reliable numerical computation*, vol. 44, pp. 35–55, 1990.

[9] J. Nocedal, "Numerical optimization," 2006.

[10] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," *Evolutionary computation*, vol. 9, no. 2, pp. 159–195, 2001.

[11] N. Hansen, "The cma evolution strategy: A tutorial," 2023. [Online]. Available: https://arxiv.org/abs/1604.00772

[12] E. Visser, C. E. van Daalen, and J. Schoeman, "Labcat: Locally adaptive bayesian optimization using principal-component-aligned trust regions," *Swarm and Evolutionary Computation*, vol. 97, p. 101986, 2025.

[13] S. Bagheri, W. Konen, and T. Bäck, "Sacobra with online whitening for solving optimization problems with high conditioning," *arXiv preprint arXiv:1904.08397*, 2019.

[14] J. Zhang and A. C. Sanderson, "Jade: adaptive differential evolution with optional external archive," *IEEE Transactions on evolutionary computation*, vol. 13, no. 5, pp. 945–958, 2009.

[15] Y.-l. Li, J. Zhang, and W.-n. Chen, "Differential evolution algorithm with pca-based crossover," in *Proceedings of the 14th annual conference companion on Genetic and evolutionary computation*, 2012, pp. 1509–1510.

[16] Y. Hariya, T. Shindo, and K. Jin'no, "A novel particle swarm optimization algorithm for non-separable and ill-conditioned problems," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2016, pp. 002 110–002 115.

[17] F. Sun, C. Liu, L. Wu, L. Wang, S. Liu, and B. Yang, "A novel velocity reinforced mechanism on improving particle swarm optimization for ill-conditioned problems," in *2020 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2020, pp. 1–8.

[18] J. Grahl, P. A. Bosman, and F. Rothlauf, "The correlation-triggered adaptive variance scaling idea," in *Proceedings of the 8th annual conference on Genetic and evolutionary computation*, 2006, pp. 397–404.

[19] P. A. Bosman, J. Grahl, and F. Rothlauf, "Sdr: A better trigger for adaptive variance scaling in normal edas," in *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, 2007, pp. 492–499.

[20] W. Dong and X. Yao, "Covariance matrix repairing in gaussian based edas," in *2007 IEEE Congress on Evolutionary Computation*. IEEE, 2007, pp. 415–422.

[21] Y. Cai, X. Sun, H. Xu, and P. Jia, "Cross entropy and adaptive variance scaling in continuous eda," in *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, 2007, pp. 609–616.

[22] P. A. Bosman, J. Grahl, and D. Thierens, "Enhancing the performance of maximum–likelihood gaussian edas using anticipated mean shift," in *International Conference on Parallel Problem Solving from Nature*. Springer, 2008, pp. 133–143.

[23] P. A. Bosman and J. Grahl, "Matching inductive search bias and problem structure in continuous estimation-of-distribution algorithms," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1246–1264, 2008.

[24] O. M. Shir, J. Roslund, D. Whitley, and H. Rabitz, "Evolutionary hessian learning: Forced optimal covariance adaptive learning (focal)," *arXiv preprint arXiv:1112.4454*, 2011.

[25] H. Shimizu and M. Toyoda, "Cma-es with coordinate selection for high-dimensional and ill-conditioned functions," in *Proceedings of the genetic and evolutionary computation conference companion*, 2021, pp. 209–210.

[26] X. He, Z. Zheng, Z. Chen, and Y. Zhou, "Adaptive evolution strategies for stochastic zeroth-order optimization," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 5, pp. 1271–1285, 2022.

[27] T. Eltaeib and A. Mahmood, "Differential evolution: A survey and analysis," *Applied Sciences*, vol. 8, no. 10, p. 1945, 2018.

[28] S. Sengupta, S. Basak, and R. A. Peters, "Particle swarm optimization: A survey of historical and recent developments with hybridization perspectives," *Machine learning and knowledge extraction*, vol. 1, no. 1, pp. 157–191, 2018.

[29] M. Hauschild and M. Pelikan, "An introduction and survey of estimation of distribution algorithms," *Swarm and evolutionary computation*, vol. 1, no. 3, pp. 111–128, 2011.

[30] M. Lu, S. Ning, S. Liu, F. Sun, B. Zhang, B. Yang, and L. Wang, "Opt-gan: A broad-spectrum global optimizer for black-box problems by learning distribution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 10, 2023, pp. 12 462–12 472.

[31] N. Hansen, A. Auger, R. Ros, O. Mersmann, T. Tušar, and D. Brockhoff, "Coco: A platform for comparing continuous optimizers in a black-box setting," *Optimization Methods and Software*, vol. 36, no. 1, pp. 114–144, 2021.