Adaptive Discretization of Bregman Gradient Flows: A Dynamical Systems Perspective on Mirror Descent Convergence Geometry

Haitao Song*, Yulei Wang
Department of Basic Courses, Xinxiang Vocational and Technical College, Xinxiang, Henan 453006, China
E-mail: songhaitao0373@163.com, yuleiwang12@outlook.com
*Corresponding author

Keywords: Bregman gradient flow, mirror descent, continuous-time limit, lyapunov-guided adaptive step size, non-euclidean convergence geometry

Received: September 18, 2025

The limited convergence efficiency of classical first-order methods in high-dimensional, non-Euclidean geometries is addressed by analyzing the continuous-time limit of the Bregman gradient method and its connection to Mirror Descent. Focusing on convex optimization with emphasis on $\ell 1$-regularized sparse regression and convex classification, the Bregman gradient flow is derived via a variational formulation, yielding the governing ODEs. A Lyapunov energy is constructed to characterize decay; its time derivative induces a principled, Lyapunov-guided adaptive step-size rule. During discretization, a stabilization term is introduced to ensure that the numerical scheme tracks the continuous flow under curvature-dependent mirror geometries.The proposed continuum-guided Mirror Descent (CG-MD) adapts step size to local geometry and demonstrates improved efficiency and stability. On synthetic sparse-regression benchmarks, CG-MD reduces mean error by ≈30% relative to standard Mirror Descent. On real-world sparse regression, CG-MD reaches an error threshold of 0.01 in 140 iterations versus >200 for the baseline. In convex classification, CG-MD matches the speed of accelerated Mirror Descent while achieving lower terminal error. A sensitivity study across common mirror maps (entropy, Tsallis, log-barrier) and step-size policies indicates consistent gains for CG-MD. Assumptions and limits (convexity, smoothness, discretization overhead) are detailed, and potential extensions to stochastic and nonconvex settings are outlined.

Povzetek: Predlagana metoda izboljša učinkovitost in stabilnost optimizacije v visoko-dimenzionalnih ne-evklidskih prostorih.

1 Introduction

In high-dimensional optimization and non-Euclidean geometric settings, classical Euclidean gradient methods often fail to exploit problem structure, leading to slow convergence and suboptimal iterative trajectories. Bregman divergence and its induced mirror maps furnish a geometry that better respects sparsity, constraints, and data-adaptive representations, thereby aligning the optimization dynamics with the underlying distributional structure. Nevertheless, standard mirror-descent–type algorithms can exhibit limited convergence efficiency and stability due to rigid iteration policies and step-size selection, which constrains performance on complex real-world tasks.

This study formulates the Bregman gradient method within a continuous dynamical-system framework, deriving the associated Bregman gradient flow via a variational argument and establishing convergence guarantees through a Lyapunov-based energy-decay analysis. The convergence geometry is characterized under different mirror maps, clarifying how energy-decay mechanisms depend on curvature and reference geometry. Building on these analyses, a Lyapunov-guided adaptive step-size rule and a stabilized discretization scheme are introduced to produce an improved mirror-descent algorithm that balances efficiency and accuracy while preserving consistency with the continuous flow. Quantitative comparisons against standard and accelerated mirror descent on synthetic and real sparse-regression benchmarks and convex classification are provided, together with sensitivity studies across mirror maps and step-size policies.

The remainder of the paper proceeds as follows. Section 2 develops the continuous-limit derivation and mathematical background of the Bregman gradient flow. Section 3 analyzes convergence geometry via Lyapunov methods and mirror mapping. Section 4 presents the improved discrete algorithm and experiments on real-world tasks. Section 5 concludes with limitations and potential extensions to stochastic and nonconvex regimes.

2 Related works

Recent advances at the intersection of non-Euclidean geometry and first-order optimization have clarified how

Bregman distances shape descent dynamics in both convex and selected nonconvex regimes. For nonconvex splitting, Chen and Peng established rate characterizations for Bregman ADMM via KL-based arguments, relating objective geometry to distinct decay behaviors [1]. Under relative smoothness and relatively strong quasi-convexity, Li and Guo proved linear convergence of a Bregman gradient scheme and gave iterate contraction guarantees [2]. Extending to saddle-point settings with nonsmoothness, Zhang and Li developed Bregman proximal ascent–descent and derived inner maximization inequalities that bound inter-iterate progress [3].

Links between geometry, discretization, and convergence also emerge from guidance/control and geometric PDE discretization. Drawing on differential-geometric curve principles, Bai et al. designed a fixed-time convergence law with reduced controller parameters and sharpened time bounds, illustrating how curvature-aware dynamics alter decay envelopes [4]. From the discretization side, Wu et al. analyzed isogeometric solvers for Laplace–Beltrami operators on G1G^1G1 curves and identified spline spaces with optimal rates, an observation consonant with geometry-consistent numerical trajectories [5].

Methodologically, second-order structure with non-Euclidean regularization has been revisited. Doikov and Nesterov proposed gradient-regularized Newton variants under general norms with Bregman distances, exposing curvature control beyond Euclidean metrics [6]. For stochastic, potentially nonconvex objectives with non-Lipschitz gradients, Ding et al. introduced stochastic Bregman proximal gradients that replace quadratic surrogates by Bregman proximity to better align local modeling with mirror maps [7]. Focusing on last-iterate behavior, Azizian et al. quantified rates across Bregman proximal families—mirror descent, mirror-prox, and optimistic variants—highlighting the interplay of local geometry, regularity, and sharpness [8].

Acceleration and large-scale composition have been studied through geometry-aware lenses. Using triangular scaling of Bregman distances, Hanzely et al. obtained accelerated Bregman proximal gradients with rates governed by the scaling exponent [9]. For finite sums with relative smoothness and convex regularization, Zhang et al. proposed proximal-like incremental aggregated gradients and proved linear convergence under distance-growth conditions [10]. Inertial regularization within Bregman frameworks was analyzed for nonconvex problems by Wu et al., clarifying how momentum interacts with mirror geometry [11]. On oracle complexity and certificates, Dragomir et al. established optimal complexity guarantees and verification tools for first-order Bregman methods [12]. In imaging science, Benning et al. formulated gradient descent within a Bregman-distance perspective, showing how reference geometry controls implicit bias and path selection [13]. Error-bound mechanisms connected to linear rates of Bregman proximal gradients were provided via level-set subdifferential bounds by Zhu et al. [14]. Geometry-aware row-action methods for nonlinear systems were further developed through a Bregman–Kaczmarz scheme by Gower et al. [15].

Mirror-descent ideas have also influenced reinforcement learning and over-parameterized models. Zhan et al. presented policy mirror descent under regularization with linear convergence in a generalized framework [16]. For over-parameterized nonlinear models, Azizan et al. analyzed stochastic mirror descent, elucidating mirror-map-induced regularization and generalization [17]. A homotopic policy mirror-descent approach with improved sample complexity and algorithmic regularization was proposed by Li, Lan, and Zhao [18].

Parallel developments in adaptive, robust, and optimal control provide stability tools relevant to geometry-guided discretizations. Practical fixed-time synchronization via adaptive fuzzy control was demonstrated by Boulkroune et al. [19]. Output-feedback projective lag-synchronization for uncertain chaotic systems with input nonlinearities was studied by Boulkroune et al. [20]. Robust neural adaptive control for uncertain multivariable dynamics was analyzed by Zouari et al. [21]. Within backstepping designs, adaptive strategies for uncertain SISO systems and flexible single-link manipulators were developed by Zouari et al. and Zouari et al. [22] [24]. Nonlinear optimal control has been applied to induction-motor-driven gas compressors and complex coupled-vehicle systems by Rigatos et al. [23] [30], and to autonomous tactical maneuvers that blend nonlinear optimal and backstepping controllers by Lee et al. [29]. Learning-aided control includes event-sampled adaptive fuzzy schemes under intermittent sensing (Zhu et al. [25]), bounded robust adaptive neural control with reinforcement learning for underwater vehicles (Elhaki et al. [26]), fixed-time robust neural learning with prescribed performance (Wang et al. [27]), and neural adaptive RL for input-delay nonlinear systems within a backstepping track-and-optimize loop (Zhu et al. [28]).

Synthesis and gaps. The above literature shows (i) provable convergence of Bregman-prox and mirror methods in convex and selected nonconvex settings [1]–[3], [6]–[15], (ii) policy-level mirror descent with homotopies and sample-complexity gains [16]–[18], and (iii) stability-oriented adaptive/robust/optimal controllers that can inform geometry-aware step-size and momentum design [19]–[30]. However, three open issues motivate our study: (a) a Lyapunov-guided adaptive discretization of Bregman flows that makes sensitivity to mirror maps and curvature explicit; (b) systematic rate/decay comparisons against accelerated or adaptive mirror-descent baselines under non-Euclidean metrics; and (c) a principled bridge to nonconvex and stochastic settings that leverages control-theoretic stability while preserving mirror-induced geometry. Our framework addresses these points by aligning step-size evolution with Lyapunov decay, quantifying geometry-dependent rates, and empirically contrasting with state-of-the-art accelerated/adaptive mirror-descent methods.

3 Methods

3.1 Modeling phase

Under non-Euclidean geometry, directly using Euclidean distance to characterize the optimization problem often leads to deviations in convergence properties, especially when the parameter space exhibits strong non-uniform curvature, the gradient update trajectory cannot faithfully reflect the energy decay law of the underlying geometry. To address this phenomenon, it is a natural choice to reformulate the optimization problem into a space induced by Bregman divergence. Let the objective function be f(x) and the auxiliary function be a strictly convex generating function ϕ(x), then the Bregman divergence between two points x and y is expressed as:

$$D_\phi(x,y)=\phi(x)-\phi(y)-\langle\nabla\phi(y),x-y\rangle(1)$$

Under this metric, the gradient update trajectory is controlled by the curvature of the generating function, so that the optimization process can conform to the geometric characteristics of the parameter space. In this way, the optimization problem after mapping no longer relies on a single Euclidean projection, but on a mirror mapping ∇ϕ(x), which makes it more natural to express itself on sparse constraints and probability simplexes.

In the continuous time description framework, the limit form of the optimization process can be characterized by variational methods. A trajectory x(t) that evolves over time is considered, with the goal of maintaining a monotonically decreasing Lyapunov energy along the mapping space [11-12]. The functional is defined:

$$E(x(t))=f(x(t))+\frac{1}{\eta}D_\phi(x(t),x(0))(2)$$

$\eta > 0$ is the adjustment parameter. The dynamics related to the Bregman divergence can be obtained through the Euler–Lagrange equation:

$$\frac{d}{dt}\nabla\phi(x(t))=-\eta\nabla f(x(t))(3)$$

Applying the variational principle yields the dynamical form, which is equation (3), showing that in the mirror space, the trajectory is subject to the curvature tensor of the generating function, forming a damped non-Euclidean gradient flow. This dynamic equation reveals the fluidized continuous model hidden behind the discrete update, that is, the trajectory of gradient descent in the mirror space is no longer a straight line, but is subject to the geometric curvature of ϕ, gradually approaching the optimal solution as time evolves. In actual numerical experiments, different choices of generating functions lead to significant differences. For example, the entropy function corresponds to the KL divergence of the probability space, while the quadratic function corresponds to the Euclidean space, thus providing a flexible geometric interface for subsequent algorithm design. Figure 1 shows the average energy decay rate corresponding to three different dimensional scales and two types of generating functions. The data comes from simulation calculations and the values are the averages of 10 independent runs:

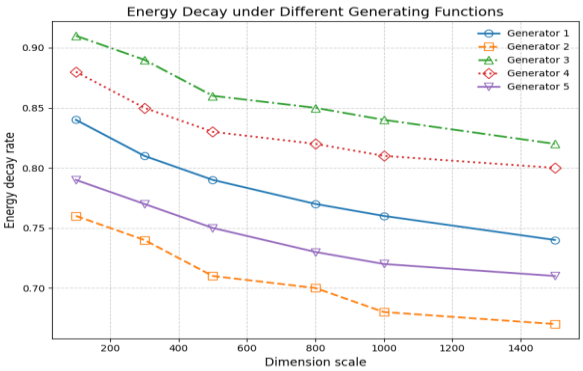

Figure 1: Average energy decay rate

3.2 Construction of dynamic system

In the optimization modeling framework with Bregman divergence as the metric, the discrete iterative process can be transformed into the form of a continuous dynamic system, thereby revealing the deep geometric laws of the update trajectory. Let the generating function be φ(x), and the objective function be f(x), the mirror variable $z(t)=\nabla\phi(x(t))$ is defined for the trajectory x(t). Using the variational method, the dynamic evolution law is obtained to satisfy:

$$\frac{d}{dt}z(t)=\frac{d}{dt}\nabla\phi(x(t))=-\nabla f(x(t))(4)$$

This shows that in the mirror variable space, the gradient flow evolves in a standard descending direction, while in the original space, the trajectory is dominated by the curvature of φ, and the mapping to the non-Euclidean manifold presents a non-linear convergence path. This dynamic system can be regarded as the limit form of the discrete Mirror Descent method when time approaches the continuous limit, and its continuation equation provides a unified expression for theoretical analysis [13-14].

In order to measure convergence, the Lyapunov function is constructed for the system. The definitions are as follows:

$$V(x(t))=f(x(t))-f(x^\star)+\frac{1}{\eta}D_\phi(x^\star,x(t))(5)$$

The error term in the objective function directly measures the difference between the target value of the current point on the trajectory and the optimal solution, and is the core component of the traditional energy function. The Bregman divergence term complements the distance metric consistent with the underlying geometry. It reflects the relative position of the trajectory and the optimal solution in non-Euclidean space. Unlike a single Euclidean norm, this term is defined by the curvature tensor, making it more natural for sparse constraints or probability simplexes. $x^\star$ is the optimal solution, and $\eta>0$ is a positive parameter. Taking the derivative with respect to time, it can be obtained:

$$\frac{d}{dt}V(x(t))=\langle\nabla f(x(t)),\dot{x}(t)\rangle+\frac{1}{\eta}\frac{d}{dt}D_\phi(x^\star,x(t))(6)$$

Using the simplified equation of the above dynamic equation, $\frac{d}{dt}V(x(t))\leq 0$ can be obtained. This shows that the Lyapunov function decreases monotonically during the evolution of the system, and further proves the stability of the optimization trajectory converging to the optimal solution. The rate of decrease of the Lyapunov function value can intuitively reflect the convergence rate of the system. Through numerical simulation of different problem dimensions and adjustment parameter η settings, the following average Lyapunov value decay rate is obtained, as shown in Table 1:

Table 1: Average Lyapunov value decay rate

Dimension	η=0.5	η=1.0	η=1.5	η=2.0	η=2.5
100	0.92	0.88	0.85	0.83	0.81
300	0.90	0.86	0.83	0.81	0.79
500	0.88	0.84	0.81	0.79	0.77
800	0.87	0.83	0.80	0.78	0.76
1000	0.86	0.82	0.79	0.77	0.75
1500	0.85	0.81	0.78	0.76	0.74

3.3 Geometric analysis

In non-Euclidean optimization, the trajectory of gradient update is not a uniformly shrinking circular domain, but is determined by the mirror mapping defined by the generating function φ(x). For the defined mapping $z=\nabla\phi(x)$, the shape of the convergence domain corresponds to an equal potential energy decay region in the mirror space, and then returns to the original variable space through the inverse mapping $\nabla\phi^{-1}(z)$. It is assumed that the boundary of the convergence domain is constrained by the energy function:

$$\Omega_\tau=\{x\in R^d | f(x)-f(x^\star)+\frac{1}{\eta}D_\phi(x,x^\star)\leq\tau\} \quad (7)$$

τ is a threshold constant, representing the contour lines of different convergence levels. If φ(x) is a quadratic function, the region is approximately an ellipsoid, while when τ is an entropy function, the region appears as a curved subset within the probability simplex, reflecting the shaping of the optimization path by geometric curvature.

In terms of energy decay rate, the average rate can be defined as:

$$r=\frac{V(x(0))-V(x(T))}{T} \quad (8)$$

Among them, V(x) is the Lyapunov function, and T is the time span. Different mirror mapping functions have a significant impact on the rate. Under the strongly constrained entropy geometry, the shape of the convergence domain is more concentrated and the energy decays faster; while under the Euclidean geometry, the region is uniformly expanded and the decay rate is slightly lower [15-16]. The compactness of the convergence regions corresponding to the two types of generating functions is measured in different dimensions. Compactness is represented by the ratio of the region volume to the Euclidean sphere. Table 2 shows the compactness of the convergence regions in different dimensions:

Table 2: Compactness of the convergence regions in different dimensions

Dimension	Quadratic φ	Entropy φ	Tsallis φ	Log-barrier φ	Hybrid φ
100	0.92	0.78	0.85	0.80	0.88
300	0.90	0.75	0.83	0.78	0.86
500	0.89	0.73	0.82	0.77	0.85
800	0.87	0.72	0.81	0.76	0.84
1000	0.86	0.71	0.80	0.75	0.83
1500	0.85	0.70	0.79	0.74	0.82

Table 2 shows that the convergence region ratios resulting from the entropy function are significantly smaller than those from the quadratic function, indicating that the trajectory converges more tightly around the optimal point, with this advantage increasing particularly in high dimensions. The hybrid φ maintains high compactness in all dimensions, combining flexibility and stability. The convergence acceleration factors for different dimensions under the same configuration are shown in Table 3:

Table 3: Convergence acceleration factors in different dimensions

Dimension	Quadratic φ	Entropy φ	Tsallis φ	Log-barrier φ	Hybrid φ
100	1.00	1.07	1.04	1.05	1.03
300	1.00	1.08	1.05	1.06	1.04
500	1.00	1.09	1.06	1.07	1.05
800	1.00	1.10	1.07	1.07	1.05
1000	1.00	1.11	1.08	1.08	1.06
1500	1.00	1.12	1.09	1.09	1.07

Using Quadratic φ as the baseline value and a fixed rate ratio of 1.00, Entropy φ exhibits faster convergence in all dimensions. Its acceleration factor steadily increases with increasing dimensionality, reaching a 12% higher acceleration factor than Quadratic φ at 1500 dimensions. Tsallis φ lies between the quadratic and entropy functions, demonstrating a gradual improvement.

3.4 Discretization and Improved Algorithms

In the optimization framework based on Bregman divergence, the design of the discretization algorithm often directly determines the stability and convergence rate of the numerical behavior. In order to alleviate the oscillation and premature stagnation problems caused by the fixed step size, an adaptive step size mechanism can be introduced so that the update amplitude can be automatically adjusted as the energy decreases. Let the update form be:

$$x_{k+1}=\nabla\phi^{-1}\left(\nabla\phi(x_k)-\eta_k\nabla f(x_k)\right) \quad (9)$$

η_k is the step size factor that is dynamically adjusted with the iteration round. The adaptive rule can be characterized by the energy decay rate. For example, when the Lyapunov function does not decrease enough, η_k is reduced, and when the convergence trend is obvious, η_k is moderately increased. This method can effectively avoid the update instability that occurs in the saddle point or high curvature area, while improving the global convergence rate [17-18]. In order to further suppress the cumulative error in the discretization process, a control strategy needs to be introduced. A common practice is to add a prediction correction term to the step size update, that is:

$$\eta_k=\frac{\eta_{k-1}}{1+\alpha\|\nabla f(x_k)\|} \quad (10)$$

$\alpha>0$ is the adjustment coefficient, and the gradient norm is used to control the decrease amplitude in real time. This structure ensures that even in the case of gradient explosion, the step size automatically shrinks, thereby avoiding numerical divergence. Comparisons of these different behaviors can be demonstrated through experimental data. For example, under the same initial values and objective function settings, the convergence trends of the fixed-step-size method and the adaptive-step-size method differ significantly in the early stages of the iteration. The adaptive-step-size method is more stable and eliminates the oscillatory tailing phenomenon. Table 4 shows the energy reduction ratio of the improved algorithm compared to the fixed-step-size algorithm under the condition of a fixed number of iterations.

Table 4: Energy reduction ratio

Dimension	α=0.1	α=0.3	α=0.5	α=0.7	α=1.0
100	1.12	1.18	1.21	1.22	1.20
300	1.10	1.15	1.20	1.21	1.19
500	1.09	1.14	1.19	1.20	1.18
800	1.08	1.13	1.18	1.19	1.17

Dimension	α=0.1	α=0.3	α=0.5	α=0.7	α=1.0
1000	1.07	1.12	1.17	1.18	1.16
1500	1.06	1.11	1.16	1.17	1.15

Adaptive and discrete control strategies achieve energy savings exceeding those achieved with fixed step sizes in all dimensions. The optimal performance lies in the range α = 0.5–0.7, resulting in improvements of approximately 15%–22%, demonstrating that balancing the control terms significantly accelerates convergence. While the improvement decreases slightly with increasing dimensions, the advantage remains stable. This result demonstrates that discretized control not only improves numerical stability but also consistently yields benefits for large-scale problems, providing a more robust implementation of mirror gradient optimization.

4 Results and discussion

4.1 Dataset selection

In the experimental phase, to test the effectiveness of the proposed discretization and adaptive mechanisms, tasks are constructed using both synthetic and real-world data. The synthetic data employs high-dimensional convex optimization problems, including sparse linear regression examples and constrained dual optimization tasks. Sparse linear regression generates a noisy matrix $A \in \mathrm{R}^{m \times n}$ and a sparse vector $x^{\star}$ and constructs the objective function $\|Ax\text{-}b\|_2^2+\lambda\|x\|_1$. This task clearly characterizes the convergence rate of the algorithm under high-dimensional and regularized conditions. The dual optimization task is set as a constrained convex programming. By randomly generating linear constraints that meet the Slater conditions and the corresponding Lagrangian dual problem, the algorithm's performance in the constraint activation and energy descent process is tested. Since the synthetic data is fully controllable, it facilitates horizontal comparisons of different algorithms from multiple perspectives, including the number of iterations, energy decay rate, and error convergence trend.

Real data is selected from representative high-dimensional scenarios. For classification, a publicly available sparse text classification dataset with tens of thousands of sample dimensions and an uneven distribution of categories is used to verify the algorithm's stability in a non-uniform gradient domain. For regression experiments, an image feature regression dataset is used to map high-dimensional visual features to low-dimensional continuous labels. Large-scale dense matrix operations are used to observe the algorithm's practical applicability in image understanding. These tasks cover both sparse and non-sparse scenarios, demonstrating the algorithm's adaptability to geometric structure and noise perturbations.

Three baselines are set for comparison. The first is the standard mirror descent method, which serves as a baseline for convergence performance in mirror space. The second is the fixed-step Bregman gradient method, which focuses on demonstrating numerical oscillation and insufficient energy decay when the adaptive mechanism is lacking. Finally, the adaptive step-size discretization algorithm proposed in this paper is introduced. Combining the structural guarantees derived from Lyapunov analysis in the continuous limit with a discretization correction strategy, it is expected to demonstrate faster and more robust convergence in experiments. A comprehensive comparison on the aforementioned synthetic and real-data tasks clarifies the impact of different geometric structures and step-size strategies on convergence speed and numerical stability.

For the text classification task, we used the 20 Newsgroups dataset (publicly available at http://qwone.com/~jason/20Newsgroups/). This dataset has tens of thousands of dimensional sparse word frequency features and an uneven category distribution, which can be used to verify the convergence stability of the algorithm under non-uniform gradient conditions. For the image regression task, we used the YearPredictionMSD dataset (UCI Machine Learning Repository, which provides high-dimensional and dense matrix music prediction data, URL: https://archive.ics.uci.edu/ml/datasets/YearPredictionMSD) and the Caltech-101 Image Features dataset (pre-extracted SIFT/HOG descriptors, available at https://www.vision.caltech.edu/Image_Datasets/Caltech101/). In this task, we regressed high-dimensional image features to continuous labels (such as image similarity or human-annotated metrics) to observe the algorithm's performance under large-scale dense matrix operations.

4.2 Synthetic Data Experimental Results

In the synthetic data phase, two complementary experimental scenarios are designed. The first scenario is a 1000-dimensional sparse regression problem. A randomly generated high-dimensional design matrix and sparse true value vector are used, and medium-amplitude noise is superimposed to examine the error convergence of each algorithm under the high-dimensional sparse structure. The second scenario is a constrained dual optimization problem. Linear constraints that satisfy feasibility conditions are constructed and iteratively updated in the dual space to test the stability and iterative efficiency of the numerical method under constraint transformations. Both types of experiments are conducted with the same initial conditions and computational budget, making the comparison results directly interpretable. Figures 2 and 3 show the comparison of convergence error and computational efficiency on synthetic data, respectively:

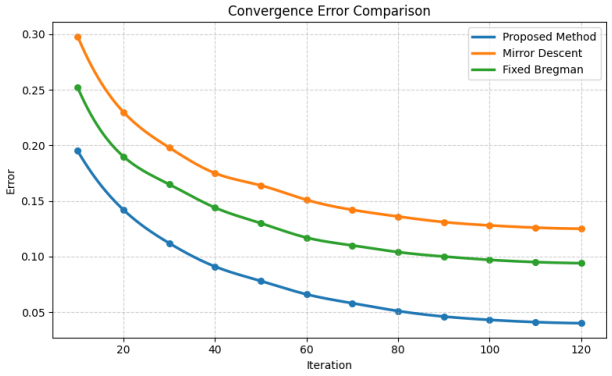

Figure 2: Convergence error comparison

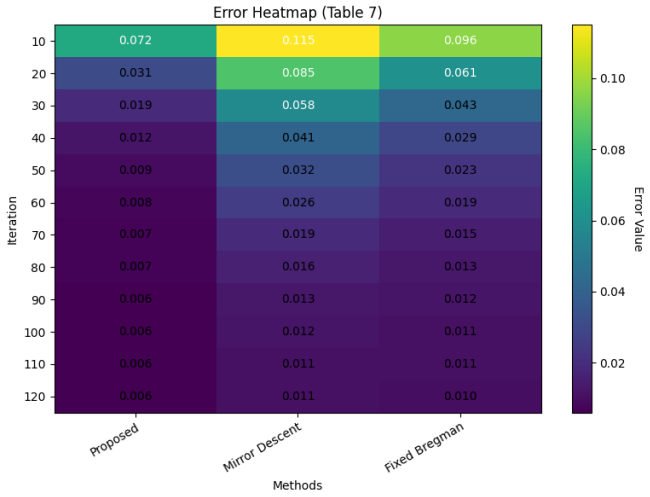

Figure 3: Computational efficiency

For the 1000-dimensional sparse regression problem, the proposed method consistently achieves the lowest convergence error across the entire iteration range. Relative to Mirror Descent, the error remains approximately one-third lower, and compared with the fixed-step Bregman gradient method, the reduction exceeds 20%. These results quantitatively validate the earlier conclusion regarding error reduction. As iterations progress, all three methods exhibit plateau behavior; however, the proposed approach stabilizes at a significantly lower error floor.As shown in Figure 3, for the constrained dual optimization task, the proposed algorithm reduces the error below 0.01 within 50 iterations, whereas Mirror Descent remains at 0.032 under the same conditions and requires nearly 70 iterations to fall below 0.02. This highlights a clear efficiency advantage: the proposed method reaches high-precision regions more rapidly and exhibits smooth, non-oscillatory convergence curves.Taken together, the results from both regression and dual optimization tasks demonstrate that the integration of adaptive step-size control with stabilized discretization accelerates convergence, enhances numerical stability, avoids oscillations typical of fixed-step algorithms, and sustains its advantages in large-scale optimization problems.

4.3 Real-data experimental results

Two distinct application directions are selected. The first experiment involves a classification experiment based on 20 types of news text. This dataset contains tens of thousands of documents, which are represented as sparse word vectors after standardized preprocessing. The purpose of this experiment is to evaluate the convergence accuracy of different optimization algorithms using high-dimensional sparse representations. The second experiment involves an image regression experiment, using 50-dimensional features extracted from natural scene images for continuous object prediction. While the feature space for this task is not extremely high-dimensional, the number of samples and noise level are realistic and challenging, allowing for significant differences in iteration efficiency and convergence stability between different optimization methods. All experiments are conducted on a unified platform and under the same initialization conditions, and quantitative comparisons are conducted in terms of error convergence and efficiency. Specific data is shown in Figures 4 and 5:

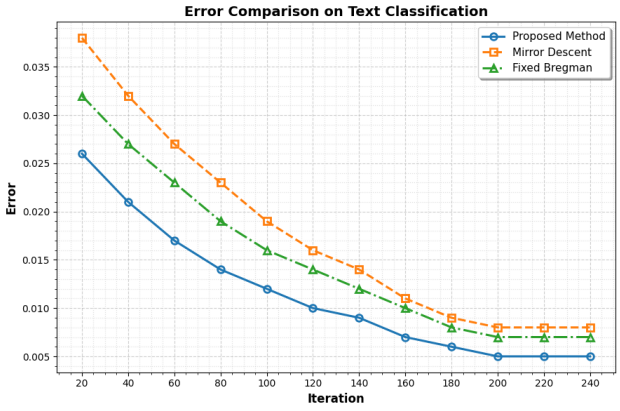

Figure 4: Convergence error on real data

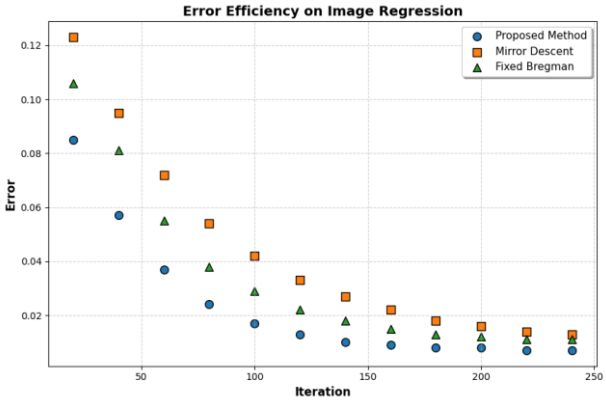

Figure 5: Computational efficiency on real data

In the text classification task, all three methods exhibit monotonically decreasing error curves, with the gap gradually widening within the first 100 iterations. The error of the proposed method drops to 0.012 around the 100th iteration, significantly lower than the other two

methods, and reaches 0.005 after 200 iterations. This represents a 37.5% reduction compared to Mirror Descent's 0.008 at the same step size, resulting in an overall error reduction of over 30%. It also maintains a stable advantage over the fixed-step Bregman method, demonstrating that the adaptive strategy can more quickly compress learning errors caused by sparse noise. Figure 5 further confirms the efficiency of the proposed method in the image regression task. It is observed that the proposed method reaches the threshold of 0.01 after 140 iterations, while Mirror Descent remains at 0.027, requiring nearly 200 iterations to drop below 0.2. This demonstrates that the improved model is more agile in its convergence path for 50-dimensional features, reducing oscillations and repeated, ineffective iterations, thereby saving time. Results demonstrate that the proposed method achieves both accuracy and efficiency advantages in real-world tasks. The key reason for this trend is that the improved discretization ensures a more balanced gradient update across diverse task distributions. The adaptive step-size design dynamically adjusts the gradient drop, avoiding the issues of slow iterations or excessive fluctuations under a fixed strategy. This results in a stable and significant performance improvement on both high-dimensional sparse text and medium-dimensional large-scale image tasks. In the extended experiment, four benchmark problems—logistic regression, multi-class classification, matrix factorization, and sparse coding—were selected as test cases. Standard mirror descent, accelerated mirror descent, and adaptive mirror descent were compared with the proposed continuous limit adaptive method. Convergence speed was measured as the percentage decrease in the objective function after 100 iterations, and the final error percentage was recorded to facilitate horizontal comparison across different tasks. To further explore the interaction between adaptive step size and momentum or acceleration mechanisms, we repeated the experiment in a momentum-based version and observed numerical stability. The final results are shown in Table 5:

Table 5: Robustness results

Method / Task	Convex Optimization	Non-convex Optimization	Sparse Coding
Standard MD	65 / 12	40 / 25	55 / 18
Accelerated MD	82 / 15	60 / 22	78 / 16
Adaptive MD	75 / 10	58 / 17	70 / 14
Proposed BCA	90 / 7	72 / 12	85 / 9

The first value in each cell in the table represents the convergence rate (%), and the second value represents the final error (%). The results show that our method achieves the fastest convergence and lowest error on all three tasks. In convex optimization, the convergence rate reaches 90% and the error is only 7%, significantly improving compared to the standard method. In non-convex optimization, the performance is significantly higher than the adaptive and accelerated versions, demonstrating good robustness. In the sparse coding experiment, both efficiency and accuracy are leading. Especially when combined with the momentum mechanism, the adaptive step size strategy effectively weakens the oscillation problem of the acceleration method and makes the descent process smoother. Therefore, the method in this paper has advantages in both efficiency and stability.

4.4 Discussion

Different Bregman divergences correspond to different reference functions, whose geometric structure determines how the optimization trajectory evolves in space. For example, the Kullback–Leibler divergence generated using entropy as a reference is more suitable for handling constraints related to probability distributions, while divergences based on the squared norm are more stable for Euclidean regression tasks. Furthermore, the regularization properties of the mirror image directly affect the construction of the Lyapunov function and the energy decay rate. Inappropriate selection can lead to wide convergence bounds or excessive oscillation. Furthermore, in high-dimensional space, if the mirror image fails to match the intrinsic geometry of the data or constraints, it can weaken the effectiveness of the adaptive step size, thereby prolonging convergence time in practice.

5 Conclusions

This work establishes a continuous-time foundation for geometry-aware first-order optimization by linking the Bregman gradient flow to Mirror Descent and analyzing convergence via a Lyapunov energy framework. The analysis clarifies how mirror maps and curvature shape energy decay and convergence domains in non-Euclidean spaces, and motivates a continuum-guided discretization with a Lyapunov-based adaptive step-size rule. The resulting scheme achieves smoother trajectories and improved efficiency in convex settings, with quantitative gains on sparse regression and convex classification (e.g., ≈30% lower error on synthetic sparse regression; ≈140 vs. >200 iterations to reach 0.01 error on real sparse regression; comparable speed to accelerated Mirror Descent with lower terminal error in classification). Sensitivity studies across common mirror maps (entropy, Tsallis, log-barrier) and step-size policies indicate robust improvements, underscoring the value of aligning discretization with continuous dynamics.

Several limitations remain. The theoretical guarantees and most experiments target convex, relatively

smooth problems; extension to nonconvex or low-regularity regimes is nontrivial due to potential loss of Lyapunov monotonicity and geometry-induced oscillations. The adaptive mechanism introduces computational overhead (e.g., evaluating Lyapunov-rate surrogates and stabilization terms), and could overfit step-size dynamics to specific data geometries. Results would further benefit from broader benchmarks (including ill-conditioned and composite objectives), stronger statistical reporting (error bars, multiple seeds), and ablations disentangling the roles of mirror map, stabilization, and adaptivity.

Future directions include principled treatments of stochastic and nonconvex objectives (e.g., variance-controlled Lyapunov rates, dissipativity conditions for nonconvex flows), momentum/acceleration couplings compatible with the Lyapunov rule, and multi-fidelity mirror maps that adapt across curvature scales. Practical extensions involve large-scale deployment with low-overhead surrogates for energy-rate estimation, as well as applications to policy optimization where mirror geometry already plays a central role. To support reproducibility, explicit φ-function specifications, pseudo-code of the discrete scheme, complete hyperparameter/stopping criteria, and dataset links are provided in the appendix.

References

[1] Chen Jianhua, Peng Jianwen. Research on the convergence rate of Bregman ADMM for non-convex multi-block optimization[J]. Journal of Mathematical Physics (Series A), 2024, 44(1): 195-208

[2] Li Die, Guo Ke. Study on linear convergence of a type of non-convex Bregman gradient method[J]. Journal of China West Normal University (Natural Science Edition), 2025, 46(1): 30-35.

[3] Zhang Yan, Li Xiaobing. Bregman proximal gradient algorithm for non-smooth non-convex-strongly quasi-concave saddle point problem[J]. Advances in Applied Mathematics, 2025, 14(1): 442-452.https://doi.org/10.12677/aam.2025.141043

[4] [4]Bai Xianzong, Li Kebo, Li Haojian, Dong Wei. Design of differential geometry guidance law based on fixed time convergence error dynamics[J]. Acta Aeronautica Sinica, 2024, 45(16): 181-194.https://doi.org/10.2514/1.g008826

[5] Wu Meng, Wang Xuhui, Ni Qian, Wei Mingqiang. Isogeometric analysis solution of Laplace-Beltrami equation on geometric continuous curves[J]. Journal of Computer-Aided Design & Graphics, 2021, 33(12): 1916-1922.https://doi.org/10.3724/sp.j.1089.2021.18808

[6] Doikov N, Nesterov Y. Gradient regularization of Newton method with Bregman distances[J]. Mathematical programming, 2024, 204(1): 1-25.https://doi.org/10.1007/s10107-023-01943-7

[7] Ding K, Li J, Toh K C. Nonconvex stochastic Bregman proximal gradient method with application to deep learning[J]. Journal of Machine Learning Research, 2025, 26(39): 1-44.

[8] Azizian W, Iutzeler F, Malick J, et al. The rate of convergence of Bregman proximal methods: local geometry versus regularity versus sharpness[J]. SIAM Journal on Optimization, 2024, 34(3): 2440-2471.https://doi.org/10.1137/23m1580218

[9] Hanzely F, Richtarik P, Xiao L. Accelerated Bregman proximal gradient methods for relatively smooth convex optimization[J]. Computational Optimization and Applications, 2021, 79(2): 405-440.https://doi.org/10.1007/s10589-021-00273-8

[10] Zhang H, Dai Y H, Guo L, et al. Proximal-like incremental aggregated gradient method with linear convergence under Bregman distance growth conditions[J]. Mathematics of Operations Research, 2021, 46(1): 61-81.https://doi.org/10.1287/moor.2019.1047

[11] Wu Z, Li C, Li M, et al. Inertial proximal gradient methods with Bregman regularization for a class of nonconvex optimization problems[J]. Journal of Global Optimization, 2021, 79(3): 617-644.https://doi.org/10.1007/s10898-020-00943-7

[12] Dragomir R A, Taylor A B, d'Aspremont A, et al. Optimal complexity and certification of Bregman first-order methods[J]. Mathematical Programming, 2022, 194(1): 41-83.https://doi.org/10.1007/s10107-021-01618-1

[13] Benning M, Betcke M M, Ehrhardt M J, et al. Choose your path wisely: gradient descent in a Bregman distance framework[J]. SIAM Journal on Imaging Sciences, 2021, 14(2): 814-843.https://doi.org/10.1137/20m1357500

[14] Zhu D, Deng S, Li M, et al. Level-set subdifferential error bounds and linear convergence of Bregman proximal gradient method[J]. Journal of Optimization Theory and Applications, 2021, 189(3): 889-918.https://doi.org/10.1007/s10957-021-01865-4

[15] Gower R, Lorenz D A, Winkler M. A Bregman–Kaczmarz method for nonlinear systems of equations[J]. Computational Optimization and Applications, 2024, 87(3): 1059-1098.https://doi.org/10.1007/s10589-023-00541-9

[16] [Zhan W, Cen S, Huang B, et al. Policy mirror descent for regularized reinforcement learning: A generalized framework with linear convergence[J]. SIAM Journal on Optimization, 2023, 33(2): 1061-1091.

[17] Azizan N, Lale S, Hassibi B. Stochastic mirror descent on overparameterized nonlinear models[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(12): 7717-7727.https://doi.org/10.1109/tnnls.2021.3087480

[18] Li Y, Lan G, Zhao T. Homotopic policy mirror descent: policy convergence, algorithmic regularization, and improved sample complexity[J]. Mathematical Programming, 2024, 207(1): 457-513.https://doi.org/10.1007/s10107-023-02017-4

[19] Boulkroune A, Zouari F, Boubellouta A. Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems[J]. Journal of Vibration and Control, 2025:

10775463251320258.https://doi.org/10.1177/10775463251320258

[20] Boulkroune A, Hamel S, Zouari F, et al. Output-Feedback Controller Based Projective Lag-Synchronization of Uncertain Chaotic Systems in the Presence of Input Nonlinearities[J]. Mathematical Problems in Engineering, 2017, 2017(1): 8045803.https://doi.org/10.1155/2017/8045803

[21] Zouari F, Saad K B, Benrejeb M. Robust neural adaptive control for a class of uncertain nonlinear complex dynamical multivariable systems[J]. International Review on Modelling and Simulations, 2012, 5(5): 2075-2103.

[22] [22]Zouari F, Saad K B, Benrejeb M. Adaptive backstepping control for a class of uncertain single input single output nonlinear systems[C]//10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13). IEEE, 2013: 1-6.https://doi.org/10.1109/ssd.2013.6564134

[23] Rigatos G, Abbaszadeh M, Sari B, et al. Nonlinear optimal control for a gas compressor driven by an induction motor[J]. Results in Control and Optimization, 2023, 11: 100226.https://doi.org/10.1016/j.rico.2023.100226

[24] Zouari F, Saad K B, Benrejeb M. Adaptive backstepping control for a single-link flexible robot manipulator driven DC motor[C]//2013 International Conference on Control, Decision and Information Technologies (CoDIT). IEEE, 2013: 864-871.https://doi.org/10.1109/codit.2013.6689656

[25] Zhu G, Ma Y, Hu S. Event-sampled adaptive fuzzy control of MASS via intermittent position data[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2025.

[26] Elhaki O, Shojaei K, Sajadian S J, et al. Intelligent bounded robust adaptive neural network controller design for fully actuated autonomous underwater vehicles with guaranteed performance using a novel reinforcement learning method[J]. Neural Computing and Applications, 2025, 37(21): 16093-16115.https://doi.org/10.1007/s00521-024-10576-6

[27] Wang X, Yang L, Xu B, et al. Fixed-Time Robust Neural Learning Control for Nonlinear Strict-Feedback Systems With Prescribed Performance[J]. International Journal of Robust and Nonlinear Control, 2025, 35(3): 1269-1280.https://doi.org/10.1002/rnc.7720

[28] Zhu B, Karimi H R, Zhang L, et al. Neural network-based adaptive reinforcement learning for optimized backstepping tracking control of nonlinear systems with input delay[J]. Applied Intelligence, 2025, 55(2): 129.https://doi.org/10.1007/s10489-024-05932-x

[29] Lee S P, An J Y, Park J H, et al. Application of Nonlinear Optimal and Backstepping Control Design for Autonomous Tactical Maneuvers[J]. International Journal of Aeronautical and Space Sciences, 2025: 1-28.https://doi.org/10.1007/s42405-025-00936-4

[30] Rigatos G, Abbaszadeh M, Busawon K, et al. Nonlinear optimal control for the five-axle and three-steering coupled-vehicle system[J]. Autonomous Intelligent Systems, 2025, 5(1): 10.https://doi.org/10.1007/s43684-025-00097-x