





Article

Mirror Descent and Exponentiated Gradient Algorithms Using Trace-Form Entropies

Andrzej Cichocki ^{1,2,3,4,*} , Toshihisa Tanaka ^{3,*} , Frank Nielsen ^{5,*}  and Sergio Cruces ^{6,*} 

¹ Systems Research Institute of Polish Academy of Science, Newelska 6, 01-447 Warsaw, Poland

² Department of Electrical Engineering, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland

³ Department of Electronic and Information Engineering, Tokyo University of Agriculture and Technology, Koganei-shi, Tokyo 184-8588, Japan

⁴ Riken Artificial Intelligence Project (AIP), 1 Chome-4-1, Nihonbashi, Tokyo 103-0027, Japan

⁵ Sony Computer Science Laboratories, Tokyo 141-0022, Japan

⁶ Department of Signal Processing and Communications, Universidad de Sevilla, 41092 Seville, Spain

* Correspondence: cichockiand@gmail.com (A.C.); tanakat@cc.tuat.ac.jp (T.T.); frank.nielsen.x@gmail.com (F.N.); sergio@us.es (S.C.)

Abstract

This paper introduces a broad class of Mirror Descent (MD) and Generalized Exponentiated Gradient (GEG) algorithms derived from trace-form entropies defined via deformed logarithms. Leveraging these generalized entropies yields MD and GEG algorithms with improved convergence behavior, robustness against vanishing and exploding gradients, and inherent adaptability to non-Euclidean geometries through mirror maps. We establish deep connections between these methods and Amari's natural gradient, revealing a unified geometric foundation for additive, multiplicative, and natural gradient updates. Focusing on the Tsallis, Kaniadakis, Sharma–Taneja–Mittal, and Kaniadakis–Lissia–Scarfone entropy families, we show that each entropy induces a distinct Riemannian metric on the parameter space, leading to GEG algorithms that preserve the natural statistical geometry. The tunable parameters of deformed logarithms enable adaptive geometric selection, providing enhanced robustness and convergence over classical Euclidean optimization. Overall, our framework unifies key first-order MD optimization methods under a single information-geometric perspective based on generalized Bregman divergences, where the choice of entropy determines the underlying metric and dual geometric structure.

Keywords: mirror descent; natural gradient; information geometry; deformed logarithms; generalized exponentiated gradient; Bregman divergences; Riemannian optimization; (q, κ) -algebra



Academic Editors: Antonio M. Scarfone and Tatsuaki Wada

Received: 24 October 2025

Revised: 27 November 2025

Accepted: 27 November 2025

Published: 8 December 2025

Citation: Cichocki, A.; Tanaka, T.; Nielsen, F.; Cruces, S. Mirror Descent and Exponentiated Gradient Algorithms Using Trace-Form Entropies. *Entropy* **2025**, *27*, 1243. <https://doi.org/10.3390/e27121243>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

This paper is dedicated to Professor Shun-Ichi AMARI in honor of his 90th birthday.

1. Introduction

Mirror descent (MD), initially proposed by Nemirovsky and Yudin [1], has become an increasingly popular topic in optimization, artificial intelligence, and machine learning domains [2–6]. Its profound success stems not merely from its algorithmic efficiency, but also from deep mathematical connections to information geometry and the natural statistical structure underlying optimization problems. These connections, particularly to Amari's Natural Gradient (NG) method, reveal that effective optimization is fundamentally about respecting the intrinsic geometry of the parameter space rather than imposing artificial Euclidean constraints [7–9].

The central motivation for our research emerges from a fundamental insight in information geometry: optimized learning algorithms should adapt to the Fisher information metric of the underlying statistical manifold. This principle, pioneered by Amari in the context of neural networks, establishes that the steepest descent direction on a statistical manifold is not the Euclidean gradient, but rather the natural gradient, the direction that accounts for the curvature induced by the Fisher information matrix [10,11].

1.1. The Information Geometry Perspective

The connection between Mirror Descent and Natural Gradient runs deeper than algorithmic similarity—it represents a fundamental mathematical equivalence that has been rigorously established [2,12,13]. Our work extends this principle by showing that trace-form entropies induce natural Fisher-like metrics that can be even more appropriate for specific problem structures. Through deformed logarithms, we can systematically explore the space of possible geometries and the possibility of discovering optimal choices of hyperparameters for given data distributions [14,15].

1.2. Challenge of Geometric Selection

While the power of geometric optimization is well established, a fundamental challenge remains: how to select the appropriate geometry for a given optimization problem? Classical approaches require domain expertise and manual tuning, limiting their applicability. Our approach addresses this through parameterized entropy families, where hyperparameters control the geometric structure [2].

The theoretical foundation of our approach rests on the connection between Bregman divergences and exponential families established in information geometry [8]. This connection indicates that choosing a generalized entropy is equivalent to selecting an appropriate exponential family structure for specific optimization problems. See also [16] for an extension of the logarithmic divergences extending to Bregman divergences. Our deformed logarithms allow us to systematically explore the space of possible exponential families, enabling discovery of optimal statistical models.

The Exponentiated Gradient (EG) and its extensions emerge as a specific and powerful instantiation of the Mirror Descent framework when the mirror map is constructed from generalized entropies and deformed logarithms. This connection is far from superficial—it represents a fundamental mathematical relationship that unifies additive and multiplicative gradient updates within a single theoretical framework [15,17–21].

It is important to note that Mirror Descent updates can be reparameterized as Gradient Descent in appropriately chosen coordinate systems [2,12]. This reveals that the computational complexity of Natural Gradient can be considerably reduced. Our deformed logarithms approach gives some insight by showing that deformed logarithms naturally induce reparameterizations that preserve geometric structure while enabling efficient computation and provide implicit regularization through the choice of entropy or deformed logarithm.

1.3. Research Contributions and Scope

In this work, we systematically investigate the theoretical foundations and practical implications of employing trace-form entropies and deformed logarithms in the Mirror Descent framework.

Our primary contributions include:

- **Mathematical Framework:** We establish a comprehensive mathematical foundation connecting generalized entropies, deformed logarithms, and Mirror Descent updates, providing explicit formulations for numerous well-established trace entropy families.

- **Algorithmic Innovations:** We derive novel Generalized Exponentiated Gradient (GEG) algorithms with generalized multiplicative updates that leverage the flexibility of hyperparameter-controlled deformed logarithms, enabling adaptation to problem geometry.
- The significance of this research extends beyond algorithmic development; it opens new avenues for understanding the geometric foundations of optimization and provides practical tools for addressing increasingly complex machine learning challenges. The unifying theoretical framework connects optimization theory, information geometry, statistical physics, and practical machine learning, opens up new research directions, and provides principled approaches to algorithm design that respect the natural geometric structure of optimization problems.

2. Preliminaries: Mirror Descent (MD) and Standard Exponentiated Gradient (EG) Updates

Notations: Vectors are denoted by boldface lowercase letters, e.g., $\mathbf{w} \in \mathbb{R}^N$, where for any vector \mathbf{w} , we denote its i -th entry by w_i . For any vectors $\mathbf{w}, \mathbf{v} \in \mathbb{R}^N$, we define the Hadamard product as $\mathbf{w} \odot \mathbf{v} = [w_1 v_1, \dots, w_N v_N]^T$ and $\mathbf{w}^\alpha = [w_1^\alpha, \dots, w_N^\alpha]^T$. All operations for vectors like multiplications and additions are performed component-wise. The function of a vector is also applied for any entry of the vectors, e.g., $f(\mathbf{w}) = [f(w_1), f(w_2), \dots, f(w_N)]^T$. The N -dimensional real vector space with nonnegative real numbers is denoted by \mathbb{R}_+^N . We let $\mathbf{w}(t)$ denote the weight or parameter vector as a function of time t . The learning process advances in iterative steps, where during step t we start with the weight vector $\mathbf{w}(t) = \mathbf{w}_t$ and update it to a new vector $\mathbf{w}(t+1) = \mathbf{w}_{t+1}$. We define $[x]_+ = \max\{0, x\}$, and the gradient of a differentiable cost function as $\nabla_{\mathbf{w}} L(\mathbf{w}) = \partial L(\mathbf{w}) / \partial \mathbf{w} = [\partial L(\mathbf{w}) / \partial w_1, \dots, \partial L(\mathbf{w}) / \partial w_N]^T$. In contrast to deformed logarithms defined later, the classical natural logarithm will be denoted by $\ln(x)$.

2.1. Problem Statement

We consider the constrained optimization problem:

$$\mathbf{w}_{t+1} = \arg \min_{\mathbf{w} \in \mathbb{R}_+^N} \left\{ L(\mathbf{w}) + \frac{1}{\eta} D_F(\mathbf{w} || \mathbf{w}_t) \right\}, \quad (1)$$

where $L(\mathbf{w})$ is a continuously differentiable loss function, $\eta > 0$ is the learning rate and $D_F(\mathbf{w} || \mathbf{w}_t)$ is the Bregman divergence induced by a strictly convex generating function $F(\mathbf{w})$ (used here as a regularizer) [2,22]. The Bregman divergence provides the geometric foundation for Mirror Descent algorithms:

$$D_F(\mathbf{w} || \mathbf{w}_t) = F(\mathbf{w}) - F(\mathbf{w}_t) - (\mathbf{w} - \mathbf{w}_t)^T f(\mathbf{w}_t), \quad (2)$$

where the generating (or potential) function $F(\mathbf{w})$ is a continuously-differentiable, strictly convex function defined on the convex domain, while $f(\mathbf{w}) = \nabla_{\mathbf{w}} F(\mathbf{w})$ is the mirror map, called also the link function, which is a strictly monotonically increasing function. For fundamental properties and for some extensions, see, e.g., [23–26].

The Bregman divergence measures the difference between $F(\mathbf{w})$ and its first-order Taylor approximation around \mathbf{w}_t , providing a natural measure of geometric proximity that respects the curvature induced by F . This geometric structure is intimately connected to information geometry—different choices of F correspond to different Riemannian metrics on the parameter manifold. The Bregman divergence $D_F(\mathbf{w} || \mathbf{w}_t)$ arising from generating (potential) function $F(\mathbf{w})$ can be viewed as a measure of curvature. The Bregman divergence includes many well-known divergences commonly used in practice, namely, the squared

Euclidean distance, Kullback–Leibler divergence (relative entropy), Itakura-Saito distance, beta divergence and many more [10,14,27–29].

2.2. Mirror Descent Update Rules and Geometric Interpretation

Setting the gradient of the objective in Equation (1) to zero yields the implicit update:

$$f(\mathbf{w}_{t+1}) = f(\mathbf{w}_t) - \eta \nabla_{\mathbf{w}} L(\mathbf{w}_{t+1}), \quad (3)$$

or equivalently

$$\mathbf{w}_{t+1} = f^{(-1)}[f(\mathbf{w}_t) - \eta \nabla_{\mathbf{w}} L(\mathbf{w}_{t+1})], \quad (4)$$

where $f^{(-1)}$ is inverse function of the link function. Note that when F is separable and continuous, the inverse function F^{-1} is defined globally (by the inverse function theorem). In general, the implicit function theorem only guarantees local inversion of multivariate functions but not the existence of global inverse functions. However, when F is a multivariate convex function of Legendre-type so is its convex conjugate F^* and their gradients are reciprocal to each others globally: $\nabla F = (\nabla F^*)^{-1}$ and $\nabla F^* = (\nabla F)^{-1}$.

Assuming that $\nabla_{\mathbf{w}} L(\mathbf{w}_{t+1}) \approx \nabla_{\mathbf{w}} L(\mathbf{w}_t)$, we obtain the explicit Mirror Descent update [2,5]:

$$\mathbf{w}_{t+1} = f^{(-1)}[f(\mathbf{w}_t) - \eta \nabla_{\mathbf{w}} L(\mathbf{w}_t)] \quad (5)$$

$$= \nabla F^{(-1)}(\nabla F(\mathbf{w}_t) - \eta \nabla_{\mathbf{w}} L(\mathbf{w}_t)). \quad (6)$$

In MD, we map our primal point \mathbf{w} to the dual space (through the mapping via the link function $f(\mathbf{w}) = \nabla F(\mathbf{w})$) and take a step in the direction given by the gradient of the function, then we map back to the primal space by using the inverse function of the link function. The advantage of using Mirror Descent (MD) over Gradient Descent is that it takes into account the geometry of the problem through suitable choice of a link function.

Dual Space Interpretation: Mirror Descent operates by

1. Mapping to dual space: $\Theta = f(\mathbf{w}) = \nabla F(\mathbf{w})$,
2. Taking gradient step: $\Theta_{t+1} = \Theta_t - \eta \nabla L(\mathbf{w}_t)$,
3. Mapping back to primal: $\mathbf{w}_{t+1} = f^{-1}(\Theta_{t+1})$.

This three-step process naturally incorporates problem geometry through the choice of link function f (Figure 1).

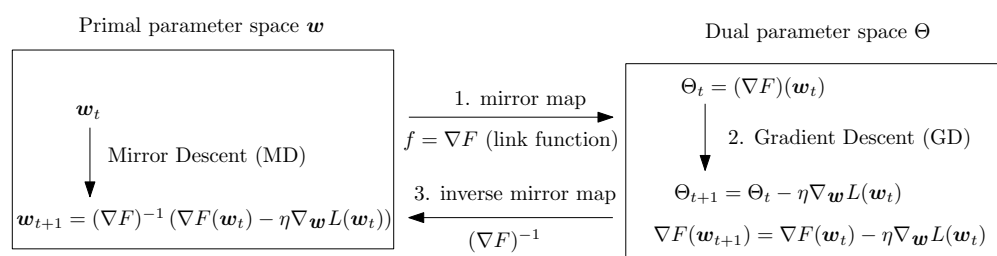


Figure 1. Mirror descent is a three-step process: 1. Map parameter with link function to dual space (mirror space), 2. Perform gradient descent in dual space, and 3. Maps back to primal parameter space using the inverse of the link function.

For example, consider $F(\mathbf{w}) = \sum_i \mathbf{w}_i \log \mathbf{w}_i$, the Shannon negative entropy. The link function is $f(\mathbf{w}) = \nabla F(\mathbf{w}) = [1 + \log \mathbf{w}_i]_i$ with inverse map $f^{-1}(\Theta) = \left[\frac{e^{\Theta_i}}{\sum_j e^{\Theta_j}} \right]_i$. The corresponding mirror update is the exponentiated gradient update:

$$\mathbf{w}_{t+1} = \mathbf{w}_t \odot \exp(-\eta \nabla_{\mathbf{w}} L(\mathbf{w}_t)).$$

This is a standard and useful algorithm for optimization on the probability simplex that is recovered as the mirror descent with respect to the Kullback–Leibler (KL) divergence (a Bregman divergence). The underlying geometric structure is the KL Hessian geometry, an example of dually flat space in information geometry.

Note that when the generating function F is separable across its coordinates (i.e., $F(\mathbf{w}) = \sum_i F(w_i)$), the Hessian matrix $\nabla^2 F(\mathbf{w})$ is diagonal.

2.3. Continuous-Time Formulation and Natural Gradient Connection

The continuous-time limit (as $\Delta t \rightarrow 0$) yields the mirror flow ODE:

$$\boxed{\frac{d f(\mathbf{w}(t))}{dt} = -\mu \nabla_{\mathbf{w}} L(\mathbf{w}(t))}, \quad (7)$$

where $\mu = \eta/\Delta t > 0$ is the learning rate for continuous-time learning, and $f(\mathbf{w}) = \nabla F(\mathbf{w})$ is a suitably chosen link function [2]. Using the chain rule, we can write mirror flow as follows

$$\frac{d f(\mathbf{w})}{dt} = \frac{d f(\mathbf{w})}{d \mathbf{w}} \odot \frac{d \mathbf{w}}{dt} = \text{diag} \left\{ \frac{d f(\mathbf{w})}{d \mathbf{w}} \right\} \frac{d \mathbf{w}}{dt} = -\mu \nabla_{\mathbf{w}} L(\mathbf{w}(t)). \quad (8)$$

Hence, we can obtain continuous-time MD update in alternative form:

$$\frac{d \mathbf{w}}{dt} = -\mu \text{diag} \left\{ \left(\frac{d f(\mathbf{w})}{d \mathbf{w}} \right)^{-1} \right\} \nabla_{\mathbf{w}} L(\mathbf{w}_t) = -\mu [\nabla^2 F(\mathbf{w})]^{-1} \nabla_{\mathbf{w}} L(\mathbf{w}(t)). \quad (9)$$

This reveals that Mirror Descent in continuous-time is equivalent to Natural Gradient descent with the Riemannian metric $H_F(\mathbf{w}) = [\nabla^2 F(\mathbf{w})]^{-1}$. This connection, established rigorously in [2,13], shows that geometric optimization methods are fundamentally unified.

2.4. Discrete Natural Gradient Form

The discrete version becomes

$$\boxed{\mathbf{w}_{t+1} = \left[\mathbf{w}_t - \eta \text{diag} \left\{ \left(\frac{d f(\mathbf{w}_t)}{d \mathbf{w}_t} \right)^{-1} \right\} \nabla_{\mathbf{w}} L(\mathbf{w}_t) \right]_+} \quad (10)$$

where $\text{diag} \left\{ \left(\frac{d f(\mathbf{w})}{d \mathbf{w}} \right)^{-1} \right\} = \text{diag} \left\{ \left(\frac{d f(\mathbf{w})}{d w_1} \right)^{-1}, \dots, \left(\frac{d f(\mathbf{w})}{d w_N} \right)^{-1} \right\}$, which we term Mirror-less Mirror Descent (MMD), representing a first-order approximation to second-order Natural Gradient methods (10) [30].

It should be noted that the above-defined diagonal matrix can be considered as the inverse of the Hessian matrix, if it exists and has positive diagonal entries for a specific set of parameters. The MMD is a special form of Natural Gradient Descent (NGD) [2,7,13].

F. Nielsen provided a geometric interpretation of NG and its connections with the Riemannian gradient, the mirror descent, and the ordinary additive gradient descent [31].

2.5. Canonical Examples and Some Geometric Insights

Case 1: For $F(\mathbf{w}) = \|\mathbf{w}\|_2^2/2 = \frac{1}{2} \sum_{i=1}^N w_i^2$ and link function $f(\mathbf{w}) = \nabla_{\mathbf{w}} F(\mathbf{w}) = \mathbf{w}$, we obtain the standard (additive) gradient descent

$$\frac{d \mathbf{w}(t)}{dt} = -\mu_t \nabla_{\mathbf{w}} L(\mathbf{w}(t)) \quad (11)$$

and its time-discrete approximate version

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \nabla_{\mathbf{w}} L(\mathbf{w}_t). \quad (12)$$

Case 2: For $F(\mathbf{w}) = \sum_{i=1}^N w_i \ln(w_i) - w_i$ and corresponding (componentwise) link function $f(w) = \ln(w)$ we obtain a (multiplicative) Exponentiated Gradient (EG) update also called the unnormalized EG update (or EGU) [20]:

$$\frac{d \ln(\mathbf{w}(t))}{dt} = -\mu \nabla_{\mathbf{w}} L(\mathbf{w}(t)), \quad \mathbf{w}(t) > 0 \quad \forall t. \quad (13)$$

In this sense, the unnormalized exponentiated gradient update (EGU) corresponds to the discrete-time version of the continuous ODE, obtained via Euler's rule:

$$\begin{aligned} \mathbf{w}_{t+1} &= \exp(\ln(\mathbf{w}_t) - \mu \Delta t \nabla_{\mathbf{w}} L(\mathbf{w}_t)) \\ &= \mathbf{w}_t \odot \exp(-\eta \nabla_{\mathbf{w}} L(\mathbf{w}_t)), \end{aligned}$$

where \odot and \exp are componentwise multiplication and componentwise exponentiation, respectively, and $\eta = \mu \Delta t > 0$ is the learning for discrete-time updates. This multiplicative update naturally preserves positivity constraints and corresponds to the natural geometry of the probability simplex.

2.6. Motivation for Using Parameterized Deformed Logarithms

Traditional Mirror Descent methods suffer from geometric rigidity—the fixed choice of mirror map f cannot adapt to diverse problem structures or data distributions. This limitation motivates our investigation of parameterized mirror maps based on trace-form entropies.

Adaptive Geometric Framework: Our approach addresses this fundamental limitation by introducing hyperparameter-controlled mirror maps $f_{\Theta}(\mathbf{w})$ that can:

- Adapt to statistical properties of training distributions.
- Interpolate between different geometries (e.g., Euclidean, exponential family, power-law).
- Provide automatic regularization through geometric bias.
- Enable systematic geometry exploration rather than ad-hoc selection.

Information-Theoretic Foundation: The connection between exponential families and Bregman divergences suggest that optimal mirror maps should reflect the underlying statistical structure of optimization problems. In fact, trace-form entropies provide systematic frameworks for discovering these optimal geometric structures.

There are many potential choices of mirror map $f(\mathbf{w})$ that can model the geometry of various optimization problems and adapt to the distribution of training data. In high dimensions (large-scale optimization), it can be advantageous to abandon the Euclidean geometry to improve convergence rates and performance. Using mirror descent with an appropriately chosen function we can obtain a considerable improvement.

3. Why Trace Entropies and Deformed Logarithms in MD and GEG?

Entropy measures provide natural regularization mechanisms and geometric structures for optimization algorithms. The connection between entropies, information theory, and geometry runs deep; each entropy functional induces a deformed logarithm and a unique Riemannian manifold structure through its associated Fisher information metric [32–35].

Trace entropies are functionals expressible in the explicit summation form [36–41]

$$S(\mathbf{p}) = \sum_i p_i f(1/p_i), \quad (14)$$

where p_i are probability values and $f(\cdot)$ is a suitable monotonically increasing function. The term “trace” refers to their mathematical structure, which resembles the trace operation for matrices, i.e., a direct summation over individual components.

Trace entropies are intimately connected to deformed logarithms through $f(x) = \log_D(x)$, where \log_D represents a deformed logarithm function with specific mathematical properties ensuring proper entropic behavior [34,38].

A function $\log_D(x)$ qualifies as a deformed logarithm if it satisfies the following conditions:

- Domain $\log_D(x): \mathbb{R}^+ \rightarrow \mathbb{R}$
- Strictly monotonically increasing: $\frac{d \log_D(x)}{dx} > 0$
- Concavity (optional): $\frac{d^2 \log_D(x)}{dx^2} < 0$
- Scaling and normalization: $\log_D(1) = 0$, $\left. \frac{d \log_D(x)}{dx} \right|_{x=1} = 1$
- Duality: $\log_D(1/x) = -\log_D(x)$.

These axioms ensure that deformed logarithms generate well-behaved entropy functionals while providing sufficient mathematical flexibility for geometric adaptation. The concavity requirement ensures that resulting entropies satisfy the maximum entropy principle, while duality guarantees symmetric treatment of probabilities and their reciprocals, which are essential for consistent statistical interpretation.

Remark 1. It should be noted that since the generating (potential) function $F(\mathbf{w})$, (which is the integral of the link function $f(\mathbf{w})$) must be strictly convex, it is sufficient that the link function $f(\mathbf{w}) = \log_D(\mathbf{w})$ is a strictly monotonically increasing function, i.e., its first derivative must be positive but the negativity of its second derivative is in the general not necessary.

It should be noted that deformed logarithms and their corresponding deformed exponential functions can be flexibly tuned by one or more hyperparameters, whose optimization enables the adaptation to specific data distributions and problem geometries. By tuning/learning these hyperparameters, we adapt to the distribution of training data and/or we can adjust them to achieve desired properties of gradient descent algorithms.

It is of great importance to understand the mathematical structure of the generalized logarithms and its inverse generalized exponentials in order to obtain more insight into the proposed MD or EG update schemes. Motivated by this fact, and to make this paper more self-contained, we systematically revise fundamental properties of the deformed logarithms and their inverses, generalized exponentials and investigate links between them.

We provide in the Appendix A the basics of q -algebra and κ -algebra and calculus in Appendix A.1 [42].

4. MD and GEG Updates Using the Tsallis Entropy and Its Extensions

4.1. Properties of the Tsallis q -Logarithm and q -Exponential

In physics, the Tsallis entropy is a generalization of the standard Boltzmann–Gibbs entropy [34,43,44]. It is proportional to the expectation of the deformed q -logarithm (referred here as the Tsallis logarithm or termed logarithm) of a distribution.

The Tsallis q -logarithm is defined for $x > 0$ as [45]

$$\log_q^T(x) = \begin{cases} \frac{x^{1-q} - 1}{1-q} & \text{for } x > 0 \text{ if } q \neq 1, \\ \ln(x) & \text{for } x > 0 \text{ if } q = 1 \\ -\infty & \text{for } x = 0 \forall q. \end{cases} \quad (15)$$

The inverse function of the Tsallis q -logarithm is the deformed q -exponential function $\exp_q^T(x)$, defined as follows [45]:

$$\exp_q^T(x) = \begin{cases} [1 + (1-q)x]_+^{1/(1-q)} & \text{for } \begin{cases} x \in (-\infty, 1/(1-q)) & \text{if } q < 1, \\ x \in (1/(q-1), \infty) & \text{if } q > 1, \end{cases} \\ \exp(x) & \text{for } q = 1. \end{cases} \quad (16)$$

It is easy to check that these functions satisfy the following relationships:

$$\log_q^T(\exp_q^T(x)) = x, \quad (0 < \exp_q^T(x) < \infty), \quad (17)$$

$$\exp_q^T(\log_q^T(x)) = x, \quad \text{for } x > 0. \quad (18)$$

Remark 2. The q -deformed exponential and logarithmic functions were introduced in Tsallis statistical physics in 1994 [44]. However, the q -deformation is related to the Box–Cox transformation (for $q = 1 - \lambda$), which was proposed in 1964 [46].

The plots of the q -logarithm and q -exponential functions for various values of q are illustrated in Figures 2 and 3.

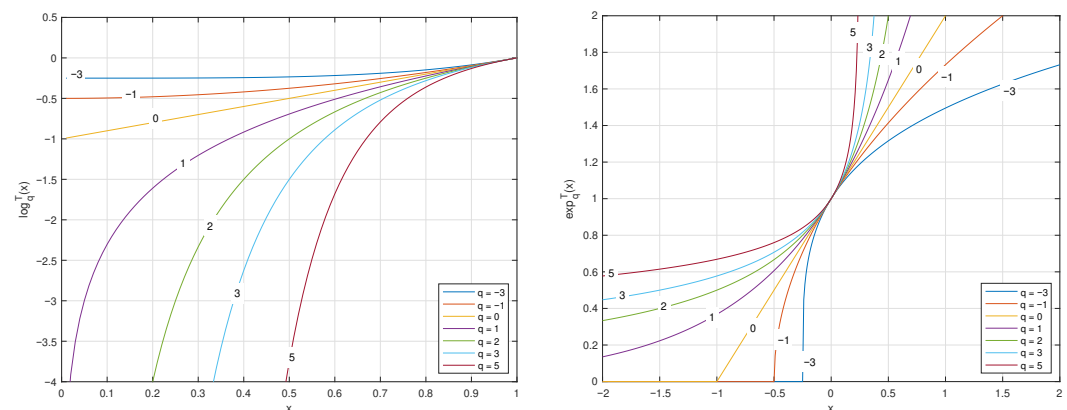


Figure 2. Plots of the q -logarithm $\log_q^T(x)$ and q -exponential $\exp_q^T(x)$ functions for different values of parameter q . From the figure, one can observe how the q parameter controls the degree of concavity/convexity of the q -logarithm as well as the degree of convexity/concavity of the q -exponential. Since the q -logarithm is convex for $q < 0$, linear for $q = 0$, and strictly concave for $q > 0$, particularizing to the classical logarithm for $q = 1$.

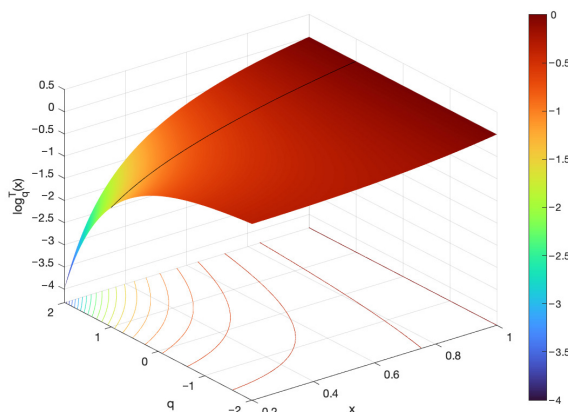


Figure 3. A 3D plot of the q -logarithm $\log_q^T(x)$. The black continuous line represents the reference of the classical logarithm $\ln(x)$, which is obtained for $q = 1$.

It should be noted that q -functions can be approximated by power series as follows:

$$\log_q^T(x) \approx \ln(x) + \frac{1}{2}(|1-q|)(\ln(x))^2 + \frac{1}{6}(1-q)^2(\ln(x))^3 + \dots, \quad (19)$$

and

$$\exp_q^T(x) \approx 1 + x + \frac{1}{2}qx^2 + \frac{1}{6}(2q^2 - q)x^3 \dots \quad (20)$$

$$= \exp(x) + \frac{1}{2}(q-1)x^2 + \frac{1}{6}(2q^2 - q - 1)x^3 + O(x^4). \quad (21)$$

These functions have the following basic properties [44,47–49]:

$$\log_q^T(x) = -\log_{2-q}^T(1/x) \quad (22)$$

$$\frac{\partial \log_q^T(x)}{\partial x} = \frac{1}{x^q} > 0 \quad (23)$$

$$\frac{\partial^2 \log_q^T(x)}{\partial x^2} = \frac{-q}{x^{q+1}} < 0 \text{ for } q > 0. \quad (24)$$

It is easy to prove the following fundamental properties:

$$\log_q^T(xy) = \log_q^T(x) + \log_q^T(y) + (1-q)\log_q^T(x)\log_q^T(y) \text{ if } x > 0, y > 0 \quad (25)$$

$$\exp_q^T(x)\exp_q^T(y) = \exp_q^T(x + y + (1-q)xy). \quad (26)$$

Using these properties we can define nonlinear generalized algebraic forms q -sum and the q -product (for more details about q -algebra, see [47,50])

$$x \oplus_q^T y = x + y + (1-q)xy, \quad (x \oplus_1^T y = x + y), \quad (27)$$

$$x \otimes_q y = [x^{1-q} + y^{1-q} - 1]_+^{1/(1-q)} \text{ if } x > 0, y > 0 \quad (x \otimes_1^T y = xy). \quad (28)$$

Using this notation, and definitions of q -exponential function (16) and q -logarithm (15), we can write the following formulas

$$\exp_q^T(x+y) = \exp_q^T(x) \otimes_q \exp_q^T(y), \quad (29)$$

$$\text{for } 1+(1-q)x > 0, 1+(1-q)y > 0, 1+(1-q)(x+y) > 0,$$

$$\exp_q^T(\log_q^T(x)+y) = x \otimes_q \exp_q^T(y), \quad (30)$$

$$\text{for } x > 0, 1+(1-q)y > 0, x^{1-q} + (1-q)y > 0.$$

which play a key role in this paper.

4.2. MD and GEG Updates Using the Tsallis q -Logarithm

Let us assume that the link function in Mirror Descent can take the following componentwise form

$$f_q(\mathbf{w}) = \log_q^T(\mathbf{w}), \quad \mathbf{w} = [w_1, \dots, w_N]^T \in \mathbb{R}_+^N. \quad (31)$$

In this case the generating function $F(\mathbf{w}) = \sum_i w_i \log_q(w_i) - \log_{q-1}(w_i)$ and the Bregman divergence is a well-known beta divergence [28]:

$$\begin{aligned} D_{F_q}(\mathbf{w}_{t+1} || \mathbf{w}_t) &= \sum_{i=1}^N w_{i,t+1} (\log_q(w_{i,t+1}) - \log_q(w_{i,t})) - \log_{q-1}(w_{i,t+1}) + \log_{q-1}(w_{i,t}) \\ &= \sum_{i=1}^N w_{i,t+1} \frac{w_{i,t+1}^{1-q} - w_{i,t}^{1-q}}{1-q} - \frac{w_{i,t+1}^{2-q} - w_{i,t}^{2-q}}{2-q}, \end{aligned} \quad (32)$$

where $\beta = 1 - q$, $\beta \neq 0$.

Applying Equation (6) and taking into account Formula (31), we obtain a novel generalized exponentiated gradient update referred to as q -GEG or q -MD update

$$\mathbf{w}_{t+1} = \exp_q^T[\log_q^T(\mathbf{w}_t) - \eta \nabla_{\mathbf{w}} L(\mathbf{w}_t)] = \mathbf{w}_t \otimes_q \exp_q^T(-\eta \nabla_{\mathbf{w}} L(\mathbf{w}_t)) \quad (33)$$

where \otimes_q q -product defined by Equation (28) is performed componentwise.

The above q -MD update can be written in a scalar (componentwise) form as

$$\begin{aligned} w_{i,t+1} &= w_{i,t} \otimes_q \exp_q^T(-\eta \nabla_{w_i} L(\mathbf{w}_t)) \\ &= \left[w_{i,t}^{1-q} + \left(\exp_q^T(-\eta \nabla_{w_i} L(\mathbf{w}_t)) \right)^{1-q} - 1 \right]_+^{1/1-q} \end{aligned} \quad (34)$$

By applying the property (31) and substituting $y \rightarrow y/(1 - (1-q)x)$, we can obtain the following identity

$$\exp_q^T(x+y) = \exp_q^T(x) \exp_q^T\left(\frac{y}{1+(1-q)x}\right) \text{ for } \begin{cases} x, y \in (-\infty, 1/(1-q)) & \text{if } q < 1, \\ x, y \in (1/(q-1), \infty) & \text{if } q > 1, \end{cases} \quad (35)$$

Hence, we obtain a simplified generalized q -GEG update

$$w_{i,t+1} = w_{i,t} \exp_q^T\left(\frac{-\eta \nabla_{w_i} L(\mathbf{w}_t)}{w_{i,t}^{1-q}}\right), \quad (36)$$

which can be written in a compact vector form

$$\mathbf{w}_{t+1} = \mathbf{w}_t \odot \exp_q^T[-\eta_t \odot \nabla_{\mathbf{w}} L(\mathbf{w}_t)] \quad (37)$$

where a vector of learning rates

$$\eta_t = [\eta_{1,t}, \dots, \eta_{N,t}]^T,$$

has entries $\eta_{i,t} = \eta / \left((1 + (1 - q) \log_q^T(w_{i,t})) \right) = \eta w_{i,t}^{q-1}$.

Remark 3. Assuming that learning rate is time variable and it is represented by a vector, i.e., $\eta \rightarrow \eta w_t^{1-\alpha}$, where $\eta_t = \eta w_t^{1-\alpha-\beta}$, the proposed MD update takes particular form derived and extensively experimentally tested in our recent publication [15], however using a different approach employing alpha-beta divergences [14,27].

4.3. MD and EG Using Schwämmle–Tsallis (ST) Entropy

Schwämmle and Tsallis proposed two-parameter entropy [51]

$$S_{q,q'}^{ST}(p) = \sum_{i=1}^W p_i \log_{q,q'}^{ST}(1/p_i), \quad q \neq 1, \quad q' \neq 1, \quad (38)$$

where the deformed logarithm referred to as the ST-logarithm or ST (q, q') -logarithm is defined as

$$\log_{q,q'}^{ST}(x) = \log_{q'}^T([x]_q) = \log_{q'}^T(e^{\log_q^T(x)}) = \frac{1}{1-q'} \left[\exp\left(\frac{1-q'}{1-q}(x^{1-q} - 1)\right) - 1 \right] \quad (39)$$

for $x > 0$ and $q \neq q'$ (typically $q > 1$ and $q' < 1$), where $[x]_q = \exp(\log_q(x))$ and its inverse function is formulated as a two-parameter deformed exponential, $\exp_{q,q'}(x)$

$$\exp_{q,q'}^{ST}(x) = \left[1 + \frac{1-q}{1-q'} \ln(1 + (1-q')x) \right]^{1/(1-q)}. \quad (40)$$

Note that if either parameter q or q' , or both, take the value of one, the above functions simplifies to Tsallis q -functions (logarithm and exponential), so we can write

$$\log_{q,1}^{ST}(x) = \log_{1,q}^{ST}(x) = \log_q^T(x), \quad \log_{1,1}^{ST}(x) = \ln(x) \quad (41)$$

$$\exp_{q,1}^{ST}(x) = \exp_{1,q}^{ST}(x) = \exp_q^T(x), \quad \exp_{1,1}^{ST}(x) = \exp(x). \quad (42)$$

In the special case for $q = q'$, we obtain

$$\begin{aligned} \log_{q,q}^{ST}(x) &= \frac{1}{1-q} \left[\exp(x^{1-q} - 1) - 1 \right] \\ &= \frac{1}{1-q} \left[\exp\left((1-q) \log_q^T(x)\right) - 1 \right], \quad x > 0, \quad q > 0 \end{aligned} \quad (43)$$

and

$$\exp_{q,q}^{ST}(x) = [\ln(x(1-q) + 1) + 1]_+^{1/(1-q)}. \quad (44)$$

The plots of the (q, q') -logarithm and (q, q') -exponential for various values of $q = q'$ are illustrated in Figure 4.

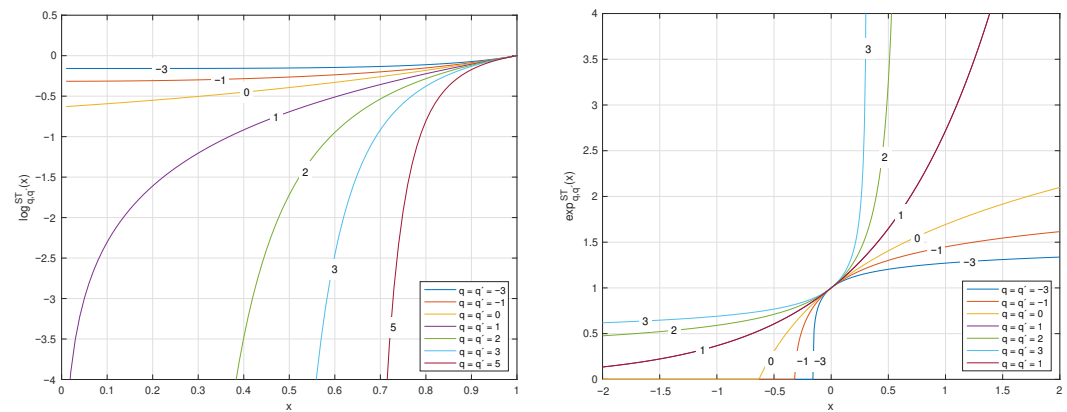


Figure 4. Plots of the (q, q') -logarithm and (q, q') -exponential functions for different values of the parameters in special cases when $q = q'$.

Moreover, it is easy to prove the following useful properties

$$\log_{q,q'}^{ST}(1/x) = -\log_{2-q,2-q'}^{ST}(x), \quad (45)$$

$$\frac{d \log_{q,q'}^{ST}(x)}{dx} = x^{-q} \exp\left(\frac{1-q'}{1-q}(x^{1-q} - 1)\right) > 0 \text{ for } x > 0, \quad \forall q, q', \quad q \neq 1. \quad (46)$$

Defining (q, q') -product as [52]:

$$x \otimes_{q,q'}^{ST} y = \exp_{q,q'}^{ST}(\log_{q,q'}^{ST}(x) + \log_{q,q'}^{ST}(y)) \quad (47)$$

we have the key formulas

$$\exp_{q,q'}^{ST}(x + y) = \exp_{q,q'}^{ST}(x) \otimes_{q,q'} \exp_{q,q'}^{ST}(y), \quad (48)$$

$$\exp_{q,q'}^{ST}(\log_{q,q'}^{ST}(x) + y) = x \otimes_{q,q'} \exp_{q,q'}^{ST}(y). \quad (49)$$

Let us consider now that the link function is defined componentwise as

$$f_{q,q'}(\mathbf{w}) = \log_{q,q'}^{ST}(\mathbf{w}). \quad (50)$$

The novel (q, q') -GEG update can take the following form

$$\mathbf{w}_{t+1} = \mathbf{w}_t \otimes_{q,q'} \exp_{q,q'}^{ST}(-\eta \nabla_{\mathbf{w}} L(\mathbf{w}_t)), \quad (51)$$

In this case, the update is more complex than in the previous case. An alternative approach is to apply the MMD/NG Formula (10):

$$\mathbf{w}_{t+1} = \left[\mathbf{w}_t - \eta \operatorname{diag} \left\{ \mathbf{w}^q \odot \exp \left(\frac{1-q'}{1-q} (1 - \mathbf{w}^{1-q}) \right) \right\} \nabla_{\mathbf{w}} L(\mathbf{w}_t) \right]_+ \quad (52)$$

which can be written equivalently in a scalar form as

$$w_{i,t+1} = \left[w_{i,t} - \eta w_{i,t}^q \exp \left(\frac{1-q'}{1-q} (1 - w_{i,t}^{1-q}) \right) \frac{\partial L(\mathbf{w}_t)}{\partial w_i} \right]_+ \quad (53)$$

Remark 4. Extension to Three Parameters (q, q', r) -Logarithm. Note that by using definition $[x]_q = \exp(\log_q^T(x))$ we can write the ST logarithm in a compact form

$$\log_{q,q'}^{ST}(x) = \frac{[x]_q^{1-q'} - 1}{1 - q'} = \frac{[\exp(\log_q^T(x))]^{1-q'} - 1}{1 - q'} \quad (54)$$

Analogously, we can define

$$[x]_{q,q'} = \exp(\log_{q,q'}^{ST}(x)). \quad (55)$$

Hence, we can formulate a three-parameter logarithm as proposed in [53]

$$\log_{q,q',r}^{CC}(x) = \frac{([x]_{q,q'}^{1-q'})^{1-r} - 1}{1 - r} = \frac{(\exp(\log_{q,q'}^{ST}(x)))^{1-r} - 1}{1 - r} \quad (56)$$

The plots of the (q, q', r) -logarithm and (q, q', r) -exponential for coincident values of q , q' , and r are illustrated in Figure 5.

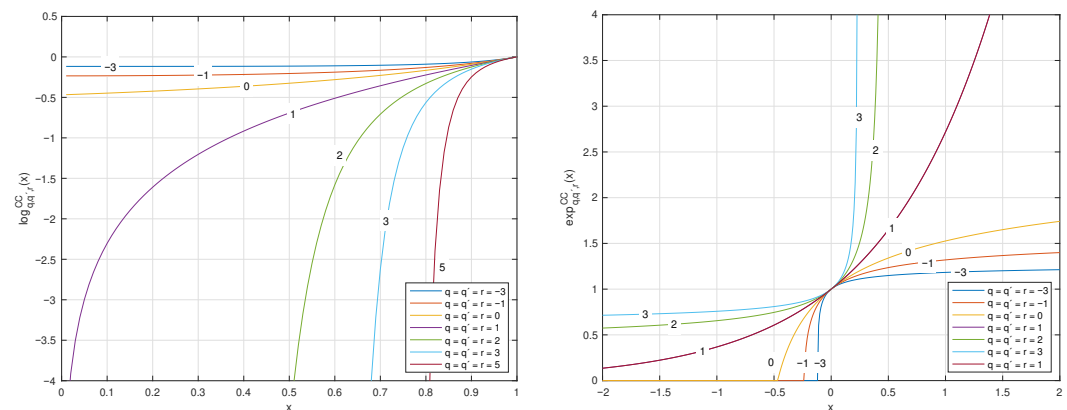


Figure 5. Plots of the (q, q', r) -logarithm and (q, q', r) -exponential functions when the parameters are coincident $q = q' = r$.

In a similar way, as in the previous section, we can drive MD updates using as a link function the above-defined three-parameter logarithm.

5. MD and GEG Using the Kaniadakis Entropy and Its Extensions and Generalizations

5.1. Basic Properties of κ -Logarithm and κ -Exponential

An entropic structure emerging in the context of special relativity, is the one defined by Kaniadakis [36,37] as follows

$$S_\kappa(p_i) = \sum_i p_i \log_\kappa^K(1/p_i), \quad (57)$$

where a deformed κ -logarithm referred to as the Kaniadakis κ -logarithm is defined as [36,37]:

$$\log_\kappa^K(x) = \begin{cases} \frac{x^\kappa - x^{-\kappa}}{2\kappa} = \frac{1}{\kappa} \sinh(\kappa \ln(x)) & \text{if } x > 0 \text{ and } 0 < \kappa^2 < 1 \\ \ln(x) & \text{if } x > 0 \text{ and } \kappa = 0. \end{cases} \quad (58)$$

The inverse function of the Kaniadakis κ -logarithm is the deformed exponential function $\exp_{\kappa}^K(x)$, represented as

$$\begin{aligned} \exp_{\kappa}^K(x) &= \exp\left(\int_0^x \frac{dy}{\sqrt{1+\kappa^2 y^2}}\right) \\ &= \begin{cases} \left(\sqrt{1+\kappa^2 x^2} + \kappa x\right)^{1/\kappa} = \exp\left(\frac{1}{\kappa} \operatorname{arsinh}(\kappa x)\right) & -1 < \kappa < 1 \\ \exp(x) & \kappa = 0. \end{cases} \end{aligned} \quad (59)$$

The plots of the κ -logarithm and κ -exponential functions for various values of κ are illustrated in Figures 6 and 7.

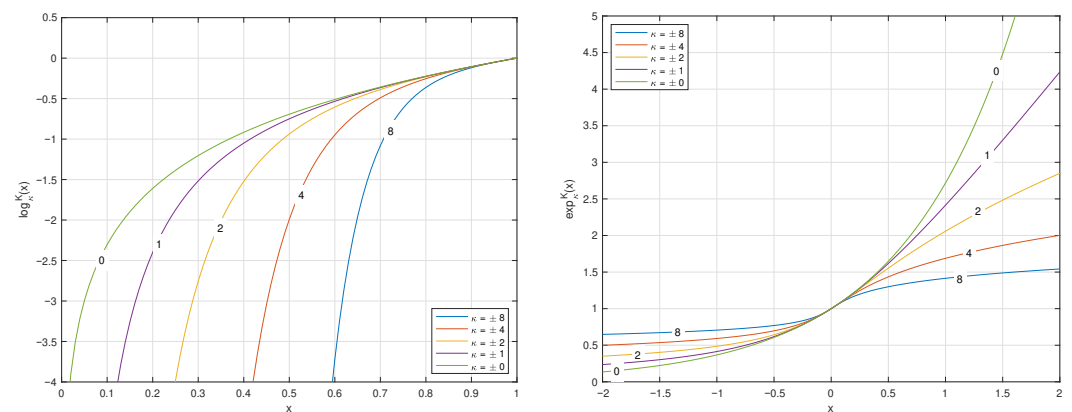


Figure 6. Plots of the κ -logarithm and κ -exponential functions for different values of the parameter κ .

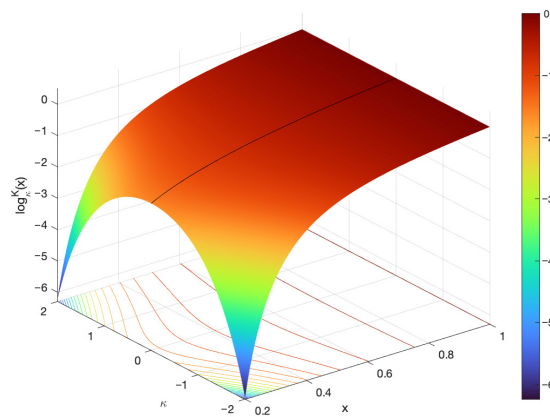


Figure 7. Surface plots of the κ -logarithm. The black continuous line represents the reference of the classical logarithm $\ln(x)$, which is obtained for $\kappa = 0$.

Note that the Kaniadakis logarithm can also be expressed in terms of the Tsallis logarithm as

$$\log_{\kappa}^K(x) = \frac{\log_{1+\kappa}^T(x) + \log_{1-\kappa}^T(x)}{2}. \quad (60)$$

These functions have the following fundamental and useful properties [37,41]:

$$\log_{\kappa}^K(1) = 1, \log_{\kappa}^K(0^+) = -\infty, \log_{\kappa}^K(\infty) = +\infty, \quad (61)$$

$$\log_{\kappa}^K(1/x) = -\log_{\kappa}^K(x), \quad (62)$$

$$\log_{\kappa}^K(x^{\lambda}) = \lambda \log_{\kappa}^K(x), \quad (63)$$

$$\log_{\kappa}^K(xy) = y^{\kappa} \log_{\kappa}^K(x) + x^{-\kappa} \log_{\kappa}^K(y), \quad (64)$$

$$\log_{\kappa}(\exp(x)) = (1/\kappa) \sinh(\kappa x), \quad (65)$$

$$\ln(\exp_{\kappa}(x)) = (1/\kappa) \operatorname{arsinh}(\kappa x), \quad (66)$$

$$\frac{\partial \log_{\kappa}^K(x)}{\partial x} = \frac{x^{\kappa} + x^{-\kappa}}{2x} = \frac{\cosh[\kappa \ln(x)]}{x} > 0, \quad (67)$$

$$\frac{\partial^2 \log_{\kappa}^K(x)}{\partial x^2} = \frac{\kappa - 1}{2} x^{\kappa-2} - \frac{\kappa + 1}{2} x^{-\kappa-2} < 0 \text{ for } \kappa \in [-1, 1]. \quad (68)$$

The last two properties indicate that the Kaniadakis κ -logarithm is monotonically increasing for any value of κ , and for $|\kappa| < 1$, it is additionally a strictly concave function.

The κ -logarithm can be approximated as a power series

$$\log_{\kappa}(x) \approx \ln(x) + \frac{1}{3!} \kappa^2 [\ln(x)]^3 + \frac{1}{5!} \kappa^4 [\ln(x)]^5 + \frac{1}{7!} \kappa^6 [\ln(x)]^7 \dots, \quad x > 0. \quad (69)$$

Furthermore, it is important to note that applying the Taylor series expansion of the κ -exponential we can obtain a simple approximation as

$$\exp_{\kappa}^K(x) = 1 + x + \frac{1}{2!} x^2 + \frac{1}{3!} (1 - \kappa^2) x^3 + \frac{1}{4!} (1 - 4\kappa^4) x^4 + \dots \quad (70)$$

$$= \exp(x) - \frac{1}{3!} \kappa^2 x^3 - \frac{1}{4!} 4\kappa^4 x^4 + \dots \quad (71)$$

Two notable features of the κ -exponential function are that it asymptotically approaches a regular exponential function for small x and asymptotically approaches a power law for large value $|x|$ [37,54]. Specifically,

$$\lim_{x \rightarrow 0} \exp_{\kappa}(x) \sim \exp(x), \quad (72)$$

$$\lim_{x \rightarrow \pm\infty} \exp_{\kappa}(x) \sim |2\kappa x|^{\pm 1/|\kappa|}. \quad (73)$$

The κ -exponential function has the following basic properties [36,37,41]

$$\exp_{\kappa}^K(0) = 1, \exp_{\kappa}^K(-\infty) = 0^+, \exp_{\kappa}^K(\infty) = +\infty, \quad (74)$$

$$\exp_{\kappa}^K(-\infty) = 0^+, \exp_{\kappa}^K(+\infty) = +\infty, \quad (75)$$

$$\exp_{\kappa}^K(x) \exp_{\kappa}^K(-x) = 1, \quad (76)$$

$$[\exp_{\kappa}^K(x)]^r = \exp_{\kappa/r}^K(rx), \quad r \in \mathbb{R}, \quad (77)$$

$$\frac{\partial \exp_{\kappa}^K(x)}{\partial x} > 0, \quad (78)$$

$$\frac{\partial^2 \exp_{\kappa}^K(x)}{\partial x^2} > 0 \text{ for } \kappa \in [-1, 1]. \quad (79)$$

The last two properties mean that the Kaniadakis κ -exponential is a monotonically increasing and convex function for a specific range of the parameter κ .

The property (79) emerges as a particular case of the more general one

$$\exp_{\kappa}^K(x) \exp_{\kappa}^K(y) = \exp_{\kappa}^K(x) \oplus_{\kappa} \exp_{\kappa}^K(y), \quad (80)$$

$$\log_{\kappa}^K(xy) = \log_{\kappa}^K(x) \oplus_{\kappa} \log_{\kappa}^K(y), \quad (81)$$

where κ -addition is defined as

$$x \oplus_{\kappa} y = x \sqrt{1 + \kappa^2 y^2} + y \sqrt{1 + \kappa^2 x^2} \quad (82)$$

$$\approx x + y + \frac{\kappa^2}{2}(xy^2 + x^2y) - \frac{\kappa^4}{8}(xy^4 + x^4y) \dots \quad (83)$$

By defining and evaluating the κ -product

$$x \otimes_{\kappa} y = \exp_{\kappa}[\log_{\kappa}(x) + \log_{\kappa}(y)] \quad (84)$$

$$= \left(\frac{x^{\kappa} - x^{-\kappa}}{2} + \frac{y^{\kappa} - y^{-\kappa}}{2} + \sqrt{1 + \left(\frac{x^{\kappa} - x^{-\kappa}}{2} + \frac{y^{\kappa} - y^{-\kappa}}{2} \right)^2} \right)^{1/\kappa} \quad (85)$$

$$= \exp\left(\frac{1}{\kappa} \operatorname{arsinh}((x^{\kappa} - x^{-\kappa} + y^{\kappa} - y^{-\kappa})/2)\right) \quad (86)$$

$$= \frac{1}{\kappa} \sinh\left(\frac{1}{\kappa} \operatorname{arsinh}(\kappa x) \operatorname{arsinh}(\kappa y)\right), \quad (87)$$

we have the key formulas for our MD application

$$\exp_{\kappa}^K(x + y) = \exp_{\kappa}^K(x) \otimes_{\kappa} \exp_{\kappa}^K(y), \quad (88)$$

$$\exp_{\kappa}^K(\log_{\kappa}^K(x) + y) = x \otimes_{\kappa} \exp_{\kappa}^K(y). \quad (89)$$

Appendix A.2 gives an overview of the κ -algebra and calculus.

5.2. MD and GEG Using the Kaniadakis Entropy and κ -Logarithm

Let us assume that the link function in Mirror Descent can take the following componentwise form

$$f_{\kappa}(\mathbf{w}_t) = \log_{\kappa}^K(\mathbf{w}_t), \quad \mathbf{w} = [w_{1,t}, \dots, w_{N,t}]^T \in \mathbb{R}_+^N. \quad (90)$$

Note that since the first derivative of $\log_{\kappa}^K(\mathbf{w})$ is positive and second derivative is negative, the link function is a (componentwise) increasing function and additionally a concave function for $\kappa \in (-1, 1]$. In this case the generating function:

$$F_{\kappa}(\mathbf{w}_t) = \sum_{i=1}^N \frac{1}{2\kappa} \left[\frac{w_{i,t}^{1+\kappa}}{1+\kappa} - \frac{w_{i,t}^{1-\kappa}}{1-\kappa} \right]. \quad (91)$$

Taking into account Formula (6), we obtain a novel κ -GEG update

$$\boxed{\mathbf{w}_{t+1} = \mathbf{w}_t \otimes_{\kappa} \exp_{\kappa}^K(-\eta \nabla_{\mathbf{w}} L(\mathbf{w}_t))} \quad (92)$$

where \otimes_{κ} is a κ -product defined by Equation (87) and is performed componentwise.

5.3. Two-Parameter Logarithms Based on Generalized Kaniadakis–Lissia–Scarfone Entropy

Another important generalized entropy has the following form

$$S_{\kappa,r}(p) = - \sum_{i=1}^W (p_i)^{r+1} \frac{p_i^{\kappa} - p_i^{-\kappa}}{2\kappa} = - \sum_{i=1}^W p_i \log_{\kappa,r}(p_i), \quad (93)$$

which was introduced by Sharma, Taneja and Mittal (STM) in [55–57], and also investigated, independently, by Kaniadakis, Lissia and Scarfone (KLS) in [38,39,58].

Equation (93) mimics the expression of the Boltzmann–Gibbs entropy by replacing the standard natural logarithm $\ln(x)$ with the two-parametric deformed logarithm $\log_{\kappa,r}(x)$ defined as

$$\log_{\kappa,r}(x) = x^r \frac{x^\kappa - x^{-\kappa}}{2\kappa}, \quad x > 0, \quad r \in \mathbb{R}, \quad \text{for } -|\kappa| \leq r \leq 1/2 - |1/2 - |\kappa||. \quad (94)$$

The surface plots of the (κ, r) -logarithm for various values of hyperparameters κ and r are illustrated in Figures 8 and 9.

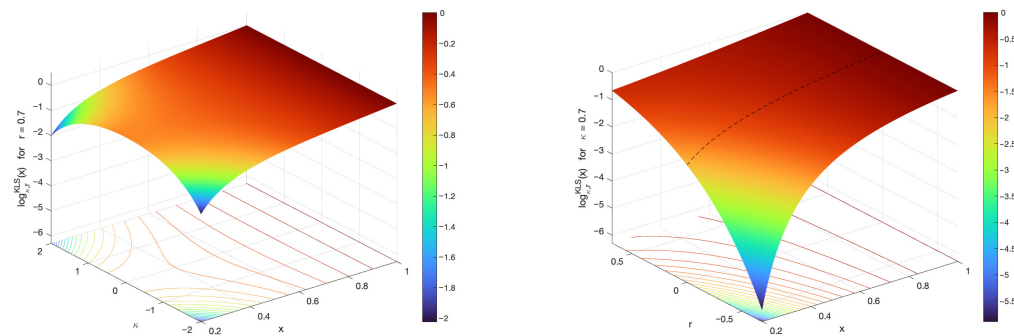


Figure 8. Surface plots of the (κ, r) -logarithm for various values of hyperparameters κ and r . The left-hand-side figure illustrates the (κ, r) -logarithm in terms of κ and x when $r = 0.7$. The right-hand-side figure illustrates the (κ, r) -logarithm, now in terms of r and x , when $\kappa = 0.7$. The black dashed line coincides with the κ -logarithm for $\kappa = 0.7$, since in this case $r = 0$.

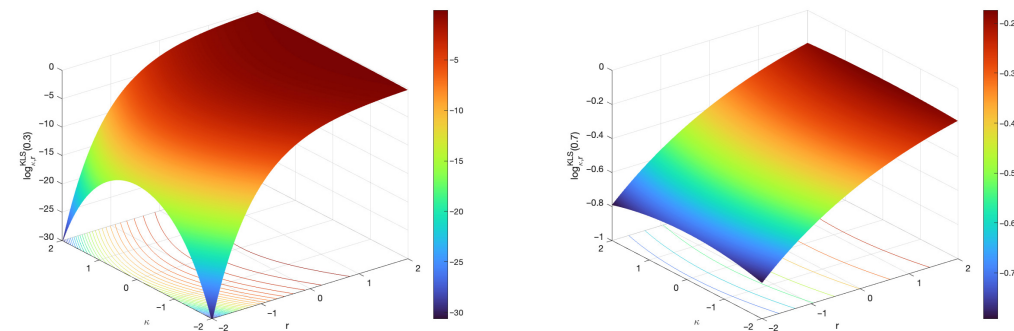


Figure 9. Surface plots of the (κ, r) -logarithm in terms of the hyperparameters κ and r , when $x \in \{0.3, 0.7\}$. From the drawings, it is apparent how the changes of the hyperparameter r have a much stronger influence in the magnification of the response, in comparison with the changes in the hyperparameter κ that correspond with more subtle elongations. The figure on the left-hand-side evaluates the (κ, r) -logarithm for $x = 0.3$, whereas the figure on the right-hand-side evaluates it for $x = 0.7$.

The (κ, r) -logarithm can be expressed via the Kaniadakis κ -logarithm and the Tsallis q -logarithm

$$\log_{\kappa,r}(x) = x^r \frac{x^\kappa - x^{-\kappa}}{2\kappa} = x^r \log_\kappa^K(x) = x^{r-\kappa} \log_{1-2\kappa}^T(x). \quad (95)$$

Obviously, for $r = 0$ the (κ, r) -logarithm simplifies to a Kaniadakis κ -logarithm, and for $r = \pm|\kappa|$ one recovers the Tsallis q -logarithm with $q = 1 \mp 2|\kappa|$, ($0 < q < 2$).

By introducing a new parameter $\omega = r/\kappa$ or replacing $r = \omega \kappa$ we can represent the logarithm as

$$\log_{\kappa,\omega}(x) = \frac{x^{\kappa(\omega+1)} - x^{\kappa(\omega-1)}}{2\kappa}, \quad (96)$$

which for $\omega = 0$ simplifies to κ -logarithm and with $\omega = 1$ and $\kappa = (1 - q)/2$ we have a q -logarithm. This formula indicates that this logarithm smoothly interpolates between a Kaniadakis logarithm and Tsallis logarithm.

Summarizing, the (κ, r) -logarithm can be described as follows [38,39]

$$\log_{\kappa,r}^{KLS}(x) = \begin{cases} \frac{x^{r+\kappa} - x^{r-\kappa}}{2\kappa} & \text{if } x > 0 \text{ for } r \in \mathbb{R}, \text{ and } -|\kappa| \leq r \leq |\kappa|, \\ \log_{\kappa}^K(x) = \frac{x^{\kappa} - x^{-\kappa}}{2\kappa} & \text{if } x > 0 \text{ for } r = 0, \kappa \in [-1, 1], \kappa \neq 0, \\ \log_q^T(x) = \frac{x^{1-q} - 1}{1-q}, & \text{if } x > 0 \text{ for } r = \kappa = (1-q)/2, q \neq 1, q > 0, \\ \ln(x) & \text{if } x > 0 \text{ and } r = \kappa = 0. \end{cases} \quad (97)$$

It should also be noted the (κ, r) -logarithm can be represented approximately by the following power series for relatively small κ :

$$\log_{\kappa,r}^{KLS} \approx \ln(x) + \frac{1}{2}(2r)[\ln(x)]^2 + \frac{1}{6}(\kappa^2 + 3r^2)[\ln(x)]^3 \dots \quad (98)$$

The (κ, r) logarithm has the following basic properties

$$\log_{\kappa,r}(1) = 0, \log_{\kappa,r}(0^+) = -\infty, \text{ for } r < |\kappa|, \log_{\kappa,r}(+\infty) = +\infty, \text{ for } r > -|\kappa|, \quad (99)$$

$$\log_{\kappa,r}(x) = -\log_{\kappa,-r}(1/x) = \log_{-\kappa,r}(x), \quad (99)$$

$$\log_{\kappa,r}(x^\lambda) = \lambda \log_{\lambda\kappa}(x), \quad (100)$$

$$\frac{\partial \log_{\kappa,r}(x)}{\partial x} > 0, \text{ for } -|\kappa| \leq r \leq |\kappa|, \quad (101)$$

$$\frac{\partial^2 \log_{\kappa,r}(x)}{\partial x^2} < 0 \text{ for } -|\kappa| \leq r \leq 1/2 - |1/2 - |\kappa||. \quad (102)$$

The last two properties indicate that the (κ, r) -logarithm is a strictly monotonically increasing function for $-|\kappa| \leq r \leq |\kappa|$ and additionally a concave function for $-|\kappa| \leq r \leq 1/2 - |1/2 - |\kappa||$.

Remark 5. Relation to the Euler Logarithm: It is interesting to note that for $r + k = a$ and $r - k = b$ the KLS κ, r -logarithm can be represented as the Euler logarithm [17,59]:

$$\log_{a,b}^{Eu}(x) = \frac{x^a - x^b}{a - b}, \quad x > 0, a < 0, 0 < b < 1, \quad (103)$$

which is related to the Borges–Rodity entropy [50,60].

Connection to the Schwämmle–Tsallis logarithm: By applying nonlinear transformation in (103) $x \rightarrow \exp(\log_q(x))$ and $a = 1 - q'$, $b = 0$, we obtain the Schwämmle–Tsallis logarithm (39).

Connection to the Mean Value Theorem: The function has deep connections to the Mean Value Theorem applied to power functions. For the power function $g(t) = x^t$, the Mean Value Theorem guarantees the existence of some parameter $c \in (a, b)$ such that $g'(c) = \frac{g(b) - g(a)}{b - a}$, which yields:

$$\log_{a,b}^{Eu}(x) = \frac{x^b - x^a}{b - a} = c \cdot x^{c-1} \cdot \ln(x) \text{ for } c \in (a, b). \quad (104)$$

Logarithmic Mean Connection: The function relates to the logarithmic mean $L(u, v) = \frac{u-v}{\ln u - \ln v}$ through the substitution $u = x^a$, $v = x^b$. These connections provide alternative computational approaches and theoretical insights.

Exponential Function Theory: The underlying structure connects to exponential function differentiation rules, where $\frac{d}{dt}x^t = x^t \ln(x)$, explaining the limiting behavior observed in the analysis.

Computational and numerical considerations: The numerical analysis reveals several important computational aspects:

1. *Numerical Stability: The KLS logarithm becomes increasingly stable as $x \rightarrow 1$, but exhibits potential numerical instability for x values far from unity.*
2. *Parameter Sensitivity: Small x values create higher sensitivity to parameter changes, requiring careful numerical handling.*
3. *Convergence Properties: The limiting behavior requires special computational treatment using L'Hôpital's rule.*

5.4. Exponential KLS Function and Its Properties

With the existence of $\exp_{\kappa,r}(x)$, the inverse function of $\log_{\kappa,r}(x)$ follows from its monotonicity in \mathbb{R} although an explicit expression, in general, cannot be given. In other words, the inverse function can not be expressed in a closed analytical form but it can be approximated and expressed, for example, in terms of the Lambert–Tsallis W_q -functions, which are the solution of equations $W_q(z)[1 + (1 - q)W_q(z)]_+^{1/(1-q)} = z$:

$$\exp_{\kappa,r}(x) = \left(\frac{W_{\frac{\lambda+1}{\lambda}}(\lambda \tilde{x}^{-\lambda})}{\lambda} \right)^{-1/(2\kappa)}, \quad (105)$$

where $\lambda = (2\kappa)/(r + \kappa)$, $\tilde{x} = 2\kappa x$ and W_q is the Lambert–Tsallis function [61].

Another much simpler approach is to use Lagrange's inversion theorem around 1 to obtain the following rough power series approximation (which may be sufficient for most of our applications):

$$\exp_{\kappa,r}(x) \approx 1 + x + \frac{1}{2}(1 - 2r)x^2 + \left(\frac{1}{6} - r + \frac{3}{2}r^2 - \frac{1}{6}\kappa^2 \right)x^3 + \dots \quad (106)$$

$$= \exp(x) - rx^2 + \left(\frac{3}{2}r^2 - r - \frac{1}{6}\kappa^2 \right)x^3 + O(x^4). \quad (107)$$

Hence, we can represent a (κ, r) -exponential as follows:

$$\exp_{\kappa,r}(x) = \begin{cases} \approx \exp(x) - rx^2 + \left(\frac{3}{2}r^2 - r - \frac{1}{6}\kappa^2 \right)x^3 & \text{for } r \in \mathbb{R}, \quad -|\kappa| \leq r \leq |\kappa|, \quad |\kappa| < 1, \\ \exp_{\kappa}^K(x) = [\kappa x + \sqrt{1 + \kappa^2 x^2}]^{1/\kappa} & \text{for } r = 0, \quad \kappa \in [-1, 1], \quad \kappa \neq 0, \\ \exp_q^T(x) = [1 + (1 - q)x]_+^{1/(1-q)} & \text{for } r = \kappa = (1 - q)/2, \quad q \neq 1, \\ \exp(x) & \text{for } r = \kappa = 0. \end{cases} \quad (108)$$

Furthermore, the (κ, r) -exponential function has the following fundamental properties:

$$\begin{aligned} \exp_{\kappa,r}(0) &= 1, \quad \exp_{\kappa,r}(-\infty) = 0^+, \text{ for } r < |\kappa|, \quad \exp_{\kappa,r}(\infty) = +\infty, \text{ for } r > -|\kappa|, \\ \exp_{\kappa,r}(x) \exp_{\kappa,r}(-x) &= 1, \end{aligned} \quad (109)$$

$$(\exp_{\kappa,r}(x))^\lambda = \exp_{\kappa/\lambda, r/\lambda}(\lambda x), \quad (110)$$

$$\frac{\partial \exp_{\kappa,r}(x)}{\partial x} > 0, \text{ for } -|\kappa| \leq r \leq |\kappa|, \quad (111)$$

$$\frac{\partial^2 \exp_{\kappa,r}(x)}{\partial x^2} > 0, \text{ for } -|\kappa| \leq r \leq 1/2 - |1/2 - |\kappa|| \quad (112)$$

The last two properties mean that the (κ, r) -exponential is a monotonically increasing convex function.

Two notable features of the (κ, r) -logarithm and exponential function are that it asymptotically approaches a regular exponential function for small x and asymptotically approaches a power law for large absolute x :

$$\lim_{x \rightarrow 0^+} \log_{\kappa,r}(x) \sim \frac{-1}{2|\kappa| x^{|\kappa|+r}}, \quad (113)$$

$$\lim_{x \rightarrow +\infty} \log_{\kappa,r}(x) \sim \frac{x^{|\kappa|+r}}{2\kappa}, \quad (114)$$

$$\lim_{x \rightarrow 0} \exp_{\kappa,r}(x) \sim \exp(x), \quad (115)$$

$$\lim_{x \rightarrow \pm\infty} \exp_{\kappa,r}(x) \sim |2\kappa x|^{1/(r \pm |\kappa|)}. \quad (116)$$

By defining

$$x \otimes_{\kappa,r} y = \exp_{\kappa,r}[\log_{\kappa,r}(x) + \log_{\kappa,r}(y)] \quad (117)$$

we have the key formulas for our MD (GEG) implementations

$$\exp_{\kappa,r}(x + y) = \exp_{\kappa,r}(x) \otimes_{\kappa,r} \exp_{\kappa,r}^K(y), \quad (118)$$

$$\exp_{\kappa,r}(\log_{\kappa,r}^K(x) + y) = x \otimes_{\kappa,r} \exp_{\kappa,r}^K(y). \quad (119)$$

Let us assume that the link function is defined as $f(w) = \log_{\kappa,r}(w)$ and its inverse (if approximated version is accepted) $f^{(-1)}(w) = \exp_{\kappa,r}(w)$, then using a general MD Formula (6), and fundamental properties described above, we obtain a general MD formula employing a wide family of deformed logarithms arising from group entropies or trace-form entropies:

$$\boxed{w_{t+1} = \exp_{\kappa,r}[\log_{\kappa,r}(w_t) - \eta_t \nabla L(w_t)] = w_t \otimes_{\kappa,r} \exp_{\kappa,r}(-\eta_t \nabla L(w_t))} \quad (120)$$

where $\otimes_{\kappa,r}$ -multiplication is defined/determined as follows

$$x \otimes_{\kappa,r} y = \exp_{\kappa,r}(\log_{\kappa,r}(x) + \log_{\kappa,r}(y)). \quad (121)$$

Alternatively, due to some complexity of computing precisely $\exp_{\kappa,r}(x)$, in the general case, we can use formula MMD/NG (10) to derive a quite flexible and general NG gradient update:

$$\boxed{w_{t+1} = \left[w_t - \eta_t \text{diag} \left\{ \left(\frac{\partial \log_{\kappa,r}(w_t)}{\partial w_t} \right)^{-1} \right\} (\nabla L(w_t)) \right]_+} \quad (122)$$

where $\text{diag} \left\{ \left(\frac{\partial \log_{\kappa,r}(\mathbf{w}_t)}{\partial \mathbf{w}_t} \right)^{-1} \right\}$ is a positive-definite diagonal matrix, with the diagonal entries

$$\left(\frac{\partial \log_{\kappa,r}(\mathbf{w}_t)}{\partial w_{i,t}} \right)^{-1} = \frac{2\kappa}{(r + \kappa)w_{i,t}^{r+\kappa-1} - (r - \kappa)w_{i,t}^{r-\kappa-1}} > 0, \quad -|\kappa| \leq r \leq |\kappa|. \quad (123)$$

6. Generalization and Normalization of Mirror Descent

Summarizing, all of the GEG updates proposed in this paper can be presented in normalized form (by projecting to unit simplex) in the following general and flexible form

$$\tilde{\mathbf{w}}_{t+1} = \mathbf{w}_t \otimes_D \exp_D \left(-\boldsymbol{\eta}_t \odot \nabla \hat{L}(\mathbf{w}_t) \right) \quad (\text{Generalized multiplicative update}), \quad (124)$$

$$\mathbf{w}_{t+1} = \frac{\tilde{\mathbf{w}}_{t+1}}{\|\tilde{\mathbf{w}}_{t+1}\|_1}, \quad (\text{Projection to unit simplex}) \quad (125)$$

where $\exp_D(x)$ ($\log_D(x)$) is a generalized exponential (logarithm), $\hat{L}(\mathbf{w}_t) = L(\mathbf{w}_t)/\|\mathbf{w}_t\|_1$ is normalized/scaled loss function, $\boldsymbol{\eta}_t$ is a vector of the learning rates, $\nabla \hat{L}(\mathbf{w}_t) = \nabla L(\mathbf{w}_t) - (\mathbf{w}^T \nabla L(\mathbf{w}_t))\mathbf{1}$ and the generalized D -multiplication is computed as

$$\mathbf{w}_t \otimes_D \exp_D(\mathbf{g}_t) = \exp_D(\log_D(\mathbf{w}_t) + \mathbf{g}_t). \quad (126)$$

Here, $\log_D(\mathbf{w}_t)$ and its inverse $\exp_D(\mathbf{w}_t)$ mean any deformed logarithm and exponential investigated in this paper (i.e., the Tsallis, Kaniadakis, ST, KLS and KS exponential/logarithm).

Alternatively, when the inverse function can not be precisely computed, we can use an MMD/NG additive natural gradient Formula (10), which is expressed in general as

$$\mathbf{w}_{t+1} = \left[\mathbf{w}_t - \eta \text{diag} \left\{ \left(\frac{d \log_D(\mathbf{w}_t)}{d \mathbf{w}_t} \right)^{-1} \right\} \nabla_w \hat{L}(\mathbf{w}_t) \right]_+, \quad (127)$$

$$\mathbf{w}_{t+1} = \frac{\tilde{\mathbf{w}}_{t+1}}{\|\tilde{\mathbf{w}}_{t+1}\|_1}, \quad \mathbf{w}_t \in \mathbb{R}_+^N, \quad \forall t \quad (128)$$

where $\text{diag} \left\{ \left(\frac{d \log_D(\mathbf{w})}{d \mathbf{w}} \right)^{-1} \right\} = \text{diag} \left\{ \left(\frac{d \log_D(\mathbf{w})}{d w_1} \right)^{-1}, \dots, \left(\frac{d \log_D(\mathbf{w})}{d w_N} \right)^{-1} \right\}$ is a diagonal positive-definite matrix.

7. Conclusions and Discussion

This study establishes a comprehensive framework for applying trace-form entropies and associated deformed logarithms in both Mirror Descent and equivalently Generalized Exponentiated Gradient algorithms. By systematically exploring trace-form entropies, especially Tsallis, Kaniadakis, Scarfone, and Sharma–Taneja–Mittal forms as regularization terms, we unveil new families of mirror gradient descent algorithms that can be tailored to the optimization landscape through suitably chosen hyperparameters. The adoption of these generalized entropies opens the door to obtaining advantageous properties such as improved convergence rates, robustness against vanishing/exploding gradients, and inherent flexibility for handling non-Euclidean geometries. Table 1 summarizes the main results obtained from our study and lists the generalized exponentiated gradient update induced by the deformed exponential functions corresponding to the deformed logarithms used to define the various trace-form entropies.

Table 1. Overview of the generalized exponentiated gradient (GEG) updates.

Entropy	Deformed Exponential	MD/GEG Update
Shannon	$\exp(x) = \sum_{i=0}^{\infty} \frac{x^i}{i!}$	$\mathbf{w}_{t+1} = \mathbf{w}_t \odot \exp(-\eta \nabla L(\mathbf{w}_t))$ (EG)
Tsallis	$\exp_q^T(x) = \begin{cases} [1 + (1-q)x]_+^{1/(1-q)} & q \neq 1 \\ \exp(x) & q = 1 \end{cases}$	$\mathbf{w}_{t+1} = \exp_q^T[\log_q^T(\mathbf{w}_t) \otimes_q (-\eta \nabla_w L(\mathbf{w}_t))]$ $= \mathbf{w}_t \otimes_q \exp_q^T(-\eta \nabla_w L(\mathbf{w}_t))$ (q-GEG)
Schwämmle-Tsallis	$\exp_{q,q'}^{\text{ST}}(x) = \left[1 + \frac{1-q}{1-q'} \ln(1 + (1-q')x)\right]^{1/(1-q)}$	$\mathbf{w}_{t+1} = \left[\mathbf{w}_t - \eta \text{diag}\left\{\mathbf{w}^q \odot \exp\left(\frac{1-q'}{1-q}(1 - \mathbf{w}^{1-q})\right)\right\} \nabla_w L(\mathbf{w}_t)\right]_+$
Kaniadakis	$\exp_{\kappa}^K(x) = \begin{cases} \text{arcsinh}(\kappa x) & -1 < \kappa < 1 \\ \exp(x) & \kappa = 0. \end{cases}$	$\mathbf{w}_{t+1} = \mathbf{w}_t \otimes_{\kappa} \exp_{\kappa}^K(-\eta \nabla_w L(\mathbf{w}_t))$ (κ -GEG)
KLS	$\exp_{\kappa,r}(x)$	$\mathbf{w}_{t+1} = \left[\mathbf{w}_t - \eta_t \text{diag}\left\{\left(\frac{\partial \log_{\kappa,r}(\mathbf{w}_t)}{\partial \mathbf{w}_t}\right)^{-1}\right\} (\nabla L(\mathbf{w}_t))\right]_+$
Generic	$\exp_D(x)$	$\mathbf{w}_{t+1} = \left[\mathbf{w}_t - \eta \text{diag}\left\{\left(\frac{d \log_D(\mathbf{w}_t)}{d \mathbf{w}_t}\right)^{-1}\right\} \nabla_w \hat{L}(\mathbf{w}_t)\right]_+$ $\mathbf{w}_{t+1} = \frac{\tilde{\mathbf{w}}_{t+1}}{\ \tilde{\mathbf{w}}_{t+1}\ _1}$

The theoretical developments presented not only unify additive and multiplicative gradient update rules via Bregman divergences but also pave the way for designing robust machine learning algorithms that have the ability to adapt precisely to the structure of training data distributions via hyperparameters. Future work will investigate broader classes of entropic functions, extending the framework to non-convex and stochastic optimization settings, applying the proposed approach to practical problems, and performing systematic comparisons through computer simulation experiments.

Author Contributions: Conceptualization, A.C.; writing—original draft, A.C. and T.T.; editing and visualization, S.C. and F.N.; validation, F.N. and S.C.; formal analysis, A.C. T.T. F.N. and S.C.; supervision and project administration, A.C. All authors have read and agreed to the published version of the manuscript.

Funding: S. Cruces was supported in part by the MICIU/AEI/10.13039/501100011033 under Grant PID2021-123090NB-I00, in part by ERDF/EU.

Data Availability Statement: No new data were created or analyzed in this study due theoretical character of this study.

Acknowledgments: We wish to express our sincere appreciation to Shun-ichi Amari for his profound influence and invaluable contributions and scientific collaborations. His expertise proved instrumental in shaping our collective understanding and continues to inspire our future endeavors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Deformed Algebra and Calculus

Appendix A.1. q -Algebra and Calculus

In this section we briefly summarize q -Algebra [42,47]:

- q -sum: $x \oplus_q y = x + y + (1-q)xy$
- Neutral element of q -sum: $x \oplus_q 0 = 0 \oplus_q x = x$
- q -subtraction: $x \ominus_q y = \frac{x-y}{1+(1-q)y}$, $y \neq -1/(1-q)$
- q -product: $x \otimes_q y = [x^{1-q} + y^{1-q} + 1]_+^{1/(1-q)}$
- Neutral element of q -product: $x \otimes_q 1 = 1 \otimes_q x = x$
- q -division: $x \oslash_q y = x \otimes_q (1/y) = [x^{1-q} - y^{1-q} + 1]_+^{1/(1-q)}$

- Inverse of q -product $1 \oslash_q x = [2 - x^{1-q}]_+^{1/(1-q)}$.

Appendix A.2. κ -Algebra and Calculus

In this section we briefly summarize κ -Algebra, especially κ -product properties [36–39]:

- κ -sum: $x \oplus_\kappa y = x\sqrt{1 + \kappa^2 y^2} + y\sqrt{1 + \kappa^2 x^2}$,
- Neutral element of κ -sum: $x \oplus_\kappa 0 = 0 \oplus_\kappa x = x$,
- κ -subtraction: $x \ominus_\kappa y = x\sqrt{1 + \kappa^2 y^2} - y\sqrt{1 + \kappa^2 x^2}$,
- κ -product: $x \otimes_\kappa y = \exp\left(\frac{1}{\kappa} \operatorname{arsinh}((x^\kappa - x^{-\kappa} + y^\kappa - y^{-\kappa})/2)\right)$,
- Admits the unity as a neutral element of κ -product: $x \otimes_\kappa 1 = 1 \otimes_\kappa x = x$.
- κ -product is commutative: $x \otimes_\kappa y = y \otimes_\kappa x$
- κ -product is associative: $(x \otimes_\kappa y) \otimes_\kappa z = x \otimes_\kappa (y \otimes_\kappa z)$
- The inverse element of x is $1/x$, i.e., $x \otimes_\kappa (1/x) = 1$
- κ -division: $x \oslash_\kappa y = x \otimes_\kappa (1/y) = \exp\left(\frac{1}{\kappa} \operatorname{arsinh}((x^\kappa - x^{-\kappa} - y^\kappa + y^{-\kappa})/2)\right)$
- Inverse of κ -product $1 \oslash_\kappa x = x^{-1}$.

References

1. Nemirovsky, A.; Yudin, D.B. *Problem Complexity and Method Efficiency in Optimization*; John Wiley and Sons: Hoboken, NJ, USA, 1983. <https://doi.org/10.1137/1027074>.
2. Amid, E.; Warmuth, M.K. Reparameterizing Mirror Descent as Gradient Descent. In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20), Vancouver, BC, Canada, 6–12 December 2020; Curran Associates Inc.: Red Hook, NY, USA, 2020; pp. 8430–8439.
3. Amid, E.; Warmuth, M.K. Winnowing with Gradient Descent. In Proceedings of the 33rd International Conference on Algorithmic Learning Theory, PMLR 125, Graz, Austria, 9–12 July 2020; pp. 163–182.
4. Ghai, U.; Hazan, E.; Singer, Y. Exponentiated Gradient Meets Gradient Descent. In Proceedings of the 31st International Conference on Algorithmic Learning Theory, PMLR 117, San Diego, CA, USA, 8–11 February 2020; pp. 386–407. <https://doi.org/10.48550/arXiv.1902.01903>.
5. Shalev-Shwartz, S. Online learning and online convex optimization. *Found. Trends Mach. Learn.* **2011**, *4*, 107–194.
6. Beck, A.; Teboulle, M. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Oper. Res. Lett.* **2003**, *31*, 167–175.
7. Amari, S. Natural gradient works efficiently in learning. *Neural Comput.* **1998**, *10*, 251–276.
8. Amari, S. *Information Geometry and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 194.
9. Amari, S. Information Geometry and Its Applications: Convex Function and Dually Flat Manifold. In *Emerging Trends in Visual Computing*; Nielsen, F., Ed.; Springer Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2009; pp. 75–102.
10. Amari, S. Alpha-divergence is unique, belonging to both f -divergence and Bregman divergence classes. *IEEE Trans. Inf. Theory* **2009**, *55*, 4925–4931.
11. Amari, S.; Cichocki, A. Information geometry of divergence functions. *Bull. Pol. Acad. Sci.* **2010**, *58*, 183–195.
12. Amid, E.; Nielsen, F.; Nock, R.; Warmuth, M.K. Optimal transport with tempered exponential measures. *Proc. AAAI Conf. Artif. Intell.* **2024**, *38*, 10838–10846.
13. Raskutti, G.; Mukherjee, S. The information geometry of mirror descent. *IEEE Trans. Inf. Theory* **2015**, *61*, 1451–1457.
14. Cichocki, A.; Cruces, S.; Amari, S.I. Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. *Entropy* **2014**, *13*, 134–170.
15. Cichocki, A.; Cruces, S.; Sarmiento, A.; Tanaka, T. Generalized Exponentiated Gradient Algorithms and Their Application to On-Line Portfolio Selection. *IEEE Access* **2024**, *12*, 197000–197020.
16. Kainth, A.S.; Wong, T.-K.L.; Rudzicz, F. Conformal mirror descent with logarithmic divergences. *Inf. Geom.* **2024**, *7* (Suppl. 1), 303–327.
17. Cichocki, A. Generalized Exponentiated Gradient Algorithms Using the Euler Two-Parameter Logarithm. *arXiv* **2025**, arXiv:2502.17500.
18. Cichocki, A. Mirror Descent Using the Tempesta Generalized Multi-parametric Logarithms. *arXiv* **2025**, arXiv:2506.13984.

19. Helmbold, D.P.; Schapire, R.E.; Singer, Y.; Warmuth, M.K. On-line Portfolio Selection Using Multiplicative Updates. *Math. Financ.* **1998**, *8*, 325–347. <https://doi.org/10.1111/1467-9965.00058>.
20. Kivinen, J.; Warmuth, M.K. Exponentiated Gradient versus Gradient Descent for Linear Predictors. *Inf. Comput.* **1997**, *132*, 1–63. <https://doi.org/10.1006/inco.1996.2612>.
21. Kivinen, J.; Warmuth, M.K. Additive Versus Exponentiated Gradient Updates for Linear Prediction. In Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing, Las Vegas, NV, USA, 29 May–1 June 1995; pp. 209–218. <https://doi.org/10.1145/225058.225121>.
22. Bregman, L. The relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming. *Comp. Math. Phys. USSR* **1967**, *7*, 200–217.
23. Burachik, R.S.; Dao, M.N.; Lindstrom, S.B. The generalized Bregman distance. *SIAM J. Optim.* **2021**, *31*, 404–424.
24. Martinez-Legaz, J.E.; Tamadoni Jahromi, M.; Naraghirad, E. On Bregman-type distances and their associated projection mappings. *J. Optim. Theory Appl.* **2022**, *193*, 107–117.
25. Nielsen, F.; Nock, R. Generalizing skew Jensen divergences and Bregman divergences with comparative convexity. *IEEE Signal Process. Lett.* **2017**, *24*, 1123–1127.
26. Nock, R.; Nielsen, F. Bregman divergences and surrogates for learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *31*, 2048–2059.
27. Cichocki, A.; Amari, S.I. Families of α - β - and γ -divergences: Flexible and robust measures of similarities. *Entropy* **2010**, *12*, 1532–1568.
28. Cichocki, A.; Zdunek, R.; Phan, A.H.; Amari, S.I. Multiplicative Iterative Algorithms for NMF with Sparsity Constraints. In *Nonnegative Matrix and Tensor Factorizations*; Chapter 3; John Wiley and Sons: Hoboken, NJ, USA, 2009; pp. 131–202. <https://doi.org/10.1002/9780470747278.ch3>.
29. Cichocki, A.; Zdunek, R.; Amari, S. Csiszár’s Divergences for Nonnegative Matrix Factorization: Family of New Algorithms. In *Independent Component Analysis and Signal Separation*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 3889, pp. 32–39.
30. Gunasekar, S.; Woodworth, B.; Srebro, N. Mirrorless Mirror Descent: A Natural Derivation of Mirror Descent. In Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, CA, USA, 13–15 April 2021; Volume 130, pp. 2305–2313.
31. Nielsen, F. A Note on the Natural Gradient and Its Connections with the Riemannian Gradient, the Mirror Descent, and the Ordinary Gradient. Github Report. Available online: <https://franknielsen.github.io/blog/NaturalGradientConnections/NaturalGradientConnections.pdf> (accessed on 1 August 2020).
32. Hristopulos, D.T.; da Silva, S.L.E.; Scarfone, A.M. Twenty Years of Kaniadakis Entropy: Current Trends and Future Perspectives. *Entropy* **2025**, *27*, 247.
33. Naudts, J. Deformed exponentials and logarithms in generalized thermostatics. *Phys. A Stat. Mech. Its Appl.* **2002**, *316*, 323–334.
34. Tsallis, C. Entropy. *Encyclopedia* **2022**, *2*, 264–300.
35. Wada, T.; Scarfone, A.M. Finite difference and averaging operators in generalized entropies. *J. Phys. Conf. Ser.* **2010**, *201*, 012005.
36. Kaniadakis, G.; Scarfone, A.M. A new one-parameter deformation of the exponential function. *Phys. A Stat. Mech. Its Appl.* **2002**, *305*, 69–75.
37. Kaniadakis, G. Statistical mechanics in the context of special relativity. *Phys. Rev. E* **2002**, *66*, 056125.
38. Kaniadakis, G.; Lissia, M.; Scarfone, A.M. Deformed logarithms and entropies. *Phys. A Stat. Mech. Its Appl.* **2004**, *340*, 41–49.
39. Kaniadakis, G.; Lissia, M.; Scarfone, A.M. Two-parameter deformations of logarithm, exponential, and entropy: A consistent framework for generalized statistical mechanics. *Phys. Rev. E* **2005**, *71*, 046128.
40. Tempesta, P. A theorem on the existence of trace-form generalized entropies. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2015**, *471*, 20150165.
41. Wada, T.; Scarfone, A.M. On the Kaniadakis distributions applied in statistical physics and natural sciences. *Entropy* **2023**, *25*, 292.
42. Gomez, I.S.; Borges, E.P. Algebraic structures and position-dependent mass Schrödinger equation from group entropy theory. *Lett. Math. Phys.* **2021**, *111*, 43.
43. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.
44. Tsallis, C. What are the numbers that experiments provide. *Quim. Nova* **1994**, *17*, 468–471.
45. Ishige, K.; Salani, P.; Takatsu, A. Hierarchy of deformations in concavity. *Inf. Geom.* **2024**, *7* (Suppl. 1), 251–269.
46. Box, G.E.P.; Cox, D.R. An Analysis of Transformations. *J. R. Stat. Soc. Ser. B* **1964**, *26*, 211–252.
47. Borges, E.P. A possible deformed algebra and calculus inspired in nonextensive thermostatics. *Phys. A Stat. Mech. Its Appl.* **2004**, *340*, 95–101.
48. Yamano, T. Some properties of q -logarithm and q -exponential functions in Tsallis statistics. *Phys. A Stat. Mech. Its Appl.* **2002**, *305*, 486–496.
49. Nock, R.; Amid, E.; Warmuth, M.K. Boosting with Tempered Exponential Measures. *arXiv* **2023**, arXiv:2306.05487. <https://doi.org/10.48550/arXiv.2306.05487>.
50. Borges, E.P.; Roditi, I. A family of nonextensive entropies. *Phys. Lett. A* **1998**, *246*, 399–402.

51. Schwämmle, V.; Tsallis, C. Two-parameter generalization of the logarithm and exponential functions and Boltzmann-Gibbs-Shannon entropy. *J. Math. Phys.* **2007**, *48*, 113301.
52. Cardoso, P.G.; Borges, E.P.; Lobao, T.C.; Pinho, S.T. Nondistributive algebraic structures derived from nonextensive statistical mechanics. *J. Math. Phys.* **2008**, *49*, 093509.
53. Corcino, C.B.; Corcino, R.B. Three-Parameter Logarithm and Entropy. *J. Funct. Spaces* **2020**, *2020*, 9791789.
54. Kaniadakis, G. Maximum entropy principle and power-law tailed distributions. *Eur. Phys. J. B* **2009**, *70*, 3–13.
55. Mittal, D.P. On some functional equations concerning entropy, directed divergence and inaccuracy. *Metrika* **1975**, *22*, 35–45.
56. Sharma, B.D.; Taneja, I.J. Entropy of type (α, β) and other generalized measures in information theory. *Metrika* **1975**, *22*, 205–215.
57. Taneja, I.J. On generalized information measures and their applications. *Adv. Electron. Electron Phys.* **1989**, *76*, 327–413.
58. Scarfone, A.M.; Suyari, H.; Wada, T. Gauss law of error revisited in the framework of Sharma-Taneja-Mittal information measure. *Cent. Eur. J. Phys.* **2009**, *7*, 414–420.
59. Kaniadakis, G.; Scarfone, A.M.; Sparavigna, A.; Wada, T. Composition law of κ -entropy for statistically independent systems. *Phys. Rev. E* **2017**, *95*, 052112.
60. Furuichi, S. An axiomatic characterization of a two-parameter extended relative entropy. *J. Math. Phys.* **2010**, *51*, 123302.
61. Da Silva, G.B.; Ramos, R.V. The Lambert-Tsallis W_q function. *Phys. A Stat. Mech. Its Appl.* **2019**, *525*, 164–170.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.