
A PHYSICS-INSPIRED MOMENTUM-BASED GRADIENT METHOD

A PREPRINT

Jianing Zhang
Northeastern University, China

Rumei Liu
Liaoning University, China

November 21, 2025

ABSTRACT

In this work, a nonlinear momentum method is introduced to improve the convergence performance of momentum-based gradient optimization algorithms. The method is motivated by the dynamics of non-Newtonian mechanical systems, where conventional momentum schemes can be interpreted as a dynamical model with quadratic kinetic energy and linear damping. Based on this analogy, a generalized optimization dynamics is constructed by extending the kinetic energy formulation and incorporating a nonlinear damping term. An anharmonic kinetic energy function can be employed to represent the inertial effect of accumulated gradient information during the iterations, while the nonlinear damping mechanism enables a more flexible control of the momentum contribution along the convergence trajectory. Numerical experiments indicate that the method exhibits faster convergence and higher robustness compared to classical momentum algorithms. Moreover, its strong performance on nonconvex objectives makes it particularly suitable for inverse photonic design problems. The results suggest that dynamical systems from physics can provide a view towards the development of efficient optimization methods.

Keywords momentum method · optimization dynamics · non-Newtonian flow · nonlinear damping · inverse design

1 Introduction

Physics has long served as a fundamental source of inspiration for numerical optimization. Classical examples include Newton-type and quasi-Newton methods [1], originally derived from physical models of force balance and dynamical relaxation. Momentum-based optimization methods, such as Polyak’s Heavy Ball (HB) [2], can be interpreted as damped mechanical systems, while modern algorithms—like simulated annealing [3], Langevin dynamics [4], and approximate message passing [5]—are formulated from statistical physics principles. These physics-inspired algorithms have not only enriched optimization theory but also profoundly influenced computational physics, machine learning, and materials and chemical design.

Gradient descent, one of the simplest first-order methods, has drawn much attention with the rise of machine learning, especially deep learning [6]. Its low computational cost makes it suitable for large-scale and distributed optimization. Momentum-based methods improve gradient descent by introducing a linear damping term. Its accelerated variants, including Nesterov’s Accelerated Gradient (NAG) [7] and adaptive methods such as Adam, further extend the momentum idea to improve convergence. From a physics perspective, a natural question arises: can nonlinear damping mechanisms achieve comparable or superior acceleration? Recent studies have incorporated dry friction [8], mixed viscous-Coulomb damping [9], and Hessian-dependent mechanisms, demonstrating promising performance in nonsmooth optimization.

Inverse photonic design has emerged as a central challenge in computational physics and engineering [10]. Applications [10, 11]—from nanophotonics [12, 13] and near-field radiative heat transfer (NFRHT) to metamaterials—seek to automatically infer optimal design parameters achieving target performance. These problems are typically high-dimensional, strongly nonconvex, and computationally expensive, posing challenges for conventional optimization schemes. This motivates the exploration of physics-inspired nonlinear momentum dynamics as effective solvers for inverse design.

In this work, we develop a nonlinear momentum-based optimization algorithm and apply it to inverse design. The method accelerates convergence, enhances global search capability, and exhibits robustness under physics-constrained landscapes.

Contributions:

1. Algorithmic: We propose a gradient-based optimization method with a nonlinear momentum mechanism, inspired by non-Newtonian dynamics, which achieves faster convergence than classical momentum methods while maintaining stability.
2. Application: The algorithm is applied to inverse design in nanophotonics. Theoretical analysis and numerical experiments demonstrate its efficiency and robustness in high-dimensional, constrained problems, suggesting broader applicability to other fields such as machine learning.

The remainder of this paper is organized as follows. Section 2 introduces the dynamical modeling of momentum methods and presents the nonlinear damping construction and discretization. Section 3 provides a theoretical analysis of local convergence and computational aspects. Section 4 gives numerical experiments on representative nonconvex objectives and near field thermal radiation design to validate the proposed method. Finally, Section 5 concludes the paper and outlines future works.

2 Algorithm

2.1 Momentum Methods and Optimization Dynamics

Momentum methods play a central role in both optimization theory and practice. The HB method [14] treats the gradient descent update as the motion of a particle in a potential field, accelerating convergence by introducing a linear damping term. Within the framework of classical mechanics, momentum methods can be connected to the Euler–Lagrange equation :

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_j} \right) - \frac{\partial L}{\partial x_j} = D_j \quad (1)$$

here L is the Lagrangian, consisting of kinetic energy K and potential energy V :

$$L(x, \dot{x}, t) = K(\dot{x}) - V(x, t) \quad (2)$$

D_j denotes generalized forces not derived from a potential, such as friction, which commonly appears in damped mechanical systems. In the case of linear damping, the frictional force is proportional to the particle’s velocity, so the generalized force component in the x direction is given by:

$$D_j = -\frac{\partial \phi}{\partial \dot{x}_j} = -\gamma \dot{x}(t) \quad (3)$$

where γ is the damping coefficient, $\phi(\dot{x})$ is the dissipative function. Substituting (3) into (1) yields the Lagrangian with a dissipative term:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_j} \right) - \frac{\partial L}{\partial x_j} + \frac{\partial \phi}{\partial \dot{x}_j} = 0, \quad (4)$$

The classical Heavy Ball method corresponds to linear damping, represented by the second-order ODE [14]:

$$\ddot{x}(t) + \gamma \dot{x}(t) + \nabla V(x(t)) = 0. \quad (5)$$

2.2 Anharmonic Kinetic Energy and Nonlinear Damping

The aforementioned methods are fundamentally based on linear damping and Newtonian mechanics. However, in nonconvex optimization, traditional momentum methods often exhibit large oscillations and poor stability, especially near extrema, which can result in overshooting. In many complex physical systems—such as non-Newtonian fluids, viscoelastic materials, or oscillators with nonlinear damping—the inertial and damping behaviors are not well described by linear assumptions. Recently, several studies [8, 9] have modeled optimization dynamics as nonlinear dynamical systems to tackle more challenging objectives. We expect that by introducing more sophisticated inertial or damping structures, optimization methods can achieve improved performance, particularly for nonconvex problems.

From the perspective of non-Newtonian mechanics, the Heavy Ball method can naturally be extended to a nonlinear momentum-based gradient descent method, described by the following nonlinear dynamical system:

$$\ddot{x}(t) + \nabla \phi(\dot{x}(t)) + \nabla V(x(t)) = 0 \quad (6)$$

This equation generalizes classical momentum optimization by extending the damping term to a velocity-dependent nonlinear dissipation function $\sim \frac{2}{s} \|\dot{x}\|_s^s$. The parameter $s \geq 0$ controls the degree of nonlinearity: when $s = 0$, the system reduces to linear damping (i.e., classical momentum); when $s > 0$, damping increases with velocity, consistent with the behavior of non-Newtonian media.

The nonlinear momentum optimization method proposed in this work is inspired by non-Newtonian mechanics, incorporating nonlinear damping terms and non-quadratic kinetic energy terms into the dynamical equations, thereby forming a class of nonlinear dissipative optimization dynamical systems. In the following, we focus primarily on the system's physical dynamical behavior, energy dissipation structure, and algorithmic stability, providing a physical perspective for interpreting and qualitatively analyzing its optimization mechanisms. Rigorous and complete mathematical proofs of convergence are left to the field of optimization theory.

Consider a finite-dimensional mechanical system with generalized coordinate $x \in \mathbb{R}^n$ and velocity $v = \dot{x}$. Let $K(v)$ be the kinetic energy function. Assume $K(v) : \mathbb{R}^n \rightarrow \mathbb{R}$ is \mathcal{C}^2 and strictly convex in v . And $V(x)$ be a \mathcal{C}^1 potential energy function; the dissipation (generalized damping) is given by a force $D(v)$ such that $v \cdot D(v) \geq 0$ for all v .

The equation of motion with dissipation is

$$\frac{d}{dt}(\nabla_v K(v)) + \nabla_x V(x) + D(v) = 0, \quad (7)$$

We would like to show that the total energy decreases under these assumptions. Define the momentum:

$$p := \nabla_v K(v) \quad (8)$$

The Legendre transform of K is

$$K^*(p) := \sup_v (p \cdot v - K(v)). \quad (9)$$

By convex duality, the supremum is attained at the unique $v^* = v(p)$ satisfying $p = \nabla_v K(v)$, and we have the identity

$$K^*(p) + K(v) = p \cdot v, \text{ with } v = \nabla_p K^*(p) \quad (10)$$

Define the Hamiltonian

$$H(x, p) := K^*(p) + V(x) \quad (11)$$

which equals the usual energy $E = H$. The dynamics (1) can be written in Hamiltonian form with dissipation:

Thus the energy function we track is

$$E(t) = H(x(t), p(t)). \quad (12)$$

The corresponding continuous-time optimization dynamics are then described by Hamilton's equations:

$$\dot{x} = \frac{\partial H}{\partial p} = v(p) \quad (13)$$

$$\dot{p} = -\frac{\partial H}{\partial x} - D(v(p)) = -\nabla_x V(x) - D(v) \quad (14)$$

This is equivalent to (1) under the Legendre duality.

Differentiate $H(x(t), p(t))$ with respect to t :

$$\frac{d}{dt}H(x(t), p(t)) = \frac{\partial H}{\partial x} \cdot \dot{x} + \frac{\partial H}{\partial p} \cdot \dot{p} \quad (15)$$

Insert $\dot{x} = \partial_p H$ and $\dot{p} = -\partial_x H - D(v)$:

$$\frac{dH}{dt} = \partial_x H \cdot \partial_p H + \partial_p H \cdot (-\partial_x H - D(v)) = -v \cdot D(v). \quad (16)$$

But $\partial_p H = \dot{x} = v$, Hence

$$\frac{dE}{dt} = \frac{dH}{dt} = -v \cdot D(v). \quad (17)$$

By the assumed dissipation criterion $v \cdot D(v) > 0$, we obtain

$$\frac{d}{dt}E \leq 0. \quad (18)$$

If $v \cdot D(v) > 0$ for every nonzero v , then E is strictly decreasing along any trajectory with $v \neq 0$; if $v \cdot D(v) = 0$ only at $v = 0$, then E is nonincreasing and can decrease strictly until the system settles with $v = 0$. That completes the proof under the stated assumptions. Let the dissipation force be

$$D(v) = \mu |v|^{\alpha-1} \odot \text{sgn}(v) \quad (19)$$

Where $|v|^{\alpha-1} := (|v_1|^{\alpha-1}, |v_2|^{\alpha-1}, \dots, |v_n|^{\alpha-1})'$, and \odot denotes componentwise product. Then

$$v \cdot D(v) = \mu \sum_i |v_i|^\alpha = \mu \|v\|_\alpha^\alpha \geq 0. \quad (20)$$

so $dE/dt \leq 0$.

Hence, the kinetic energy becomes,

$$K(v) = \frac{1}{sm^{s-1}} \sum_{i=1}^n |v_i|^s = \frac{1}{sm^{s-1}} \|v\|_s^s, s > 1, m > 0 \quad (21)$$

Compute the generalized momentum

$$p_i = \frac{\partial K}{\partial v_i} = \frac{1}{sm^{s-1}} s |v_i|^{s-1} \text{sgn}(v_i) = m^{1-s} |v_i|^{s-1} \text{sgn}(v_i) \quad (22)$$

Hence

$$p_i = m^{1-s} |v_i|^{s-1} \implies |v_i|^{s-1} = m^{s-1} |p_i| \implies |v_i| = m |p_i|^{\frac{1}{s-1}}. \quad (23)$$

That is the mapping $v = \nabla_p K^*(p)$. Then,

$$p_i v_i = p_i (m |p_i|^{\frac{1}{s-1}}) = m |p_i|^{\frac{s}{s-1}} \quad (24)$$

Also

$$|v_i|^s = (m |p_i|^{\frac{1}{s-1}})^s = m^s |p_i|^{\frac{s}{s-1}} \quad (25)$$

So

$$K(v) = \frac{1}{sm^{s-1}} \sum_{i=1}^n |v_i|^s = \frac{1}{sm^{s-1}} \sum_i m^s |p_i|^{\frac{s}{s-1}} = \frac{m}{s} \sum_i |p_i|^{\frac{s}{s-1}} \quad (26)$$

Therefore,

$$K^*(p) = \sum_i p_i v_i - K(v) = m(1 - \frac{1}{s}) \sum_{i=1}^n |p_i|^{\frac{s}{s-1}} \quad (27)$$

Introduce the conjugate exponent $r = \frac{s}{s-1}$ (so $1/s + 1/r = 1$). In vector norm form:

$$K^*(p) = \frac{m(s-1)}{s} \|p\|_r^r. \quad (28)$$

Differentiate K^* with respect to p_i :

$$\frac{\partial K^*(p)}{\partial p_i} = \frac{m(s-1)}{s} r |p_i|^{r-1} \text{sgn}(p_i) = m |p_i|^{r-1} \text{sgn}(p_i). \quad (29)$$

But $r - 1 = \frac{s}{s-1} - 1 = \frac{1}{s-1}$, hence

$$\partial_{p_i} K^*(p) = m \text{sgn}(p_i) |p_i|^{\frac{1}{s-1}} = v_i, \quad (30)$$

which recovers the inversion $v = \nabla_p K^*(p)$ derived above. The Hamiltonian is

$$H(x, p) = K^*(p) + V(x) = \frac{m(s-1)}{s} \sum_i |p_i|^r + V(x). \quad (31)$$

And the monotone decay of the Hamiltonian follows since $v \cdot D(v) \geq 0$.

It should be noted that in conventional gradient based optimization dynamics, the kinetic energy term is quadratic, corresponding to linear velocity dependence in a Newtonian system. In this work, the restriction can be lifted by generalizing the kinetic energy to:

$$K(\dot{x}) = \frac{1}{sm^{s-1}} \|\dot{x}\|_s^s \quad (32)$$

where m is the particle mass. While alternative kinetic energy formulations are possible, we adopt the power-law form in this study.

2.3 Discretization and Update Scheme

To apply the continuous-time dynamics to iterative optimization, we use an explicit discretization scheme. Let $x_k \approx x(t_k)$ and $p_k \approx p(t_k)$, the Heavy Ball method naturally arises as:

$$p_k = -h\nabla V(x_k) + (1 - \gamma h)p_{k-1} \quad (33)$$

$$x_{k+1} = x_k + hp_k \quad (34)$$

Let $\alpha = h^2$, $\beta = 1 - \gamma h$, and $hp_k \rightarrow p_k$, we recover the standard Heavy Ball update. Replacing $\nabla V(x_k) \rightarrow \nabla V(x_k + \beta p_{k-1})$, we get Nesterov's accelerated gradient method.

$$x_{k+1} = x_k - \alpha \nabla V(x_k + \beta(x_k - x_{k-1})) + \beta(x_k - x_{k-1}) \quad (35)$$

For the proposed algorithm, the update rules are:

$$p_k = p_{k-1} - h\nabla V(x_k) - h\nabla\phi(p_{k-1}) \quad (36)$$

$$x_{k+1} = x_k + h\nabla K^*(p_k) \quad (37)$$

here $\phi(p_k)$ controls the damping, and $K^*(p_k)$ determines the anharmonic kinetic energy. In the update scheme above, $\gamma > 0$ is the nonlinear damping coefficient, s controls the nonlinearity of the damping, and $h > 0$ is the step size (time interval or learning rate).

The discrete updates emulate the dynamics of a velocity-dependent nonlinear dissipative system, with the following properties: when $s = 2$, the algorithm reduces to the classical Heavy Ball method; when $s = 1$, damping scales with the velocity magnitude, effectively suppressing high-speed oscillations; when $1 < s < 2$, momentum is retained in low-speed regions, helping the system escape saddle points or flat plateaus.

It should be noted that both the proposed nonlinear momentum method and mirror descent [15] reinterpret gradient-based optimization from a generalized dynamical or geometric perspective, they differ in their underlying principles and objectives. Mirror descent reformulates the optimization process in a non-Euclidean geometry defined by a mirror map or a Bregman divergence. It effectively performs gradient descent in the dual space, adapting the update direction and step size according to the local curvature or information geometry of the problem. The goal is to ensure better alignment between the optimization geometry and the structure of the objective function.

In contrast, the nonlinear momentum method extends the temporal dynamics rather than the spatial geometry of optimization. It originates from the analogy between gradient methods and physical systems governed by kinetic energy and damping. By introducing an anharmonic (non-quadratic) kinetic energy and nonlinear damping, the method generalizes classical momentum schemes to capture richer inertial behaviors and adaptive momentum attenuation. Thus, while mirror descent adapts the geometry of the descent trajectory, the nonlinear momentum method adapts its dynamical law.

Despite this distinction, the two frameworks are conceptually connected. Both can be viewed as geometric or dynamical regularizations of gradient descent: mirror descent modifies the metric tensor of the optimization manifold, whereas the nonlinear momentum method modifies the kinetic energy metric of the underlying dynamical system. In principle, a unified interpretation can be achieved by formulating the nonlinear momentum dynamics in a Riemannian space defined by a mirror map, suggesting a possible bridge between the two approaches.

Since mirror descent can naturally handle constraints and possesses better convergence and stability properties, we propose to combine the nonlinear momentum mechanism with the mirror descent framework. This yields a generalized nonlinear momentum algorithm, which integrates the advantages of adaptive momentum dynamics and geometry-aware updates.

$$p_k = p_{k-1} - h\nabla V(x_k) - h\nabla\phi(p_{k-1}) \quad (38)$$

$$z_{k+1} = \psi(z_k) + h\nabla K^*(p_k) \quad (39)$$

$$x_{k+1} = \psi^*(x_k) \quad (40)$$

where $\psi(\cdot)$ is the mirror potential function and $\psi^*(\cdot)$ its conjugate. For example, in the l_p geometry, $\psi(z) = \frac{1}{p} \sum |z_i|^p$ and $\psi^*(x) = \frac{1}{q} \sum |x_i|^q$ ($\frac{1}{p} + \frac{1}{q} = 1$).

3 Numerical experiments

To validate the effectiveness and stability of the proposed nonlinear momentum optimization method, this section presents several numerical experiments comparing the convergence performance of different momentum mechanisms on representative optimization tasks, with a particular focus on the adaptability of nonlinear damping and non-quadratic kinetic energy structures in complex optimization problems.

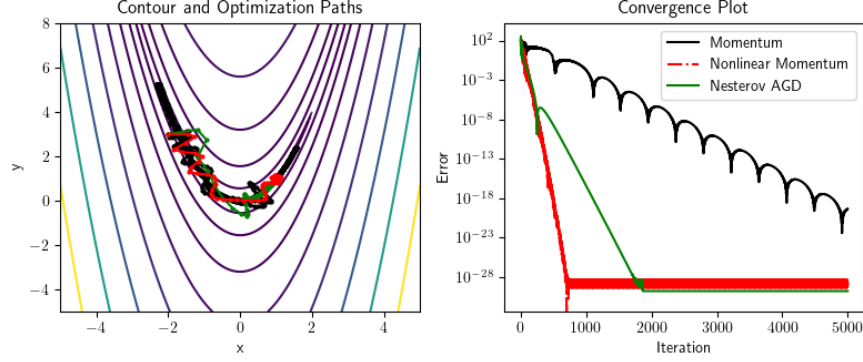


Figure 1: Convergence plot of Rosenbrock function.

3.1 Rosenbrock Function

As the first example, we consider the Rosenbrock function, a widely used benchmark nonconvex test function in optimization:

$$V(x) = \sum_{i=1}^{d-1} [100(x_{i+1} - x_i^2)^2 - (x_i - 1)^2] \quad (41)$$

The Rosenbrock function features a highly curved valley, which can be interpreted as a special potential energy landscape, making it suitable for evaluating algorithmic convergence under nonlinear coupling. In this task, we compare the following optimization algorithms: the classical Heavy Ball (HB) method, representing a momentum mechanism with linear damping and quadratic kinetic energy; the proposed nonlinear momentum method; and Nesterov's Accelerated Gradient Decent method. As shown in Figure 1, the proposed algorithm achieves faster convergence.

3.2 Inverse Design of Multilayer Nanospheres

In the 2nd example, we consider an inverse design problem for a multilayer nanosphere composed of alternating silver (Ag) and silica (SiO_2) shells. The objective is to optimize the layer thicknesses so that the resulting structure exhibits high average emissivity within the visible range (400-800 nm). The design variable is the set of shell thicknesses

$$x = (d_1, d_2, \dots, d_N)', \quad (42)$$

where N is the number of layers. Physical constraints include manufacturability (minimum layer thickness 5 nm) and an upper bound on the overall radius $R_{\max} = 300\text{ nm}$.

The spectral absorption efficiency $Q_{abs}(\lambda; x)$ of the multilayer sphere is computed using the modified recursive transfer matrix algorithm (mRTMA), a method previously developed by [16] for concentric layered spheres. The broadband objective is defined as the spectral average:

$$J(x) = \frac{1}{\lambda_2 - \lambda_1} \int_{\lambda_1}^{\lambda_2} \sigma_{abs}(\lambda; x) d\lambda, \quad (43)$$

where $\lambda_1 = 400\text{ nm}$, $\lambda_2 = 800\text{ nm}$. And we are concerned with the weighted average absorptivity, which is defined as the arithmetic mean. The design is subject to the following layer thickness bounds:

$$d_{\min} \leq d \leq d_{\max} \quad (44)$$

where the lower bound ensures the validity of the effective continuous dielectric description and the upper bound reflects fabrication limits. The overall particle radius must not exceed a prescribed maximum value. For silver layers, the minimum thickness must be sufficiently large to form a continuous film. For $N = 6$ layers, the optimized configuration yields an average absorptivity $J(x_*) \approx 0.72$, compared to ~ 0.45 for a homogeneous SiO_2 sphere of the same radius. The optimal design exhibits a balance between plasmonic resonance enhancement from Ag shells and dielectric confinement from SiO_2 layers. Figures 2 and 3 present the results. We can see that the proposed method converges much faster than the projected gradient descent, which becomes trapped in local minima.

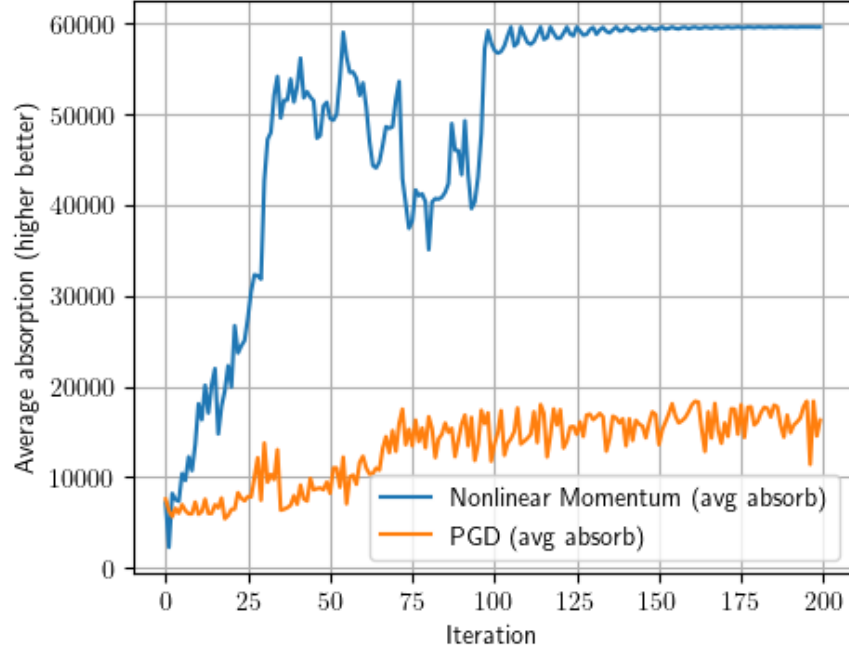


Figure 2: Performance profiles of nonlinear momentum and projected gradient decent on the numerical example 2.

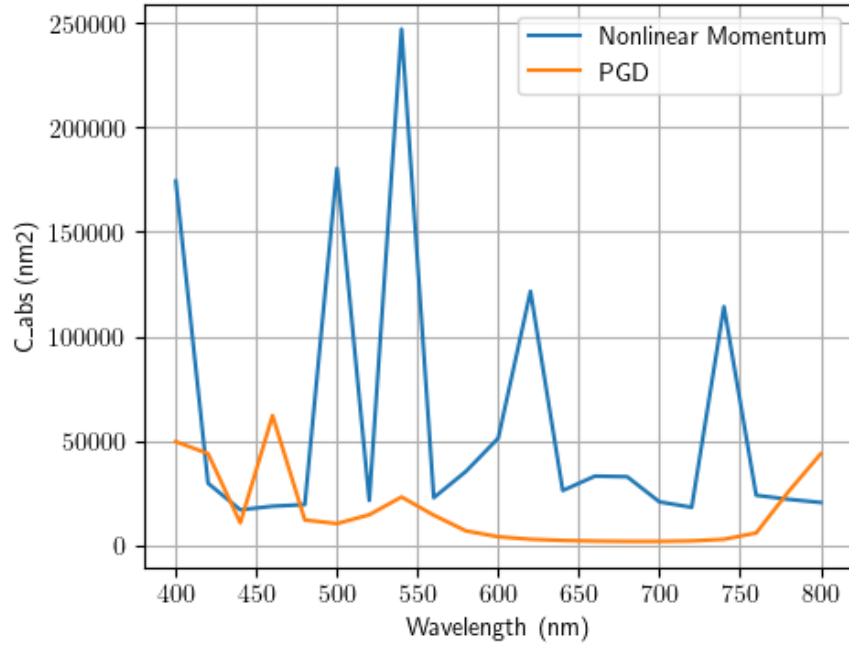


Figure 3: Optimization of average absorption cross-sections over visible wavelength range. The object under consideration is a silver/ SiO_2 multilayer sphere. The method for computing the absorption cross section is mRTMA [16].

It should be noted that the results obtained in this numerical example correspond to a local optimum only. In practical inverse photonic design problems, identifying the global optimum of design parameters typically requires the use of global optimization strategies, such as the Multi-Level Single-Linkage (MLSL) algorithm, which selects the best solution among multiple local optima.

4 Conclusion

Based on the dynamics of non-Newtonian mechanical systems, this work proposes a novel nonlinear momentum optimization method. The method enhances the convergence speed and robustness of momentum-based gradient optimization algorithms by generalizing the kinetic energy form and introducing a nonlinear damping mechanism. Specifically, an anharmonic kinetic energy function is employed to characterize the inertial effect of accumulated gradient information during iterations, while the nonlinear damping adaptively regulates the influence of the momentum term on the convergence trajectory. Numerical experiments illustrate that, compared with classical momentum-based methods, the proposed algorithm exhibits notable advantages in parameter flexibility, stability, and robustness against gradient noise. Moreover, the method establishes new a physics-based perspective complementary to traditional numerical optimization analysis.

This study not only shows superior performance over conventional gradient methods in inverse photonic design tasks but also represents a new attempt to develop physics-inspired optimization algorithms. Future research directions include employing higher-order numerical integration schemes, introducing more general forms of kinetic energy, and combining advanced sampling and optimization strategies within Hamiltonian Monte Carlo frameworks to further improve efficiency. Additionally, systematic theoretical analysis is needed to deepen the understanding of the convergence and stability properties of nonlinear momentum mechanisms.

5 Acknowledgements

References

- [1] Nocedal J. and Wright S. J. *Numerical Optimization*. Springer, New York, 2006.
- [2] B.T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.
- [3] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
- [4] Max Welling and Yee Whye Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML’11, page 681–688, Madison, WI, USA, 2011. Omnipress.
- [5] David L. Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.
- [6] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1139–1147, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.
- [7] Y. Nesterov. A method for solving the convex programming problem with convergence rate $o(1/k^2)$, 1983.
- [8] Samir Adly and Hedy Attouch. First-order inertial algorithms involving dry friction damping. *Mathematical Programming*, 193(1):405–445, 2022.
- [9] Samir Adly, Hedy Attouch, and Manh Hung Le. First order inertial optimization algorithms with threshold effects associated with dry friction. *Computational Optimization and Applications*, 86(3):801–843, 2023.
- [10] Sean Molesky, Zin Lin, Alexander Y. Piggott, Weiliang Jin, Jelena Vucković, and Alejandro W. Rodriguez. Inverse design in nanophotonics. *Nature Photonics*, 12(11):659–670, 2018.
- [11] S.-A. Biehs, R. Messina, P. S. Venkataram, A. W. Rodriguez, J. C. Cuevas, and P. Ben-Abdallah. Near-field radiative heat transfer in many-body systems. *Rev. Mod. Phys.*, 93:025009, Jun 2021.
- [12] Zhichao Ruan and Shanhui Fan. Design of subwavelength superscattering nanospheres. *Applied Physics Letters*, 98(4):043101, 01 2011.

- [13] Wenjun Qiu, Brendan G. DeLacy, Steven G. Johnson, John D. Joannopoulos, and Marin Soljačić. Optimization of broadband optical response of multilayer nanospheres. *Opt. Express*, 20(16):18494–18504, Jul 2012.
- [14] Nikola B. Kovachki and Andrew M. Stuart. Continuous time analysis of momentum methods. *Journal of Machine Learning Research*, 22(17):1–40, 2021.
- [15] Arkadiĭ Semenovich Nemirovskiĭ and David Berkovich Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Chichester, 1983.
- [16] Jianing Zhang. A modified recursive transfer matrix algorithm for radiation and scattering computation of a multilayered sphere. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 338:109401, 2025.