https://opennovelty.org/

Novelty Assessment Report

Paper: Nesterov Finds GRAAL: Optimal and Adaptive Gradient Method for Convex Optimization

PDF URL: https://openreview.net/pdf?id=vPSiCA3CkD

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-27

Abstract

In this paper, we focus on the problem of minimizing a continuously differentiable convex objective function, $\min_x f(x)$. Recently, Malitsky (2020); Alacaoglu et al. (2023) developed an adaptive first-order method, GRAAL. This algorithm computes stepsizes by estimating the local curvature of the objective function without any line search procedures or hyperparameter tuning, and attains the standard iteration complexity $\mathcal{O}(L\Vert x_0-x^ \Vert^2/\epsilon)$ of fixed-stepsize gradient descent for L-smooth functions. However, a natural question arises: is it possible to accelerate the convergence of GRAAL to match the optimal complexity $\mathcal{O}(\sqrt{L\Vert x_0-x^\Vert^2/\epsilon})$ of the accelerated gradient descent of Nesterov (1983)? Although some attempts have been made by Li and Lan (2025); Suh and Ma (2025), the ability of existing accelerated algorithms to adapt to the local curvature of the objective function is highly limited. We resolve this issue and develop GRAAL with Nesterov acceleration, which can adapt its stepsize to the local curvature at a geometric, or linear, rate just like non-accelerated GRAAL. We demonstrate the adaptive capabilities of our algorithm by proving that it achieves near-optimal iteration complexities for L-smooth functions, as well as under a more general (L_0,L_1)-smoothness assumption (Zhang et al., 2019).

Core Task Landscape

This paper addresses: **Accelerated Adaptive Gradient Methods for Convex Optimization**

A total of **37 papers** were analyzed and organized into a taxonomy with **18 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Adaptive Stepsize Selection Mechanisms**
- **Acceleration Frameworks and Momentum Techniques**
- **Specialized Problem Settings and Constraints**
- **Stochastic and Online Optimization**
- **Specialized Application Domains**

Complete Taxonomy Tree

- Accelerated Adaptive Gradient Methods for Convex Optimization Survey Taxonomy
- Adaptive Stepsize Selection Mechanisms
  - Local Curvature Estimation Without Line Search ★ (7 papers)
  - [0] Nesterov Finds GRAAL: Optimal and Adaptive Gradient Method for Convex Optimization (Anon et al., 2026) View paper
  - [2] Linesearch-free adaptive Bregman proximal gradient for convex minimization without relative smoothness (Ou Hongjia, 2025) View paper
  - [5] A simple uniformly optimal method without line search for convex optimization: T. li, g. lan (T Li, 2025) View paper
  - [13] Local Curvature Descent: Squeezing More Curvature out of Standard and Polyak Gradient Descent (RichtÃ¡rik, 2024) View paper
  - [24] Adaptive proximal algorithms for convex optimization under local Lipschitz continuity of the gradient (Puya Latafat, 2023) View paper
  - [28] Adaptive gradient descent without descent (Malitsky, 2019) View paper
  - [34] Adaptive Proximal Gradient Method for Convex Optimization (Yura Malitsky, 2023) View paper
  - Line-Search-Based Adaptive Methods (1 papers)
  - [21] Adaptive gradient descent for convex and non-convex stochastic optimization (Dvinskikh, 2019) View paper
  - Polyak and Barzilai-Borwein Stepsize Strategies (4 papers)
  - [1] Adaptive sgd with polyak stepsize and line-search: Robust convergence and variance reduction (Jiang Xiao-wen, 2023) View paper
  - [14] Adaptive stochastic gradient descent for large-scale learning problems (Zhuang Yang, 2022) View paper
  - [16] Adaptive Polyak Step-Size for Momentum Accelerated Stochastic Gradient Descent With General Convergence Guarantee (Jiawei Zhang, 2025) View paper
  - [18] AdaBB: Adaptive Barzilai-Borwein Method for Convex Optimization (Danqing Zhou, 2024) View paper
  - Universal and Parameter-Free Gradient Methods (2 papers)
  - [12] Universal Gradient Methods for Stochastic Convex Optimization (Rodomanov, 2024) View paper
  - [15] Stochastic Adaptive Gradient Descent Without Descent (Aujol, 2025) View paper
- Acceleration Frameworks and Momentum Techniques
  - Accelerated Gradient Methods with Adaptive Stepsizes (3 papers)
  - [9] Fast convex optimization via closed-loop time scaling of gradient dynamics (Radu Ioan BoÅ£, 2023) View paper
  - [20] Accelerated Convergence of Gradient Descent Using Adaptive Parameters (Mills, 2022) View paper

  - [27] A Unified Scheme to Accelerate Adaptive Cubic Regularization and Gradient Methods for Convex Optimization (Jiang Bo, 2022) View paper
  - Quasi-Newton and Second-Order Acceleration (4 papers)
  - [3] Accelerated quasi-newton proximal extragradient: Faster rate for smooth convex optimization (Jiang Rui-chen, 2023) View paper
  - [7] Simple Stepsize for Quasi-Newton Methods with Global Convergence Guarantees (Agafonov, 2025) View paper
  - [8] Gradient-Normalized Smoothness for Optimization with Approximate Hessians (Semenov, 2025) View paper
  - [17] Doubly Adaptive Scaled Algorithm for Machine Learning Using Second-Order Information (Jahani, 2021) View paper
  - Geometric and Continuous-Time Perspectives on Acceleration (1 papers)
  - [30] On the curved geometry of accelerated optimization (Defazio, 2019) View paper
- Specialized Problem Settings and Constraints
  - Proximal and Composite Optimization (2 papers)
  - [26] Projected Nesterovs proximal-gradient algorithm for sparse signal reconstruction with a convex constraint (Renliang, 2016) View paper
  - [31] Projected nesterov's proximal-gradient algorithm for sparse signal recovery (Renliang Gu, 2017) View paper
  - Frank-Wolfe and Projection-Free Methods (2 papers)
  - [6] Fast Frank--Wolfe Algorithms with Adaptive Bregman Step-Size for Weakly Convex Functions (Takahashi Shota, 2025) View paper
  - [29] Adaptive Variant of the Frank-Wolfe Algorithm for Convex Optimization Problems (G. V. Aivazian, 2023) View paper
  - Relative Smoothness and Bregman Divergence (1 papers)
  - [37] Gradient-Type Adaptive Methods for Relatively Lipschitz Convex Optimization Problems (Fedor Stonyakin, 2021) View paper
  - Weakly Smooth and Uniformly Convex Functions (1 papers)
  - [35] Fast Gradient Methods for Uniformly Convex and Weakly Smooth Problems (Park, 2022) View paper
  - High-Order Regularization and Gradient Norm Minimization (1 papers)
  - [4] High-order Accumulative Regularization for Gradient Minimization in Convex Programming (Ji Yao, 2025) View paper
  - Nonsmooth Convex Optimization and Hypodifferentials (1 papers)
  - [32] Hypodifferentials of nonsmooth convex functions and their applications to nonsmooth convex optimization (Dolgopolik, 2023) View paper
- Stochastic and Online Optimization
  - Stochastic Gradient Descent with Adaptive Stepsizes (1 papers)
  - [19] An Adaptive Stochastic Gradient Method with Non-negative Gauss-Newton Stepsizes (Orvieto, 2024) View paper
  - Online Convex Optimization and Regret Minimization (3 papers)
  - [10] Exploiting Curvature in Online Convex Optimization with Delayed Feedback (Qiu Hao, 2025) View paper
  - [11] Fast Rates in Stochastic Online Convex Optimization by Exploiting the Curvature of Feasible Sets (Shinji Ito, 2024) View paper
  - [36] Adaptive online gradient descent (Bartlett, 2007) View paper
  - Incremental and Finite-Sum Optimization (2 papers)
  - [22] Accelerating incremental gradient optimization with curvature information (Wai, 2020) View paper
  - [33] Curvature-aided incremental aggregated gradient method (Hoi-To Wai, 2017) View paper
- Specialized Application Domains
  - Distributed and Multi-Agent Optimization (1 papers)
  - [23] ZO-JADE: Zeroth-Order Curvature-Aware Distributed Multi-Agent Convex Optimization (Alessio Maritan, 2023) View paper
  - Differential Privacy and Curvature-Aware Methods (1 papers)
  - [25] Rapid DP Convex Optimization via Curvature-Aware (Second-Order) Algorithms. (Ali, 2025) View paper

Narrative

Core task: Accelerated adaptive gradient method for convex optimization with local curvature estimation. The field of accelerated adaptive gradient methods for convex optimization has evolved into a rich landscape organized around several complementary themes. Adaptive Stepsize Selection Mechanisms explore how to tune learning rates without exhaustive line search, often relying on local curvature estimates or gradient-based heuristics to balance computational cost and convergence speed. Acceleration Frameworks and Momentum Techniques build on Nesterov-style momentum and related schemes to achieve optimal rates, while Specialized Problem Settings and Constraints address structured domains such as proximal operators, feasible sets, or non-Euclidean geometries. Stochastic and Online Optimization extends these ideas to noisy or streaming data, and Specialized Application Domains tailor methods to machine learning, differential privacy, or large-scale scenarios. Representative works like Linesearch-free Bregman Proximal[2] and Uniformly Optimal Without Linesearch[5] illustrate the drive to eliminate costly subroutines, while Accelerated Quasi-Newton Extragradient[3] and High-order Accumulative Regularization[4] show how second-order information can be incorporated efficiently.

A particularly active line of research focuses on local curvature estimation without line search, where methods infer smoothness or Lipschitz constants on the fly to adapt stepsizes dynamically. This branch contrasts with classical backtracking approaches by avoiding repeated function evaluations, trading off some theoretical guarantees for practical efficiency. Nesterov GRAAL[0] sits squarely within this cluster, emphasizing acceleration combined with curvature-aware stepsize rules that do not require explicit line search. It shares conceptual ground with Local Curvature Descent[13], which also leverages gradient-based curvature proxies, and with Adaptive Proximal Local Lipschitz[24], which adapts to local geometry in proximal settings. Compared to Adaptive Without Descent[28], which relaxes monotone descent assumptions, Nesterov GRAAL[0] retains acceleration guarantees while still avoiding line search overhead. This positioning highlights an ongoing tension in the field: balancing the simplicity and speed of parameter-free methods against the convergence assurances of more conservative, line-search-based schemes.

Related Works in Same Category

The following **6 sibling papers** share the same taxonomy leaf node with the original paper:

1. Linesearch-free adaptive Bregman proximal gradient for convex minimization without relative smoothness

Authors: Ou Hongjia, Latafat, Puya, Themelis, Andreas | **Year/Venue**: 2025 | **URL**: View paper

Abstract

This paper introduces adaptive Bregman proximal gradient algorithms for solving convex composite minimization problems without relying on global relative smoothness or strong convexity assumptions. Building upon recent advances in adaptive stepsize selections, the proposed methods generate stepsizes based on local curvature estimates, entirely eliminating the need for backtracking linesearch. A key innovation is a Bregman generalization of Young's inequality, which allows controlling a critical ...

Relationship Analysis

Both papers belong to the same taxonomy category of local curvature estimation without line search, sharing the goal of computing adaptive stepsizes from local smoothness estimates. The candidate paper overlaps with the original by developing linesearch-free adaptive methods with local curvature estimation, but differs fundamentally in its approach: it uses Bregman proximal gradient methods with a distance-generating function under local relative smoothness, while the original paper develops an accelerated Nesterov-type method (GRAAL) for standard Euclidean settings under L-smoothness and (L0,L1)-smoothness assumptions.

---

2. A simple uniformly optimal method without line search for convex optimization: T. li, g. lan

Authors: T Li, G Lan | **Year/Venue**: 2025 | **URL**: View paper

Abstract

â¦ a convex optimization problem whose parameters are not given a priori. In particular, we present a novel accelerated gradient descent â¦ increased to reflect the local curvature. Therefore, â¦

Relationship Analysis

Both papers belong to the same taxonomy category of local curvature estimation without line search, developing adaptive stepsize methods that estimate local smoothness properties without backtracking procedures. The original paper (Accelerated GRAAL) focuses on achieving optimal O(√(L/ε)) complexity through Nesterov acceleration with geometric stepsize growth, adapting to local curvature at a linear rate and handling (L0,L1)-smoothness. The candidate paper (AC-FGM) achieves similar $O(1/k^2)$ optimal rates for smooth convex optimization but uses a different algorithmic framework with dual extrapolated sequences and a distinct stepsize policy that allows at most sublinear growth ($\eta_{k+1} \leq (k+1)/k \cdot \eta_k$), which limits its adaptation capabilities compared to GRAAL's geometric growth.

---

3. Local Curvature Descent: Squeezing More Curvature out of Standard and Polyak Gradient Descent

Authors: RichtÃ¡rik, Peter | **Year/Venue**: 2024 | **URL**: View paper

Abstract

We contribute to the growing body of knowledge on more powerful and adaptive stepsizes for convex optimization, empowered by local curvature information. We do not go the route of fully-fledged second-order methods which require the expensive computation of the Hessian. Instead, our key observation is that, for some problems (e.g., when minimizing the sum of squares of absolutely convex functions), certain local curvature information is readily available, and can be used to obtain surprisingly p...

Relationship Analysis

Both papers belong to the same taxonomy category of local curvature estimation without line search, developing adaptive stepsize methods that estimate local smoothness properties directly from gradient information. They overlap in their core approach of computing stepsizes from local curvature estimates to achieve adaptive convergence for convex optimization, both avoiding backtracking line search procedures. The key difference is that the original paper (Accelerated GRAAL) incorporates Nesterov acceleration to achieve near-optimal O(√(L/ε)) complexity with geometric stepsize growth, while the candidate paper (Local Curvature Descent) focuses on non-accelerated methods with matrix-valued stepsizes derived from a novel curvature mapping assumption, achieving O(L/ε) complexity for a specialized class of functions including sums of squares of absolutely convex functions.

---

4. Adaptive proximal algorithms for convex optimization under local Lipschitz continuity of the gradient

Authors: Puya Latafat, Andreas Themelis, Lorenzo Stella, Panagiotis Patrinos | **Year/Venue**: 2023 • arXiv (Cornell University) | **URL**: View paper

Abstract

Backtracking linesearch is the de facto approach for minimizing continuously differentiable functions with locally Lipschitz gradient. In recent years, it has been shown that in the convex setting it is possible to avoid linesearch altogether, and to allow the stepsize to adapt based on a local smoothness estimate without any backtracks or evaluations of the function value. In this work we propose an adaptive proximal gradient method, adaPG, that uses novel estimates of the local smoothness modu...

Relationship Analysis

Both papers belong to the same taxonomy category of local curvature estimation without line search, focusing on adaptive stepsize selection for convex optimization. They overlap in their core approach of estimating local smoothness/curvature to compute stepsizes without backtracking line search procedures. The key difference is that the original paper (Accelerated GRAAL) focuses on incorporating Nesterov acceleration with geometric stepsize growth to achieve optimal O(√(L/ε)) complexity, while the candidate paper (adaPGM/adaPDM) proposes adaptive proximal gradient and primal-dual methods with novel local smoothness estimates but does not emphasize acceleration or achieving the accelerated complexity rate.

---

5. Adaptive gradient descent without descent

Authors: Malitsky, Yura, Mishchenko, Konstantin | **Year/Venue**: 2019 | **URL**: View paper

Abstract

We present a strikingly simple proof that two rules are sufficient to automate gradient descent: 1) don't increase the stepsize too fast and 2) don't overstep the local curvature. No need for functional values, no line search, no information about the function except for the gradients. By following these rules, you get a method adaptive to the local geometry, with convergence guarantees depending only on the smoothness in a neighborhood of a solution. Given that the problem is convex, our method...

Relationship Analysis

Both papers belong to the same taxonomy category of local curvature estimation without line search, developing adaptive stepsize methods that estimate local smoothness properties. The candidate paper (AdGD/GRAAL) focuses on non-accelerated gradient descent with geometric stepsize growth and establishes O(1/k) convergence for convex functions, while the original paper extends this framework by incorporating Nesterov acceleration to achieve near-optimal O(1/√k) convergence rates. The key distinction is that the original paper addresses the acceleration challenge that the candidate paper leaves open, introducing additional coupling steps and modified stepsize rules to enable geometric adaptation while maintaining accelerated convergence guarantees.

---

6. Adaptive Proximal Gradient Method for Convex Optimization

Authors: Yura Malitsky, Konstantin Mishchenko | **Year/Venue**: 2023 • arXiv (Cornell University) | **URL**: View paper

Abstract

In this paper, we explore two fundamental first-order algorithms in convex optimization, namely, gradient descent (GD) and proximal gradient method (ProxGD). Our focus is on making these algorithms entirely adaptive by leveraging local curvature information of smooth functions. We propose adaptive versions of GD and ProxGD that are based on observed gradient differences and, thus, have no added computational costs. Moreover, we prove convergence of our methods assuming only local Lipschitzness o...

Relationship Analysis

Both papers belong to the same taxonomy category of local curvature estimation without line search, developing adaptive gradient methods that compute stepsizes from local smoothness estimates. The candidate paper (Adaptive Proximal Gradient Method) focuses on improving the AdGD algorithm with larger allowable stepsizes and extends it to the proximal case for composite objectives, while the original paper (Accelerated GRAAL) incorporates Nesterov acceleration to achieve near-optimal $O(\sqrt{(L/\varepsilon)})$ complexity. The key distinction is that the original paper achieves acceleration with geometric stepsize growth, whereas the candidate paper improves non-accelerated gradient descent with enhanced stepsize bounds but maintains $O(L/\varepsilon)$ complexity.

Contributions Analysis

Overall novelty summary. The paper develops an accelerated variant of the GRAAL algorithm that combines Nesterov momentum with adaptive stepsize selection based on local curvature estimation. It resides in the 'Local Curvature Estimation Without Line Search' leaf, which contains seven papers total, indicating a moderately populated research direction. This leaf focuses specifically on methods that infer smoothness or Lipschitz constants dynamically without backtracking procedures, distinguishing it from line-search-based approaches in a sibling leaf. The paper's core contribution—achieving accelerated convergence while maintaining GRAAL's adaptive stepsize capabilities—addresses a natural extension question in this subfield.

The taxonomy reveals that this work sits at the intersection of two major branches: 'Adaptive Stepsize Selection Mechanisms' and 'Acceleration Frameworks and Momentum Techniques'. The sibling leaf 'Accelerated Gradient Methods with Adaptive Stepsizes' contains three papers exploring similar acceleration themes but through different mechanisms. Neighboring leaves address related but distinct approaches: 'Polyak and Barzilai-Borwein Stepsize Strategies' uses gradient-difference-based rules rather than curvature estimates, while 'Universal and Parameter-Free Gradient Methods' emphasizes broader adaptivity to noise and smoothness. The paper's positioning suggests it bridges local curvature adaptation with optimal-rate acceleration, a combination less explored in adjacent branches.

Among thirty candidates examined, the contribution on near-optimal complexity for L-smooth functions shows three refutable candidates, indicating substantial prior work in this area. The core algorithmic contribution (accelerated GRAAL with adaptive stepsize) examined ten candidates with zero refutations, suggesting greater novelty in the specific combination of techniques. The contribution addressing (L0,L1)-smoothness also examined ten candidates without refutation. The limited search scope means these statistics reflect top-semantic-match overlap rather than exhaustive field coverage. The algorithmic novelty appears stronger than the complexity-bound claims, where existing accelerated adaptive methods provide closer precedents.

Based on the thirty-candidate search, the work appears to occupy a meaningful but not entirely unexplored niche. The taxonomy structure shows this is an active area with multiple related approaches, and the refutation statistics suggest the acceleration-with-adaptation combination is less saturated than the complexity guarantees themselves. The analysis does not cover the full breadth of optimization literature, so additional related work may exist beyond the top-semantic matches examined. The paper's positioning within a seven-paper leaf suggests moderate but not extreme crowding in this specific methodological direction.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Accelerated GRAAL algorithm with adaptive stepsize and Nesterov acceleration

Description: The authors propose Accelerated GRAAL (Algorithm 1), a first-order optimization method that incorporates Nesterov acceleration while maintaining the ability to adapt stepsizes to local curvature geometrically. This resolves limitations of prior accelerated adaptive methods like AC-FGM and AdaNAG, which only allow sublinear stepsize growth.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. A Totally Asynchronous Nesterovâs Accelerated Gradient Method for Convex Optimization

URL: View paper

Brief Assessment

Asynchronous Nesterov Accelerated[51] focuses on distributed/asynchronous optimization with block-based updates under communication delays, not on adaptive stepsize methods that estimate local curvature geometrically for centralized convex optimization.

---

2. Convergence rates of a momentum algorithm with bounded adaptive step size for nonconvex optimization

URL: View paper

Brief Assessment

Momentum Bounded Adaptive Stepsize[50] focuses on Adam-like algorithms with momentum for nonconvex optimization, not on accelerated gradient methods with Nesterov acceleration for convex optimization. The candidate analyzes convergence rates under bounded adaptive step sizes for a different algorithmic framework than GRAAL.

---

3. Accelerated gradient descent escapes saddle points faster than gradient descent

URL: View paper

Brief Assessment

Accelerated Escapes Saddles[52] focuses on escaping saddle points in nonconvex optimization using Nesterov's accelerated gradient descent, not on adaptive stepsize methods for convex optimization with geometric growth rates as in the original paper.

---

4. Primal-dual accelerated gradient descent with line search for convex and nonconvex optimization problems

URL: View paper

Brief Assessment

Primal-dual Accelerated Linesearch[56] uses exact line search for stepsize selection, whereas the original paper's Accelerated GRAAL adapts stepsizes through local curvature estimation without line search. These represent fundamentally different approaches to stepsize adaptation.

---

5. Stochastic gradient accelerated by negative momentum for training deep neural networks

URL: View paper

Brief Assessment

Negative Momentum Acceleration[57] focuses on negative momentum for deep neural network training, which is a distinct acceleration technique from Nesterov momentum. The candidate does not address adaptive stepsize methods for convex optimization or GRAAL-type algorithms.

---

6. Accelerated policy gradient: On the nesterov momentum for reinforcement learning

URL: View paper

Brief Assessment

Accelerated Policy Gradient[54] focuses on reinforcement learning with Nesterov momentum for policy optimization in MDPs, not on convex optimization with adaptive stepsize methods like GRAAL. The technical domains and problem settings are fundamentally different.

7. On the convergence of Nesterov's accelerated gradient method in stochastic settings

URL: View paper

Brief Assessment

Nesterov Stochastic Settings[55] focuses on Nesterov's accelerated gradient method in stochastic approximation and finite-sum settings with bounded variance assumptions, not on adaptive stepsize methods that estimate local curvature geometrically like GRAAL.

8. Tradeoffs between convergence rate and noise amplification for momentum-based accelerated optimization algorithms

URL: View paper

Brief Assessment

Momentum Noise Amplification Tradeoffs[48] focuses on analyzing noise amplification properties of momentum-based methods in optimization, not on developing adaptive stepsize algorithms with Nesterov acceleration for convex optimization.

9. Accelerated Distributed Projected Gradient Descent for Convex Optimization with Clique-wise Coupled Constraints

URL: View paper

Brief Assessment

Accelerated Distributed Clique-wise[53] addresses distributed convex optimization with clique-wise coupled constraints using Nesterov acceleration, not adaptive stepsize methods for unconstrained smooth optimization. The candidate focuses on constraint-coupled multi-agent systems with fixed stepsizes, while the original develops adaptive stepsize rules that adjust geometrically to local curvature.

10. An Adaptive and Parameter-Free Nesterov's Accelerated Gradient Method for Convex Optimization

URL: View paper

Brief Assessment

Parameter-Free Nesterov Accelerated[49] (AdaNAG) focuses on a different algorithmic approach with Lyapunov analysis and does not demonstrate that the specific GRAAL-based acceleration framework with geometric stepsize growth was previously known.

Contribution 2: Near-optimal iteration complexity for L-smooth convex functions without hyperparameter tuning

Description: The authors prove that Algorithm 1 achieves the optimal iteration complexity O(sqrt(L)||x0 - x*||^2 / epsilon) for L-smooth functions up to additive logarithmic factors, without requiring hyperparameter tuning or line search procedures, as stated in Corollary 2.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. The first optimal acceleration of high-order methods in smooth convex optimization

URL: View paper

Brief Assessment

First Optimal High-order[59] focuses on high-order methods (p≥2) for smooth convex optimization, not first-order gradient methods. The candidate addresses optimal acceleration of tensor methods with p-th order oracles, while the original paper develops an adaptive first-order gradient method (GRAAL with Nesterov acceleration).

2. Methods for Convex -Smooth Optimization: Clipping, Acceleration, and Adaptivity

URL: View paper

Brief Assessment

Convex Smooth Clipping Acceleration[63] focuses on (L0,L1)-smooth functions rather than standard L-smooth functions. While both papers address smooth convex optimization, the candidate's technical framework and smoothness assumptions differ fundamentally from the original paper's L-smoothness setting.

3. Achieving Linear Convergence with Parameter-Free Algorithms in Decentralized Optimization

URL: View paper

Brief Assessment

Linear Convergence Parameter-Free Decentralized[62] focuses on decentralized optimization over networks without centralized servers, while the original paper addresses centralized convex optimization with Nesterov acceleration. The candidate's linear convergence applies to strongly convex functions in a distributed setting, not the L-smooth convex case studied in the original work.

4. A simple uniformly optimal method without line search for convex optimization: T. li, g. lan

URL: View paper

Prior Art Analysis

Uniformly Optimal Without Linesearch[5] demonstrates that their AC-FGM algorithm achieves the optimal O(1/k^2) convergence rate for smooth convex optimization without requiring hyperparameter tuning or line search procedures. This directly refutes the novelty claim of the original paper, as the candidate paper was published earlier (accepted June 2025 in Mathematical Programming) and presents a method that achieves the same optimal complexity O(√(L)||x0-x*||^2/ε) without hyperparameter tuning. The candidate paper explicitly states their method is 'fully problem parameter free, line search free, and optimal for smooth convex problems' and provides theoretical guarantees matching the complexity claimed in the original paper.

Evidence

Evidence 1 - **Rationale**: Both papers claim to achieve optimal convergence rates for smooth convex optimization without hyperparameter tuning or line search. The candidate paper's AC-FGM achieves O(1/k^2) rate, which corresponds to the optimal complexity the original paper claims. - **Original**: we show that algorithm 1 achieves the optimal iteration complexity in eq. (4) for l-smooth functions up to additive logarithmic factors, without the requirement of hyperparameter tuning or any additional line search procedures - **Candidate**: we introduce a novel first-order algorithm, called auto-conditioned fast gradient method (ac-fgm) , which can achieve the optimal o(1/k 2) convergence rate for smooth convex optimization without knowing any problem parameters or resorting to any line search procedures

5. Towards simple and provable parameter-free adaptive gradient methods

URL: View paper

Brief Assessment

Simple Parameter-Free Adaptive[60] focuses on parameter-free adaptive gradient methods (AdaGrad++ and Adam++) with convergence rates for convex optimization, not on accelerated gradient methods with Nesterov acceleration for smooth convex optimization as in the original paper.

---

6. Optimal and parameter-free gradient minimization methods for convex and nonconvex optimization

URL: View paper

Prior Art Analysis

Optimal Parameter-Free Gradient[61] demonstrates that similar near-optimal iteration complexity results for L-smooth convex functions without hyperparameter tuning were achieved prior to the original paper. The candidate paper presents an accumulative regularization method that achieves $O(\sqrt{(L||x_0 - x^*||^2/\varepsilon)})$ gradient evaluations without requiring hyperparameter tuning or line search procedures, matching the complexity claimed in the original paper. Both papers achieve the same optimal complexity bound up to logarithmic factors without requiring hyperparameter tuning.

Evidence

Evidence 1 - **Rationale**: This demonstrates that the candidate paper's parameter-free algorithm achieves optimal complexity without hyperparameter tuning, directly addressing the same problem as the original paper's contribution. - **Original**: in section 3, we show that algorithm 1 achieves the optimal iteration complexity in eq. (4) for l-smooth functions up to additive logarithmic factors, without the requirement of hyperparameter tuning or any additional line search procedures. - **Candidate**: theorem 4.1 for any input argument m0,σ1 >0, the ar method outputs an ˆxwith ‖∇f(ˆx)‖ ≤ 5σ1 dist(x0,x∗) within 4 + 3 √ max{m0/2,2l}/σ1 + 16√2ca √ l/σ1 gradient evaluations of ∇f. specially, if m0 ≤ 4l, the total number of gradient evaluations is bounded by 4 + c1 √ l/σ1, where c1 := √ 2(3 + 16 √2ca)...

Evidence 2 - **Rationale**: This shows the candidate paper achieved the same near-optimal complexity $O(\sqrt{(Ld/\varepsilon)})$ for L-smooth functions, which is equivalent to the original paper's claimed complexity when considering the relationship between d and $||x_0-x||$. - Original: we resolve this issue and develop graal with nesterov acceleration, which can adapt its stepsize to the local curvature at a geometric, or linear, rate just like non-accelerated graal. we demonstrate the adaptive capabilities of our algorithm by proving that it achieves near-optimal iteration comple... - Candidate*: by theorem 4.1, if we have information on the upper bound of t he distance to any optimal solution, i.e., d≥ dist(x0,x∗), then by setting σ1 = ε/(5d) the output ˆxsatisfies ‖∇f(ˆx)‖ ≤ ε. the number of gradient evaluations for computing such an approx imate solution is given by o(1) √ ld/ε.

---

7. A Parameter-Free Conditional Gradient Method for Composite Minimization under HÃ¶lder Condition

URL: View paper

Brief Assessment

Parameter-Free Conditional Gradient Holder[65] addresses composite optimization with Hölder continuous gradients using conditional gradient methods, not accelerated gradient descent. The candidate focuses on weakly smooth functions and bounded/uniformly convex domains, which is a different problem setting than the original paper's focus on general convex optimization with L-smooth functions using Nesterov acceleration.

---

8. Convex-concave programming: An effective alternative for optimizing shallow neural networks

URL: View paper

Brief Assessment

Convex-concave Shallow Networks[58] focuses on shallow neural network optimization using convex-concave programming, not on general convex optimization with adaptive gradient methods. The limited context provided does not demonstrate prior work on adaptive stepsize methods achieving near-optimal complexity without hyperparameter tuning for general L-smooth convex functions.

---

9. How free is parameter-free stochastic optimization?

URL: View paper

Brief Assessment

How Free Parameter-Free[66] focuses on stochastic optimization with noisy gradients and requires knowledge of parameter ranges, while the original paper addresses deterministic gradient methods with adaptive stepsize rules that estimate local curvature without any hyperparameter tuning or line search.

---

10. DoWG Unleashed: An Efficient Universal Parameter-Free Gradient Descent Method

URL: View paper

Prior Art Analysis

DoWG Unleashed[64] demonstrates that prior work achieved similar near-optimal iteration complexity for L-smooth convex functions without hyperparameter tuning. The candidate paper proves that their algorithm (DoWG/Algorithm 1) achieves the convergence rate $O(ld^2/t \cdot \log^+(d/r\epsilon))$ for L-smooth functions, which matches the optimal rate up to logarithmic factors without requiring hyperparameter tuning. This directly challenges the novelty claim that the original paper was first to achieve near-optimal complexity without hyperparameter tuning, as DoWG Unleashed[64] explicitly states their method requires no parameter tuning and achieves the same theoretical guarantees.

Evidence

Evidence 1 - **Rationale**: This pair shows that DoWG Unleashed[64] achieved the same near-optimal complexity $O(ld^2/t)$ up to logarithmic factors for L-smooth functions, demonstrating prior work with similar theoretical guarantees. - **Original**: we show that algorithm 1 achieves the optimal iteration complexity in eq. (4) for l-smooth functions up to additive logarithmic factors, without the requirement of hyperparameter tuning or any additional line search procedures. - **Candidate**: theorem 4. (dowg, smooth f). suppose that the function f is l-smooth, convex, and has a minimizer x∗ ∈ x. suppose that the domain x is a closed convex set of diameter d >0. let rϵ < d. then the output of algorithm 1 satisfies for some t ∈ {0, 1, . . . , t1} f(¯xt) - f∗ = o ld2 t log+ d rϵ

Evidence 2 - **Rationale**: This pair demonstrates that DoWG Unleashed[64] requires minimal initialization (choosing rϵ ≤ d) and no hyperparameter tuning, directly supporting the claim that prior work achieved similar results without hyperparameter tuning. - **Original**: in section 3, we show that algorithm 1 achieves the optimal iteration complexity in eq. (4) for l-smooth functions up to additive logarithmic factors, without the requirement of hyperparameter tuning or any additional line search procedures. - **Candidate**: the only initialization dowg requires is choosing rϵ ≤ d. this can be done by choosing rϵ ≤ ‖x - y‖ for any two x ̸= y ∈ x.

Evidence 3 - **Rationale**: This shows that DoWG Unleashed[64] explicitly claims to match optimal rates up to logarithmic factors, which is the same claim made in the original paper about achieving near-optimal complexity. - **Original**: we demonstrate the adaptive capabilities of our algorithm by proving that it achieves near-optimal iteration complexities for l-smooth functions - **Candidate**: dowg matches the optimal o dg√ t rate of tuned gd and tuned ngd up to an extra logarithmic factor. we note that the recently proposed algorithms dog (ivgi et al., 2023) and d-adaptation (defazio and mishchenko, 2023) achieve a similar rate in this setting.

Contribution 3: Near-optimal iteration complexity under (L0, L1)-smoothness assumption

Description: The authors demonstrate that Algorithm 1 achieves iteration complexity matching the optimal rate for (L0, L1)-smooth functions up to additive constant factors that do not depend on precision epsilon. This is the first adaptive algorithm to achieve such results under this more general smoothness condition.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. A Comprehensive Framework for Analyzing the Convergence of Adam: Bridging the Gap with SGD

URL: View paper

Brief Assessment

Comprehensive Adam Framework[45] analyzes Adam's convergence under L-smoothness and ABC inequality assumptions for stochastic optimization, while the original paper focuses on deterministic convex optimization under (L0, L1)-smoothness. These are fundamentally different problem settings and smoothness assumptions.

2. Adaptive Gradient Normalization and Independent Sampling for (Stochastic) Generalized-Smooth Optimization

URL: View paper

Brief Assessment

Adaptive Gradient Normalization[43] focuses on stochastic generalized-smooth optimization with independent sampling and gradient clipping, not on accelerated adaptive methods for deterministic (L0, L1)-smooth optimization. The candidate addresses different algorithmic challenges (stochastic noise handling) than the original paper's focus on adaptive stepsize selection with Nesterov acceleration.

3. Provable Complexity Improvement of AdaGrad over SGD: Upper and Lower Bounds in Stochastic Non-Convex Optimization

URL: View paper

Brief Assessment

AdaGrad Complexity Improvement[42] focuses on non-convex optimization with refined smoothness assumptions and L1-norm stationarity measures for adaptive methods, while the original paper addresses convex optimization under (L0, L1)-smoothness with L2-norm convergence guarantees. These are fundamentally different problem settings and objectives.

4. Adaptive Generalized Conditional Gradient Method for Multiobjective Optimization

URL: View paper

Brief Assessment

Adaptive Generalized Conditional Gradient[46] focuses on multiobjective optimization using Frank-Wolfe methods with normalized descent directions, not on achieving optimal iteration complexity under (L0, L1)-smoothness for single-objective convex optimization.

5. Gradient-Variation Online Adaptivity for Accelerated Optimization with H\" older Smoothness

URL: View paper

Brief Assessment

Gradient-Variation Online Adaptivity[39] focuses on Hölder smoothness in online learning and offline optimization, not on (L0, L1)-smoothness. The candidate paper studies (Lν, ν)-Hölder smooth functions where smoothness is defined as $||\nabla\ell(x)-\nabla\ell(y)||_2 \leq L\nu||x-y||$^$\nu_2$, which is a different generalization of smoothness than the (L0, L1)-smoothness condition studied in the original paper.

6. On convergence of adam for stochastic optimization under relaxed assumptions

URL: View paper

Brief Assessment

Adam Relaxed Assumptions[40] focuses on stochastic optimization with Adam under generalized smoothness conditions, while the original paper develops an accelerated deterministic gradient method (GRAAL with Nesterov acceleration) achieving near-optimal complexity for (L0, L1)-smooth functions. These are fundamentally different algorithmic approaches (stochastic adaptive vs. deterministic accelerated) addressing different problem settings.

7. Provable adaptivity of adam under non-uniform smoothness

URL: View paper

Brief Assessment

Adam Non-uniform Smoothness[47] focuses on Adam optimizer with diminishing learning rate under (L0, L1)-smoothness, while the original paper develops an accelerated gradient method (GRAAL with Nesterov acceleration) that adaptively adjusts stepsizes. These are fundamentally different algorithmic approaches to the same smoothness condition.

8. Decentralized Relaxed Smooth Optimization with Gradient Descent Methods

URL: View paper

Brief Assessment

Decentralized Relaxed Smooth[44] focuses on decentralized gradient descent methods for distributed optimization, while the original paper addresses centralized adaptive gradient methods with Nesterov acceleration. These are fundamentally different problem settings.

9. Convergence Guarantees for RMSProp and Adam in Generalized-smooth Non-convex Optimization with Affine Noise Variance

URL: View paper

Brief Assessment

The candidate paper (RMSProp Adam Generalized-smooth[41]) focuses on RMSProp and Adam algorithms under coordinate-wise generalized smoothness with affine noise variance in non-convex settings, achieving O(ε^-4) complexity. The original paper addresses convex optimization under (L0, L1)-smoothness with accelerated GRAAL achieving near-optimal $O(\sqrt{(L0d^2/\varepsilon)})$ complexity. These are fundamentally different problem settings (non-convex vs. convex) and different algorithmic approaches (adaptive gradient methods vs. accelerated gradient with local curvature estimation).

10. Convergence analysis of adaptive gradient methods under refined smoothness and noise assumptions

URL: View paper

Brief Assessment

Adaptive Refined Smoothness[38] focuses on convergence analysis of adaptive gradient methods (AdaGrad) in stochastic non-convex optimization with coordinate-wise smoothness assumptions, not on accelerated methods for convex optimization under (L0, L1)-smoothness as in the original paper.

Appendix: Text Similarity Detection

Textual similarity detection checked 35 papers and found 3 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Adaptive gradient descent without descent

Detected in: Core Task (sibling)

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Nesterov Finds GRAAL: Optimal and Adaptive Gradient Method for Convex Optimization View paper
- [1] Adaptive sgd with polyak stepsize and line-search: Robust convergence and variance reduction View paper
- [2] Linesearch-free adaptive Bregman proximal gradient for convex minimization without relative smoothness View paper
- [3] Accelerated quasi-newton proximal extragradient: Faster rate for smooth convex optimization View paper
- [4] High-order Accumulative Regularization for Gradient Minimization in Convex Programming View paper
- [5] A simple uniformly optimal method without line search for convex optimization: T. li, g. lan View paper
- [6] Fast Frank--Wolfe Algorithms with Adaptive Bregman Step-Size for Weakly Convex Functions View paper
- [7] Simple Stepsize for Quasi-Newton Methods with Global Convergence Guarantees View paper
- [8] Gradient-Normalized Smoothness for Optimization with Approximate Hessians View paper
- [9] Fast convex optimization via closed-loop time scaling of gradient dynamics View paper
- [10] Exploiting Curvature in Online Convex Optimization with Delayed Feedback View paper
- [11] Fast Rates in Stochastic Online Convex Optimization by Exploiting the Curvature of Feasible Sets View paper
- [12] Universal Gradient Methods for Stochastic Convex Optimization View paper
- [13] Local Curvature Descent: Squeezing More Curvature out of Standard and Polyak Gradient Descent View paper
- [14] Adaptive stochastic gradient descent for large-scale learning problems View paper
- [15] Stochastic Adaptive Gradient Descent Without Descent View paper
- [16] Adaptive Polyak Step-Size for Momentum Accelerated Stochastic Gradient Descent With General Convergence Guarantee View paper
- [17] Doubly Adaptive Scaled Algorithm for Machine Learning Using Second-Order Information View paper
- [18] AdaBB: Adaptive Barzilai-Borwein Method for Convex Optimization View paper
- [19] An Adaptive Stochastic Gradient Method with Non-negative Gauss-Newton Stepsizes View paper
- [20] Accelerated Convergence of Gradient Descent Using Adaptive Parameters View paper
- [21] Adaptive gradient descent for convex and non-convex stochastic optimization View paper
- [22] Accelerating incremental gradient optimization with curvature information View paper
- [23] ZO-JADE: Zeroth-Order Curvature-Aware Distributed Multi-Agent Convex Optimization View paper
- [24] Adaptive proximal algorithms for convex optimization under local Lipschitz continuity of the gradient View paper
- [25] Rapid DP Convex Optimization via Curvature-Aware (Second-Order) Algorithms. View paper
- [26] Projected Nesterovs proximal-gradient algorithm for sparse signal reconstruction with a convex constraint View paper
- [27] A Unified Scheme to Accelerate Adaptive Cubic Regularization and Gradient Methods for Convex Optimization View paper
- [28] Adaptive gradient descent without descent View paper
- [29] Adaptive Variant of the Frank-Wolfe Algorithm for Convex Optimization Problems View paper
- [30] On the curved geometry of accelerated optimization View paper
- [31] Projected nesterov's proximal-gradient algorithm for sparse signal recovery View paper
- [32] Hypodifferentials of nonsmooth convex functions and their applications to nonsmooth convex optimization View paper
- [33] Curvature-aided incremental aggregated gradient method View paper
- [34] Adaptive Proximal Gradient Method for Convex Optimization View paper
- [35] Fast Gradient Methods for Uniformly Convex and Weakly Smooth Problems View paper
- [36] Adaptive online gradient descent View paper
- [37] Gradient-Type Adaptive Methods for Relatively Lipschitz Convex Optimization Problems View paper
- [38] Convergence analysis of adaptive gradient methods under refined smoothness and noise assumptions View paper
- [39] Gradient-Variation Online Adaptivity for Accelerated Optimization with H\" older Smoothness View paper
- [40] On convergence of adam for stochastic optimization under relaxed assumptions View paper
- [41] Convergence Guarantees for RMSProp and Adam in Generalized-smooth Non-convex Optimization with Affine Noise Variance View paper
- [42] Provable Complexity Improvement of AdaGrad over SGD: Upper and Lower Bounds in Stochastic Non-Convex Optimization View paper
- [43] Adaptive Gradient Normalization and Independent Sampling for (Stochastic) Generalized-Smooth Optimization View paper
- [44] Decentralized Relaxed Smooth Optimization with Gradient Descent Methods View paper
- [45] A Comprehensive Framework for Analyzing the Convergence of Adam: Bridging the Gap with SGD View paper
- [46] Adaptive Generalized Conditional Gradient Method for Multiobjective Optimization View paper
- [47] Provable adaptivity of adam under non-uniform smoothness View paper
- [48] Tradeoffs between convergence rate and noise amplification for momentum-based accelerated optimization algorithms View paper
- [49] An Adaptive and Parameter-Free Nesterov's Accelerated Gradient Method for Convex Optimization View paper
- [50] Convergence rates of a momentum algorithm with bounded adaptive step size for nonconvex optimization View paper
- [51] A Totally Asynchronous Nesterovâs Accelerated Gradient Method for Convex Optimization View paper

- [52] Accelerated gradient descent escapes saddle points faster than gradient descent View paper
- [53] Accelerated Distributed Projected Gradient Descent for Convex Optimization with Clique-wise Coupled Constraints View paper
- [54] Accelerated policy gradient: On the nesterov momentum for reinforcement learning View paper
- [55] On the convergence of Nesterov's accelerated gradient method in stochastic settings View paper
- [56] Primal-dual accelerated gradient descent with line search for convex and nonconvex optimization problems View paper
- [57] Stochastic gradient accelerated by negative momentum for training deep neural networks View paper
- [58] Convex-concave programming: An effective alternative for optimizing shallow neural networks View paper
- [59] The first optimal acceleration of high-order methods in smooth convex optimization View paper
- [60] Towards simple and provable parameter-free adaptive gradient methods View paper
- [61] Optimal and parameter-free gradient minimization methods for convex and nonconvex optimization View paper
- [62] Achieving Linear Convergence with Parameter-Free Algorithms in Decentralized Optimization View paper
- [63] Methods for Convex -Smooth Optimization: Clipping, Acceleration, and Adaptivity View paper
- [64] DoWG Unleashed: An Efficient Universal Parameter-Free Gradient Descent Method View paper
- [65] A Parameter-Free Conditional Gradient Method for Composite Minimization under HÃ¶lder Condition View paper
- [66] How free is parameter-free stochastic optimization? View paper