Novelty Assessment Report

Paper: Gradient-Normalized Smoothness for Optimization with Approximate Hessians

PDF URL: https://openreview.net/pdf?id=Epu8Lm6VMK

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-30

Abstract

In this work, we develop new optimization algorithms that use approximate second-order information combined with the gradient regularization technique to achieve fast global convergence rates for both convex and non-convex objectives. The key innovation of our analysis is a novel notion called Gradient-Normalized Smoothness, which characterizes the maximum radius of a ball around the current point that yields a good relative approximation of the gradient field. Our theory establishes a natural intrinsic connection between Hessian approximation and the linearization of the gradient. Importantly, Gradient-Normalized Smoothness does not depend on the specific problem class of the objective functions, while effectively translating local information about the gradient field and Hessian approximation into the global behavior of the method. This new concept equips approximate second-order algorithms with universal global convergence guarantees, recovering state-of-the-art rates for functions with Hölder-continuous Hessians and third derivatives, quasi-self-concordant functions, as well as smooth classes in first-order optimization. These rates are achieved automatically and extend to broader classes, such as generalized self-concordant functions. We demonstrate direct applications of our results for global linear rates in logistic regression and softmax problems with approximate Hessians, as well as in non-convex optimization using Fisher and Gauss-Newton approximations.

Core Task Landscape

This paper addresses: **Optimization with Approximate Second-Order Information**

A total of **50 papers** were analyzed and organized into a taxonomy with **19 categories**.

Taxonomy Overview

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

- **Theoretical Foundations and Convergence Analysis**
- **Hessian Approximation Techniques**
- **Algorithm Design and Variants**
- **Implementation Strategies and Computational Efficiency**
- **Specialized Problem Classes**
- **Application Domains**

Complete Taxonomy Tree

- Optimization with Approximate Second-Order Information Survey Taxonomy
- Theoretical Foundations and Convergence Analysis
  - Convergence Theory for Approximate Hessian Methods ★ (4 papers)
  - [0] Gradient-Normalized Smoothness for Optimization with Approximate Hessians (Anon et al., 2026) View paper
  - [4] Newton-type methods for non-convex optimization under inexact Hessian information (Peng Xu, 2020) View paper
  - [14] Convergence of Newton-MR under inexact Hessian information (Yang Liu, 2021) View paper
  - [39] A subsampling line-search method with second-order results (El Houcine Bergou, 2022) View paper
  - Second-Order Approximation Theory (3 papers)
  - [3] New second-order and tensor methods in convex optimization (Doikov, 2021) View paper
  - [25] Convex optimization based on global lower second-order models (Nikita Doikov, 2020) View paper
  - [47] An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods (Renato D.C. Monteiro, 2013) View paper
  - Optimality Conditions and Theoretical Characterizations (3 papers)
  - [17] KKT conditions, first-order and second-order optimization, and distributed optimization: tutorial and survey (Ghojogh, 2021) View paper
  - [32] Detecting negative eigenvalues of exact and approximate Hessian matrices in optimization (W. Hare, 2022) View paper
  - [34] A second-order sequential optimality condition for nonlinear second-order cone programming problems (Ellen H. Fukuda, 2023) View paper
- Hessian Approximation Techniques
  - Structured Matrix Approximations (4 papers)
  - [8] Eva: Practical second-order optimization with kronecker-vectorized approximation (Zhang Lin, 2023) View paper
  - [28] Distributed second-order optimization using kronecker-factored approximations (Jimmy Ba, 2017) View paper
  - [30] A Modified Daiâ□□Liao Conjugate Gradient Method Based on a Scalar Matrix Approximation of Hessian and Its Application (Branislav Ivanov, 2023) View paper
  - [49] Legendre polynomial based approximation of the Hessian for quasi-Newton optimization (Stefan PaniÄ□, 2025) View paper
  - Stochastic and Sampling-Based Approximations (4 papers)
  - [13] Fast unconstrained optimization via hessian averaging and adaptive gradient sampling methods (O'Leary-Roseberry, 2024) View paper

- [18] Adahessian: An adaptive second order optimizer for machine learning (Zhewei Yao, 2021) View paper
- [40] Newton Meets Marchenko-Pastur: Massively Parallel Second-Order Optimization with Hessian Sketching and Debiasing (Romanov, 2024) View paper
- [43] Subspace-based Approximate Hessian Method for Zeroth-Order Optimization (Kim Dong-Yoon, 2025) View paper
- Gradient-Based Hessian Estimation (3 papers)
- [15] Accelerating SVRG via second-order information (R Kolte, 2015) View paper
- [27] Faster independent component analysis by preconditioning with Hessian approximations (Pierre Ablin, 2018) View paper
- [41] Second-order Information in First-order Optimization Methods (Hu, 2022) View paper

- Algorithm Design and Variants
  - Trust-Region and Cubic Regularization Methods (2 papers)
  - [7] A second order approximation technique for robust shape optimization (Adrian Sichau, 2012) View paper
  - [11] Second-Order Optimization (Grosse, 2021) View paper
  - Newton-Type and Quasi-Newton Methods (2 papers)
  - [26] A family of second-order methods for convex -regularized optimization (RH Byrd, 2016) View paper
  - [42] A Hessian inversion-free exact second order method for distributed consensus optimization (Dusan Jakovetic, 2022) View paper
  - Accelerated and Hybrid Methods (2 papers)
  - [33] Transformers Learn to Achieve Second-Order Convergence Rates for In-Context Linear Regression (Tian Qi Chen, 2023) View paper
  - [46] Information Newton's flow: second-order optimization method in probability space (Wang Yi-fei, 2020) View paper
- Implementation Strategies and Computational Efficiency
  - Distributed and Parallel Implementations (3 papers)
  - [20] Second order optimization made practical (Rohan Anil, 2020) View paper
  - [23] Scalable second order optimization for deep learning (Anil, 2020) View paper
  - [45] FopLAHD: Federated Optimization Using Locally Approximated Hessian Diagonal (Mrinmay Sen, 2023) View paper
  - Memory and Computation Optimization (3 papers)
  - [6] The iterative optimal brain surgeon: Faster sparse recovery by leveraging second-order information (Wu, 2024) View paper
  - [10] Second-order stochastic optimization for machine learning in linear time (Naman Agarwal, 2017) View paper
  - [16] Woodfisher: Efficient second-order approximation for neural network compression (Singh, 2020) View paper
  - Hessian-Free and Inversion-Free Methods (2 papers)
  - [5] Sylva: Sparse Embedded Adapters via Hierarchical Approximate Second-Order Information (Baorun Mu, 2024) View paper
  - [9] Gradient descent on neurons and its link to approximate second-order optimization (Benzing, 2022) View paper
- Specialized Problem Classes
  - Manifold and Geometric Optimization (1 papers)
  - [1] Riemannian optimization on the symplectic Stiefel manifold using second-order information (Jensen Rasmus, 2024) View paper
  - Convex and Structured Optimization (2 papers)
  - [21] First and second order convex approximation strategies in structural optimization (C. Fleury, 1989) View paper
  - [24] Robust Decision Aggregation with Second-order Information (Yuqi Pan, 2024) View paper
  - Non-Convex and Saddle-Point Optimization (2 papers)
  - [12] Second-order optimization for non-convex machine learning: An empirical study (Xu Peng, 2020) View paper
  - [29] Second-order optimization for neural networks (Martens, 2016) View paper
- Application Domains
  - Machine Learning and Neural Networks (1 papers)
  - [35] Review of second-order optimization techniques in artificial neural networks backpropagation (H. Tan, 2019) View paper
  - Engineering Design and Structural Optimization (6 papers)
  - [2] A new second-order approximation method for optimum design of structures (H. Ahmadvand, 2021) View paper
  - [22] Robust optimization utilizing the second-order design sensitivity information (Nam-Kyung Kim, 2010) View paper
  - [36] Reliability-based topology optimization using mean-value second-order saddlepoint approximation (Dimitrios Papadimitriou, 2018) View paper
  - [38] Improved two-point function approximations for design optimization (Liping Wang, 1995) View paper
  - [44] Reliability-based design optimization of structures using the second-order reliability method and complex-step derivative approximation (Junho Chun, 2021) View paper
  - [50] Robust design optimization using a non-intrusive second-order approximation of stochastic moments (J. KrÃ¼ger, 2024) View paper
  - Computational Chemistry and Molecular Optimization (2 papers)
  - [19] Analytical ab initio hessian from a deep learning potential for transition state optimization (Eric Chung-Yueh Yuan, 2024) View paper
  - [31] Deep Learning of ab initio Hessians for Transition State Optimization (Eric C. -Y. Yuan, 2024) View paper
  - Surrogate Modeling and Multifidelity Optimization (2 papers)
  - [37] Second-order approximation and fast multigrid solution of parabolic bilinear optimization problems (A. BorzÃ¬, 2015) View paper
  - [48] Second-order corrections for surrogate-based optimization with model hierarchies (Michael Eldred, 2004) View paper

Narrative

Core task: optimization with approximate second-order information. This field addresses the challenge of exploiting curvature information to accelerate convergence without incurring the prohibitive cost of exact Hessian computation and inversion. The taxonomy organizes the landscape into six main branches. Theoretical Foundations and Convergence Analysis establishes rigorous guarantees for methods that rely on inexact or sampled curvature, examining conditions under which approximate second-order steps preserve superlinear or fast linear rates. Hessian Approximation Techniques explores diverse strategies—from low-rank factorizations and Kronecker structures to stochastic subsampling and quasi-Newton updates—that trade off fidelity against computational expense. Algorithm Design and Variants encompasses the spectrum of practical schemes, including trust-region methods, line-search Newton variants, and hybrid approaches that blend first- and second-order information. Implementation Strategies and Computational Efficiency focuses on making these methods scalable through parallelization, memory-efficient data structures, and adaptive preconditioning. Specialized Problem Classes tailors approximate Hessian techniques to settings such as constrained optimization, saddle-point problems, and manifold optimization, while Application Domains demonstrates their impact in neural network training, inverse problems, and large-scale scientific computing.

Recent work has intensified around making second-order methods practical for deep learning and high-dimensional settings. A dense cluster of studies investigates how to construct cheap yet informative curvature approximations—ranging from diagonal or block-diagonal Hessian estimates to Kronecker-factored structures—that can be computed and inverted in near-linear time. Another active line examines the interplay between stochastic sampling and convergence guarantees, asking how subsampled Hessians or gradient covariances affect iteration complexity and generalization. The original paper ```json[0] sits squarely within the convergence theory branch, contributing rigorous analysis of how approximation errors propagate through Newton-type iterations. It complements foundational work on inexact Newton methods Inexact Hessian Newton[7] and recent studies on gradient-normalized smoothness Gradient-Normalized Smoothness[3], which together clarify when and why approximate second-order information suffices for fast convergence. By contrast, neighboring efforts such as Symplectic Stiefel Optimization[4] and Saddlepoint Topology Optimization[39] emphasize specialized geometric or constrained settings, highlighting the breadth of contexts in which approximate curvature plays a pivotal role.

Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

1. Newton-type methods for non-convex optimization under inexact Hessian information

Authors: Peng Xu, Farbod Roosta-Khorasani, Michael W. Mahoney, Fred Roosta | **Year/Venue**: 2020 | **URL**: View paper

Abstract

We consider variants of trust-region and adaptive cubic regularization methods for non-convex optimization, in which the Hessian matrix is approximated. Under certain condition on the inexact Hessian, and using approximate solution of the corresponding sub-problems, we provide iteration complexity to achieve Îµ\documentclass[12pt]{minimal} \usepackage{amsmath} \usepackage{wasysym} \usepackage{amsfonts} \usepackage{amssymb} \usepackage{amsbsy} \usepackage{mathrsfs} \usepackage{upgreek} \setlength...

Relationship Analysis

Both papers belong to the convergence theory for approximate Hessian methods, analyzing optimization algorithms under inexact second-order information. They overlap in studying Newton-type methods with Hessian approximations and establishing convergence rates for non-convex optimization. The original paper introduces Gradient-Normalized Smoothness as a universal framework that adapts to problem classes and achieves state-of-the-art rates across multiple smoothness regimes, while the candidate paper focuses specifically on trust-region and cubic regularization methods with a fixed Hessian approximation condition (spectral norm bounds) and emphasizes finite-sum optimization with sub-sampling strategies.

2. Convergence of Newton-MR under inexact Hessian information

Authors: Yang Liu, Fred Roosta | **Year/Venue**: 2021 | **URL**: View paper

Abstract

Recently, there has been a surge of interest in designing variants of the classical Newton-CG in which the Hessian of a (strongly) convex function is replaced by suitable approximations. This is mainly motivated by large-scale finite-sum minimization problems that arise in many machine learning applications. Going beyond convexity, inexact Hessian information has also been recently considered in the context of algorithms such as trust-region or (adaptive) cubic regularization for general non-con...

Relationship Analysis

Both papers belong to the Convergence Theory for Approximate Hessian Methods category, analyzing convergence under inexact Hessian information. The original paper introduces Gradient-Normalized Smoothness to unify convergence analysis across multiple problem classes (convex and non-convex) with approximate Hessians, achieving universal global rates that adapt to smoothness degrees. The candidate paper focuses specifically on Newton-MR for invex problems, using matrix perturbation theory to bound subspace distances rather than operator distances, and addresses rank-deficient Hessians through Moore-Penrose pseudo-inverses rather than gradient regularization.

3. A subsampling line-search method with second-order results

Authors: El Houcine Bergou, Youssef Diouane, E. Bergou, Vladimir Kunc, Y. Diouane, et al. (10 authors total) | **Year/Venue**: 2022 | **URL**: View paper

Abstract

In many contemporary optimization problems such as those arising in machine learning, it can be computationally challenging or even infeasible to evaluate an entire function or its derivatives. This motivates the use of stochastic algorithms that sample problem data, which can jeopardize the guarantees obtained through classical globalization techniques in optimization, such as a line search. Using subsampled function values is particularly challenging for the latter strategy, which relies upon ...

Relationship Analysis

Both papers belong to the convergence theory for approximate Hessian methods category, analyzing optimization algorithms that use inexact second-order information. The original paper introduces Gradient-Normalized Smoothness to analyze gradient-regularized Newton methods with approximate Hessians, establishing universal convergence rates across multiple smoothness classes (Hölder Hessian, Quasi-Self-Concordant, etc.) for both convex and non-convex objectives. The candidate paper focuses specifically on subsampling line-search methods with stochastic Hessian estimates, deriving worst-case complexity guarantees for reaching second-order stationary points in finite-sum problems, with emphasis on probabilistic accuracy and expected decrease properties rather than universal smoothness characterizations.

Contributions Analysis

Overall novelty summary. The paper introduces Gradient-Normalized Smoothness as a unifying framework for approximate second-order optimization, aiming to establish universal convergence guarantees across convex and non-convex settings. It resides in the 'Convergence Theory for Approximate Hessian Methods' leaf, which contains four papers total—a moderately populated node within the broader 'Theoretical Foundations and Convergence Analysis' branch. This placement indicates the work targets foundational convergence analysis rather than algorithm-specific implementation or domain applications, situating it in a core but not overcrowded research direction.

The taxonomy reveals neighboring leaves focused on 'Second-Order Approximation Theory' (three papers on mathematical frameworks for curvature approximation) and 'Optimality Conditions and Theoretical Characterizations' (three papers on KKT conditions and sequential optimality). The paper's emphasis on connecting Hessian approximation quality to gradient field linearization bridges these areas, potentially drawing on approximation theory while contributing convergence guarantees. Nearby branches address practical Hessian construction techniques and algorithmic variants, suggesting the work complements rather than duplicates existing implementation-focused studies.

Among twenty-three candidates examined, the Gradient-Normalized Smoothness concept (ten candidates, zero refutations) and universal convergence theory (three candidates, zero refutations) appear relatively novel within the limited search scope. The gradient-regularized

Newton method with approximate Hessians (ten candidates, one refutation) shows overlap with prior work, indicating this algorithmic component may have precedent. The statistics suggest the conceptual framework is less anticipated than the specific algorithmic instantiation, though the search examined only top-ranked semantic matches rather than exhaustive coverage.

Based on the limited literature search, the work's novelty appears concentrated in its theoretical lens—Gradient-Normalized Smoothness as a problem-agnostic convergence criterion—rather than in the algorithmic mechanics. The analysis does not capture the full breadth of optimization theory, and deeper investigation of gradient regularization techniques or alternative smoothness characterizations could reveal additional overlaps. The taxonomy context suggests the contribution fills a gap in unifying convergence analysis across problem classes, though the extent of this gap remains partially characterized.

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

Contribution 1: Gradient-Normalized Smoothness concept

Description: The authors introduce Gradient-Normalized Smoothness, a new local characterization that measures how well the gradient field can be linearly approximated in a neighborhood of the current point. This concept unifies the treatment of Hessian approximation errors and Taylor approximation errors, providing a problem-class-free framework for analyzing second-order methods.

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. GraN-GAN: Piecewise gradient normalization for generative adversarial networks

URL: View paper

Brief Assessment

The candidate paper focuses on gradient normalization for GANs to enforce piecewise Lipschitz constraints on discriminators, not on characterizing gradient field approximation quality for second-order optimization methods as in the original paper.

---

2. Psychophysical determination of boundaries and smoothness of color gradients

URL: View paper

Brief Assessment

The candidate paper focuses on psychophysical experiments for visual perception of color gradients in images, not on mathematical optimization theory. It examines perceived smoothness of color transitions, which is unrelated to the original paper's gradient-normalized smoothness concept for analyzing second-order optimization methods.

---

3. Inertial Bregman proximal gradient under partial smoothness

URL: View paper

Brief Assessment

The candidate paper (Inertial Bregman proximal gradient under partial smoothness) focuses on Bregman proximal gradient methods with smooth adaptable property (SMAD) for non-Lipschitz gradients, not on gradient-normalized smoothness as a local characterization of gradient field approximation quality.

---

4. An algebraic approach to surface reconstruction from gradient fields

URL: View paper

Brief Assessment

The candidate paper (Gradient Field HOG[71]) focuses on surface reconstruction from gradient fields in computer vision, specifically addressing integrability constraints. This is fundamentally different from the original paper's gradient-normalized smoothness concept for optimization algorithms with approximate Hessians.

---

5. Low-pass filtering sgd for recovering flat optima in the deep learning optimization landscape

URL: View paper

Brief Assessment

The candidate paper (GraN-GAN Normalization) focuses on low-pass filtering SGD for recovering flat optima in deep learning optimization landscapes. While both papers discuss smoothness concepts in optimization, the candidate does not address gradient-normalized smoothness as a local characterization that measures gradient field linear approximation in a neighborhood, which is the core novelty of the original contribution.

---

6. Image smoothing method based on global gradient sparsity and local relative gradient constraint optimization

URL: View paper

Brief Assessment

The candidate paper focuses on image smoothing using gradient sparsity and local relative gradient constraints for texture removal, not on optimization theory or Hessian approximation analysis for second-order methods.

---

7. Fast image super-resolution via local adaptive gradient field sharpening transform

URL: View paper

Brief Assessment

The candidate paper focuses on gradient field sharpening for image super-resolution, which is a computer vision technique. This is fundamentally different from the original paper's theoretical framework for analyzing second-order optimization methods through gradient-normalized smoothness.

---

8. A second gradient formulation for a 2D fabric sheet with inextensible fibres

URL: View paper

Brief Assessment

The candidate paper focuses on second gradient formulations for fabric sheets with inextensible fibers, not on gradient-normalized smoothness for optimization algorithms. These are entirely different research domains with no conceptual overlap.

---

9. A performance evaluation of gradient field hog descriptor for sketch based image retrieval

URL: View paper

Brief Assessment

The candidate paper focuses on gradient field HOG descriptors for sketch-based image retrieval, not on gradient-normalized smoothness for optimization. These are entirely different research domains with no conceptual overlap.

---

10. Gradient and smoothness regularization operators for geophysical inversion on unstructured meshes

URL: View paper

Brief Assessment

The candidate paper focuses on gradient operators for geophysical inversion on unstructured meshes in a completely different domain (geophysics/geology). It does not address gradient-normalized smoothness as a theoretical concept for optimization analysis.

Contribution 2: Universal convergence theory for approximate second-order methods

Description: The authors develop a unified convergence theory that automatically adapts to different problem classes and Hessian approximation quality. Their framework recovers existing state-of-the-art rates for various smoothness classes and extends to broader function classes such as generalized self-concordant functions, all without requiring prior knowledge of problem-specific parameters.

This contribution was assessed against **3 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Minimizing Quasi-Self-Concordant Functions by Gradient Regularization of Newton Method

URL: View paper

Brief Assessment

The candidate paper focuses specifically on quasi-self-concordant functions with gradient regularization of Newton method, achieving global linear rates. The original paper develops a broader universal framework with gradient-normalized smoothness that adapts to multiple problem classes including Hölder-continuous Hessians and handles inexact Hessians systematically.

---

2. Minimizing Quasi-Self-Concordant Functions by Gradient Regularization of Newton Method: N. Doikov

URL: View paper

Brief Assessment

The candidate focuses specifically on quasi-self-concordant functions with gradient regularization, not on a universal framework that automatically adapts to different problem classes and Hessian approximation quality as described in the original contribution.

---

3. Improved global performance guarantees of second-order methods in convex minimization: P. Dvurechensky, Y. Nesterov

URL: View paper

Brief Assessment

Improved global performance guarantees of second-order methods in convex minimization[61] focuses on self-concordant functions and path-following schemes, not on universal convergence theory for approximate second-order methods with Hölder-continuous Hessians or quasi-self-concordant functions as in the original paper.

Contribution 3: Gradient-regularized Newton method with approximate Hessians

Description: The authors propose Algorithm 1, which combines gradient regularization with approximate Hessian information. The method uses a universal step-size rule based on Gradient-Normalized Smoothness that automatically adjusts to both the objective's smoothness class and the Hessian approximation quality, achieving fast convergence without problem-specific tuning.

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 Regularized Newton Method for -Norm Composite Optimization Problems

URL: View paper

Brief Assessment

Newton-MR Complexity,[60] focuses on ℓq-norm composite optimization with subspace regularized Newton methods, not general gradient-regularized frameworks with universal step-size rules for diverse smoothness classes.

---

2. Convergence rates of regularized quasi-Newton methods without strong convexity

URL: View paper

Brief Assessment

Convergence rates of regularized quasi-Newton methods without strong convexity (Surrogate Hierarchies Corrections",[51]) focuses on quasi-Newton methods (specifically SR1) with cubic and gradient regularization under Kurdyka-Łojasiewicz property, not on the gradient-normalized smoothness framework or the universal step-size adaptation mechanism proposed in the original paper.

---

3. Accelerated adaptive cubic regularized Quasi-Newton methods

URL: View paper

Brief Assessment

The candidate paper title mentions 'Accelerated adaptive cubic regularized Quasi-Newton methods' and 'Non-Intrusive Stochastic Moments', which appears to focus on cubic regularization and stochastic methods rather than gradient regularization. The limited context provided does not contain sufficient detail to establish that similar prior work exists for the original paper's specific approach combining gradient regularization with approximate Hessians and universal step-size rules based on Gradient-Normalized Smoothness.

---

4. Randomized subspace regularized Newton method for unconstrained non-convex optimization

URL: View paper

Brief Assessment

Randomized subspace regularized Newton Global,[55] focuses on randomized subspace methods for non-convex optimization with theoretical local convergence analysis, while the original paper develops gradient-normalized smoothness theory for approximate Hessians with universal step-size rules across multiple smoothness classes.

---

5. Reduced order model hessian approximations in newton methods for optimal control

URL: View paper

Brief Assessment

Quasi-Newton Convergence Rates,[54] focuses on reduced-order model (ROM) Hessian approximations in optimal control contexts, not the gradient-normalized smoothness framework or universal step-size adaptation proposed in the original paper.

---

6. Minimizing Quasi-Self-Concordant Functions by Gradient Regularization of Newton Method: N. Doikov

URL: View paper

Prior Art Analysis

Minimizing Quasi-Self-Concordant Functions by Gradient Regularization of Newton Method[56] demonstrates prior work on gradient-regularized Newton methods. The candidate paper presents Algorithm 1 which uses gradient regularization with approximate Hessians in the form xk+1 = xk - (hk + ||∇f(xk)||/γk * b)^(-1)∇f(xk), where the regularization parameter is proportional to the gradient norm. This is substantively similar to the original paper's claimed novel Algorithm 1, though applied to a specific function class (quasi-self-concordant functions).

Evidence

Evidence 1 - **Rationale**: While the candidate uses a different smoothness characterization (quasi-self-concordance), it addresses the same problem of handling approximate Hessians in Newton methods, establishing that this approach predates the original paper's claimed novelty. - **Original**: we consider the following condition for our method: ‖∇2f(xk) - hk‖ ≤ c1 + c2‖∇f(xk)‖1-β, 0 ≤ β ≤ 1, (2) for certain c1, c2 ≥ 0, and β is a fixed approximation degree. - **Candidate**: we say that a convex function f is quasi-self-concordant with parameter m ≥ 0, if for all u,v ∈ e it holds d3 f (x)[u]2[v]≤ m‖u‖2 x ‖v‖, x ∈ dom f . (15)

Evidence 2 - **Rationale**: Both papers describe gradient regularization parameters that control step size and can reduce to gradient descent in limiting cases, showing the candidate established this methodology prior to the original paper. - **Original**: this parametrization ensures that each step is bounded, ‖xk+1 - xk‖ ≤γk, and for hk = 0 we obtain iterations of the normalized gradient descent - **Candidate**: note that in this algorithm we have a freedom of choosing the regularization parameter σk ≥ 0 for each iteration k ≥ 0. a simple choice that follows directly from our theory, is the constant one: σk ≡ m, ∀k, where m is the parameter of quasi-self-concordant

---

7. Complexity guarantees for nonconvex Newton-MR under inexact Hessian information

URL: View paper

Brief Assessment

Reduced Order Hessian[57] focuses on Newton-MR with inexact Hessian under a specific noise model (Definition 5) for nonconvex problems, using MINRES as the sub-problem solver. The original paper proposes gradient regularization with gradient-normalized smoothness for universal step-size adaptation across smoothness classes, which is a fundamentally different algorithmic framework and theoretical contribution.

---

8. Regularized Newton Method with Global Convergence

URL: View paper

Brief Assessment

The candidate paper focuses on gradient-norm regularization (λk = √h‖∇f(xk)‖) for exact Hessians with global $O(1/k^2)$ convergence. The original paper's contribution is specifically about combining gradient regularization with approximate Hessians using gradient-normalized smoothness γ(x,g) to handle inexact Hessian information, which is a distinct technical framework not present in the candidate.

---

9. Gradient descent finds the cubic-regularized nonconvex Newton step

URL: View paper

Brief Assessment

Gradient descent finds the cubic-regularized nonconvex Newton step,[59] focuses on solving the cubic-regularized subproblem (1) using gradient descent, not on developing a gradient-regularized Newton method with approximate Hessians for general optimization. The candidate addresses subproblem solution rather than the main algorithmic framework with universal step-size rules.

---

10. Adaptive Quasi-Newton and anderson acceleration framework with explicit global (accelerated) convergence rates

URL: View paper

Brief Assessment

Randomized Subspace Newton[58] focuses on subspace minimization with random directions and multisecant approximations, not gradient regularization with universal step-size rules based on Gradient-Normalized Smoothness.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Gradient-Normalized Smoothness for Optimization with Approximate Hessians View paper
- [1] Riemannian optimization on the symplectic Stiefel manifold using second-order information View paper
- [2] A new second-order approximation method for optimum design of structures View paper
- [3] New second-order and tensor methods in convex optimization View paper
- [4] Newton-type methods for non-convex optimization under inexact Hessian information View paper
- [5] Sylva: Sparse Embedded Adapters via Hierarchical Approximate Second-Order Information View paper
- [6] The iterative optimal brain surgeon: Faster sparse recovery by leveraging second-order information View paper
- [7] A second order approximation technique for robust shape optimization View paper
- [8] Eva: Practical second-order optimization with kronecker-vectorized approximation View paper
- [9] Gradient descent on neurons and its link to approximate second-order optimization View paper
- [10] Second-order stochastic optimization for machine learning in linear time View paper
- [11] Second-Order Optimization View paper
- [12] Second-order optimization for non-convex machine learning: An empirical study View paper
- [13] Fast unconstrained optimization via hessian averaging and adaptive gradient sampling methods View paper
- [14] Convergence of Newton-MR under inexact Hessian information View paper
- [15] Accelerating SVRG via second-order information View paper
- [16] Woodfisher: Efficient second-order approximation for neural network compression View paper
- [17] KKT conditions, first-order and second-order optimization, and distributed optimization: tutorial and survey View paper
- [18] Adahessian: An adaptive second order optimizer for machine learning View paper
- [19] Analytical ab initio hessian from a deep learning potential for transition state optimization View paper
- [20] Second order optimization made practical View paper
- [21] First and second order convex approximation strategies in structural optimization View paper
- [22] Robust optimization utilizing the second-order design sensitivity information View paper

- [23] Scalable second order optimization for deep learning View paper
- [24] Robust Decision Aggregation with Second-order Information View paper
- [25] Convex optimization based on global lower second-order models View paper
- [26] A family of second-order methods for convex -regularized optimization View paper
- [27] Faster independent component analysis by preconditioning with Hessian approximations View paper
- [28] Distributed second-order optimization using kronecker-factored approximations View paper
- [29] Second-order optimization for neural networks View paper
- [30] A Modified Daiâ Liao Conjugate Gradient Method Based on a Scalar Matrix Approximation of Hessian and Its Application View paper
- [31] Deep Learning of ab initio Hessians for Transition State Optimization View paper
- [32] Detecting negative eigenvalues of exact and approximate Hessian matrices in optimization View paper
- [33] Transformers Learn to Achieve Second-Order Convergence Rates for In-Context Linear Regression View paper
- [34] A second-order sequential optimality condition for nonlinear second-order cone programming problems View paper
- [35] Review of second-order optimization techniques in artificial neural networks backpropagation View paper
- [36] Reliability-based topology optimization using mean-value second-order saddlepoint approximation View paper
- [37] Second-order approximation and fast multigrid solution of parabolic bilinear optimization problems View paper
- [38] Improved two-point function approximations for design optimization View paper
- [39] A subsampling line-search method with second-order results View paper
- [40] Newton Meets Marchenko-Pastur: Massively Parallel Second-Order Optimization with Hessian Sketching and Debiasing View paper
- [41] Second-order Information in First-order Optimization Methods View paper
- [42] A Hessian inversion-free exact second order method for distributed consensus optimization View paper
- [43] Subspace-based Approximate Hessian Method for Zeroth-Order Optimization View paper
- [44] Reliability-based design optimization of structures using the second-order reliability method and complex-step derivative approximation View paper
- [45] FopLAHD: Federated Optimization Using Locally Approximated Hessian Diagonal View paper
- [46] Information Newton's flow: second-order optimization method in probability space View paper
- [47] An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods View paper
- [48] Second-order corrections for surrogate-based optimization with model hierarchies View paper
- [49] Legendre polynomial based approximation of the Hessian for quasi-Newton optimization View paper
- [50] Robust design optimization using a non-intrusive second-order approximation of stochastic moments View paper
- [51] Convergence rates of regularized quasi-Newton methods without strong convexity View paper
- [52] Regularized Newton Method with Global Convergence View paper
- [53] Accelerated adaptive cubic regularized Quasi-Newton methods View paper
- [54] Reduced order model hessian approximations in newton methods for optimal control View paper
- [55] Randomized subspace regularized Newton method for unconstrained non-convex optimization View paper
- [56] Minimizing Quasi-Self-Concordant Functions by Gradient Regularization of Newton Method: N. Doikov View paper
- [57] Complexity guarantees for nonconvex Newton-MR under inexact Hessian information View paper
- [58] Adaptive Quasi-Newton and anderson acceleration framework with explicit global (accelerated) convergence rates View paper
- [59] Gradient descent finds the cubic-regularized nonconvex Newton step View paper
- [60] A Regularized Newton Method for -Norm Composite Optimization Problems View paper
- [61] Improved global performance guarantees of second-order methods in convex minimization: P. Dvurechensky, Y. Nesterov View paper
- [62] Minimizing Quasi-Self-Concordant Functions by Gradient Regularization of Newton Method View paper
- [63] GraN-GAN: Piecewise gradient normalization for generative adversarial networks View paper
- [64] Psychophysical determination of boundaries and smoothness of color gradients View paper
- [65] Image smoothing method based on global gradient sparsity and local relative gradient constraint optimization View paper
- [66] Low-pass filtering sgd for recovering flat optima in the deep learning optimization landscape View paper
- [67] Fast image super-resolution via local adaptive gradient field sharpening transform View paper
- [68] A performance evaluation of gradient field hog descriptor for sketch based image retrieval View paper
- [69] Gradient and smoothness regularization operators for geophysical inversion on unstructured meshes View paper
- [70] A second gradient formulation for a 2D fabric sheet with inextensible fibres View paper
- [71] An algebraic approach to surface reconstruction from gradient fields View paper
- [72] Inertial Bregman proximal gradient under partial smoothness View paper