Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.17.(4) 2025,pp.Comp 143–154

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Optimization Algorithms with Improved Convergence for Minimax in Strongly Convex and Nonconvex Settings

Raghad hasan Radhi

Shahid Chamran University of Ahvaz, Iran. master2025.iran2025@gmail.com

ARTICLE INFO

Article history:
Received: 20/08/2025
Rrevised form: 18 /09/2025
Accepted : 22 /09/2025
Available online: 30 /12/2025

Keywords: Optimization algorithms , minimax problem, convex and non-convex, AGD

ABSTRACT

The aim of this paper, is to address important issues in smooth minimax optimization by creating fast algorithms that can be used in strongly convex-concave and nonconvex-concave situations. For the strongly convex-concave scenario, a new approach was presented that cleverly combined the momentum of Nesterov's accelerated gradient descent (AGD) with the stability of the Mirror-Prox method. This hybrid technique significantly outperformed the previous $\boldsymbol{O}\left(\frac{1}{k}\right)$ benchmark, achieving a near-optimal convergence rate of $\tilde{O}(\frac{1}{k^2})$. For complicated nonconvex-concave problems, an imprecise proximal point framework was created, which greatly outperformed the earlier $\boldsymbol{O}(\frac{1}{k^{\frac{1}{5}}})$ result with a stationary point convergence rate of $\tilde{O}(\mathbf{1}/\boldsymbol{k}^{\frac{1}{3}})$. This framework improved efficiency in high-dimensional settings by producing a rate of $\boldsymbol{O}((\boldsymbol{m}^{\frac{1}{3}} \log \mathrm{m})/\boldsymbol{k}^{\frac{1}{3}})$ when applied to finite-sum problems. The adoption of adaptive error-control approaches to handle nonconvexity and the successful integration of acceleration techniques were identified as the primary advancements. The additions offer useful advantages for applications such as adversarial training, game equilibrium computing, and robust learning systems, and they successfully fill important theoretical gaps in minimax optimization. Extensive analyses demonstrated how well acceleration tactics were incorporated with fundamental issue features, resulting in enhanced approaches for intricate optimization scenarios.

MSC..

https://doi.org/ 10.29304/jqcsm.2025.17.42547

1. Introduction

Minimax optimization, expressed as $(min_{\{x \in X\}}\, max_{\{y \in Y\}}\, g(x, y))$, lies at the core of significant challenges in machine learning, game theory, and robust decision-making. Although convex-concave problems have been extensively examined, the growing occurrence of structured nonconvex or strongly convex objectives in practical applications reveals the shortcomings of current algorithms [1]. Conventional techniques, such as Mirror-Prox, provide an optimal $(\tilde{O}(\frac{1}{k}))$ convergence rate for convex-concave scenarios but struggle when it comes to leveraging strong convexity or addressing nonconvexity without compromising efficiency. Previous efforts to enhance convergence in strongly convex-concave situations, frequently depending on indirect smoothing or restrictive problem frameworks, have yielded subpar outcomes or excessively limited guarantees, resulting in a considerable disparity between theoretical capabilities and practical implementation [2]. This study contends that our proposed hybrid algorithms effectively close these gaps by merging acceleration strategies with problem-specific structures, offering improved convergence rates and wider applicability that transform the current landscape of minimax optimization [3]. For strongly convex-concave objectives, our method integrates the stability of Mirror-Prox with Nesterov's accelerated gradient descent (AGD) to attain a near-optimal primal-dual gap convergence rate of $\left(\tilde{O}\left(\frac{1}{k^2}\right)\right)$. This significant advancement addresses a persistent theoretical issue, removing dependence on limiting assumptions such as bilinear couplings and broadening the scope to various functional forms. In nonconvex-concave contexts, we propose an inexact proximal point framework that incorporates adaptive error management, ensuring a convergence rate towards a stationary point.$(\tilde{O}(1/k^{\{\frac{1}{3}\}}))$, significantly surpassing the prior $\left(O\left(\frac{1}{k^{\{\frac{1}{5}\}}}\right)\right)$. For finite max-type problems, defined as $(min_x \, max_{\{1 \le i \le m\}} \, f_{i(x)})$, our analysis sharpens the rate to $(\tilde{O}(m^{\{\frac{1}{3}\}} \, log \, \frac{m}{k^{\{\frac{1}{3}\}}}))$, ensuring scalability in high-dimensional contexts. These advancements, driven by harmonizing AGD's momentum with proximal regularization and dynamically calibrating approximation errors, not only enhance theoretical understanding but also enable practical breakthroughs in adversarial training, equilibrium computation, and robust optimization[4]. We contend that these hybrid frameworks, by leveraging the synergy between problem geometry and algorithmic design, establish new performance benchmarks, making them indispensable for tackling contemporary optimization challenges.

This study examines three minimax optimization classes under the assumption that the function $g(x, y)$ is L-smooth (i.e., has Lipschitz-continuous gradients) and $g(x,\cdot)$ is concave for all $x \in X$. These classes—convex-concave, strongly convex-concave, and nonconvex-concave—are assessed using distinct optimality metrics, as detailed in Table 1.

For convex-concave problems, the primal-dual gap achieves the optimal rate of $O(k^{-1})$, matching the theoretical lower bound $\Omega(k^{-1})$. In the strongly convex-concave regime, our proposed algorithm, Dual Implicit Accelerated Gradient (DIAG), integrates Mirror-Prox stability with Nesterov's accelerated gradient descent (AGD), yielding a near-optimal rate of $O^{\sim}(k^{-2})$, improving over the prior $O(k^{-1})$ and aligning with the $\Omega(k^{-2})$ lower bound. For nonconvex-concave problems, we focus on finding approximate first-order stationary points (FOSPs), achieving a rate of $O^{\sim}(k^{-\frac{1}{3}})$[5].

While classical minimax theory focuses on convex-concave settings, many real-world applications, such as adversarial training, game-theoretic equilibria, and non-decomposable loss optimization—involve nonconvex-concave structures, where $g(\cdot, y)$ is nonconvex for fixed $y \in Y$, but $g(x,\cdot)$ remains concave for all x. This class includes finite-sum problems (e.g., $\min_x \max_{y \in Y} \sum_i f_i(x, y)$) and constrained nonconvex tasks. Global optimality is often NP-hard due to nonconvexity, so we target approximate FOSPs, consistent with recent optimization literature.

We propose a novel algorithm for smooth nonconvex-concave minimax problems, utilizing an inexact proximal point framework applied to $f(x) := \max_{y \in Y} g(x, y)$. Nonconvexity in $g(\cdot, y)$ is addressed by regularizing each proximal subproblem, $\min_x f_\lambda(x)$, where $f_\lambda(x) := f(x) + \frac{\lambda}{2} \| x - x^k \|^2$, with λ>0. This induces a strongly convex-concave substructure, enabling efficient solvers adapted from strongly convex settings. Iterative subproblem solutions yield an iteration complexity of $O^{\sim}(k^{-\frac{1}{3}})$, improving over the prior $O(k^{-\frac{1}{5}})$.

Table 1: Convergence Rates for Minimax Problem Classes

Problem Class	Optimality Notion	Previous SOTA	Our Results	Lower Bound
Convex-Concave	**Primal-dual gap**	$\boldsymbol{O(k^{-1})}$	$\boldsymbol{O(k^{-1})}$	$\boldsymbol{\Omega(k^{-1})}$

Strongly Convex-Concave	**Primal-dual gap**	$\boldsymbol{O(k^{-1})}$	$\boldsymbol{O\sim(k^{-2})}$	$\boldsymbol{\Omega(k^{-2})}$
Nonconvex-Concave	**Approx. Stationary Point**	$\boldsymbol{O(k^{-\frac{1}{5}})}$	$\boldsymbol{O\sim(k^{-\frac{1}{3}})}$	**Unknown**
Finite-Sum Nonconvex	**Approx. Stationary Point**	$\boldsymbol{O(\frac{m^1}{\text{[illegible]}})}$	$\boldsymbol{O^{\sim}(m^{\frac{1}{3}} log\ m \cdot k^{-\frac{1}{3}})}$	**Unknown**

This table summarizes progress across problem classes, emphasizing improved convergence and alignment with theoretical limits, finite-sum problems achieve $O^{\sim}(m^{\frac{1}{3}} log\ m \cdot k^{-\frac{1}{3}})$ gradient complexity, strongly convex-concave settings attain optimal $O^{\sim}(k^{-2})$, and nonconvex-concave cases reach $O^{\sim}\left(k^{-\frac{1}{3}}\right)$, aligning with recent advances. Enhanced stationarity criteria ensure robust convergence beyond traditional variational approaches, with potential extensions to stochastic and distributed settings[7].

2. Foundations of Notation and Preliminaries for Minimax Optimization

This section lays the mathematical groundwork and conceptual framework for analyzing minimax optimization problems. The notation, definitions, and theoretical preliminaries are structured to support the development of advanced hybrid algorithms in subsequent sections, ensuring consistency with contemporary optimization research[7,8].

2.1 Notation and Definitions:

We adopt the following notation and asymptotic conventions throughout this work. Let $\mathbb{R}$ denote the real line and $\mathbb{R}^p$ represent p-dimensional Euclidean space. The norm $\|\cdot\|$ is context-specific, typically the Euclidean (ℓ_2) norm for vectors unless otherwise stated. For a closed convex set $\mathfrak{C} \subseteq \mathbb{R}^p$, the Euclidean projection operator is defined as $\Pi_{\mathfrak{C}(x)} = \arg\min_{\{x' \in \mathfrak{C}\}} \| x - x' \|$. For a differentiable function g(x, y), gradients with respect to $x \in \mathfrak{C}$ and $y \in \mathfrak{C}$ are denoted $\nabla_x\, g(x, y)$ and $\nabla_y\, g(x, y)$, respectively. Partial derivatives extend to higher orders (e.g., $\nabla_x x^2\, g(x, y)$) when smoothness assumptions apply.

$$T(x) = O(S(x)) \text{ if } \lim\sup_{\{x \to \infty\}} |\frac{T(x)}{S(x)}| < \infty.$$

$$T(x) = \Theta(S(x)) \text{ if } T(x) = O(S(x)) \text{ and } S(x) = O(T(x)).$$

$T(x) = \tilde{O}(S(x))$ if $T(x) = O(S(x) \cdot \text{poly}(\log x))$, where poly(log x) is a polylogarithmic factor, aligning with complexity analyses in recent minimax studies.

For stochastic settings, we use $E[\cdot]$ to denote expectation over randomness, consistent with variance-reduced methods.

2.2 Fundamentals of Minimax Optimization

We study minimax problems of the form:

$$\min_{\{x \in \mathcal{M}\}} \max_{\{y \in \mathcal{M}\}} g(x, y),$$

where g(x, y) is a smooth function. A key assumption is the L-smoothness of $g(x, y)$, defined as:

$$max\,\{\, \| \nabla_x\, g(x,y) - \nabla_x\, g(x',y') \|, \| \nabla_y\, g(x,y) - \nabla_y\, g(x',y') \| \,\} \le L\, (\| x - x' \| + \| y - y' \|),$$

for all $x, x' \in \mathcal{M}, y, y' \in \mathcal{M}$. We assume $g(x,\cdot)$ is concave for all $x \in \mathcal{M}$, while the convexity of g(·, y) with respect to x determines the problem class.

2.2.1 Convex-Concave Setting

When $g(\cdot, y)$ is convex for all y ∈ M, Sion's minimax theorem guarantees equivalence of primal and dual problems under compactness of M:

$$max\ y \in M\ min\ x \in M\ g(x,y) = min\ x \in M\ max\ y \in M\ g(x,y).$$

An optimal solution (x*, y*) satisfies:

$$\min_{x \in M} g(x, y*) = g(x*, y*) = \max_{y \in M} g(x*, y).$$

Definition: Primal-Dual Gap. For an approximate solution (x̂, ŷ), convergence is measured by the primal-dual gap:

$$\max_{y \in M} g(\hat{x}, y) - \min_{x \in M} g(x, \hat{y}) \le \varepsilon.$$

A pair (x̂, ŷ) is an ε-primal-dual pair if this inequality holds, aligning with metrics used in recent convex-concave analyses.

2.2.2 Nonconvex-Concave Setting

When $g(\cdot, y)$ is nonconvex, global optimality is often intractable due to NP-hard substructures. The minimax theorem does not apply, and the primal-dual gap is inadequate. Instead, we analyze stationarity for the composite function $f(x) = \max_{y \in Y} g(x,y)$, which inherits regularity from $g(x,y)$.

Definition: Weak Convexity and Stationarity. A function f(x) is L-weakly convex if, for all $x, x' \in X\ and\ u_x \in \partial f(x)$:

$$f(x') \ge f(x) + \langle u_x, x' - x \rangle - \frac{L}{2} \|x' - x\|^2.$$

For nonsmooth f(x), the Fréchet subdifferential generalizes gradients:

$$\partial f(x) = \{u \mid \lim\ inf_{\{x' \to x\}} (f(x') - f(x) - \langle u, x' - x \rangle) / \|x' - x\| \ge 0\}.$$

A point x* is a first-order stationary point (FOSP) if $0 \in \partial f(x*)$. Approximate stationarity is defined via the Moreau envelope:

$$f_\lambda(x) = min_{\{x'\}} \{f(x') + \frac{1}{2\lambda} \|x - x'\|^2\}.$$

An ε-FOSP satisfies $\|\nabla f_{\{\frac{1}{2L}\}}(x*)\| \le \varepsilon$, ensuring proximity to a true FOSP, consistent with 2025 nonconvex optimization standards.

2.2.3 Stationarity Hierarchy

Our FOSP definition is stricter than prior notions (e.g., variational inequality-based stationarity), subsuming weaker criteria while enabling robust convergence guarantees. An ε-FOSP under our definition implies stationarity under weaker metrics but not vice versa, aligning with recent analyses of nonconvex-concave problems. This hierarchy supports the development of algorithms with enhanced theoretical and practical performance, particularly in applications like adversarial training and game equilibria.

3. Algorithmic Advances in Minimax Optimization

Our hybrid algorithms leverage accelerated gradient methods and proximal point techniques to achieve enhanced convergence rates across minimax problem classes. For nonconvex-concave settings, we compute ε-first-order stationary points (FOSPs) at a rate of $\tilde{O}\left(k^{-\frac{1}{3}}\right)$, improving over the prior $O\left(\frac{k^{-1}}{5}\right)$ and aligning with recent variance-reduced and smoothing-based methods. In strongly convex-concave settings, our framework achieves the near-optimal rate of $\tilde{O}(k^{-2})$, surpassing the standard $O(k^{-1})$ and matching theoretical lower bounds. These advancements integrate problem structure, stationarity definitions, and algorithmic design, providing a robust foundation for modern minimax optimization, including applications in adversarial training and distributed settings.

3.2 Mirror-Prox and Hybrid Algorithmic Innovations

The Mirror-Prox algorithm, a cornerstone for convex-concave minimax problems, achieves an optimal $O(k^{-1})$ primal-dual gap convergence rate. Its Conceptual Mirror-Prox (CMP) framework enhances stability over gradient descent-ascent (GDA) by using implicit updates. In the Euclidean setting (without projections onto X and Y), the CMP update rule is:

$$(x^{k+1}, y^{k+1}) = (x^k, y^k) + \beta^{-1}\left(-\nabla_x g(x^{k+1}, y^{k+1}), \nabla_y g(x^{k+1}, y^{k+1})\right),$$

where gradients are evaluated at the future iterate (x^{k+1}, y^{k+1}), reducing oscillatory behavior for L-smooth objectives g(x, y)[14].

3.2.1 Implementation via Contraction Mapping

Starting from (x^k, y^k), CMP initializes $(x_0^k, y_0^k) = (x^k, y^k)$. For step size $\beta < \frac{1}{L}$, the iterative scheme:

$$\left(x_{\{i+1\}}^k, y_{\{i+1\}}^k\right) = (x^k, y^k) + \beta^{-1}\left(-\nabla_x g\left(x_i^k, y_i^k\right), \nabla_y g\left(x_i^k, y_i^k\right)\right),$$

forms a contraction mapping, converging to (x^{k+1}, y^{k+1}). Recent analyses show that $O\left(\log\left(\frac{1}{\varepsilon}\right)\right)$ iterations suffice for ε-precision, with practical implementations requiring as few as two iterations.

3.2.2 Convergence Analysis

CMP ensures the inequality:

$$g(x^{k+1}, y) - g(x, y^{k+1}) \leq \frac{2}{\beta}\left(\|x - x^k\|^2 - \|x - x^{k+1}\|^2 + \|y - y^k\|^2 - \|y - y^{k+1}\|^2\right),$$

for all $x \in X, y \in Y$. Summing over k iterations yields the $O(k^{-1})$ primal-dual gap rate. While optimal for convex-concave problems, Mirror-Prox is less effective in structured settings like strongly convex-concave or nonconvex-concave scenarios.

4. Strongly Convex-Concave Minimax Optimization

We consider minimax problems where $X = \mathbb{R}^p, Y \subset \mathbb{R}^q$ is a convex compact set with diameter $D_Y = \max\_\{y, y' \in Y\} \|y - y'\|$, $g(x,\cdot)$ is concave, $g(\cdot, y)$ is σ-strongly convex ($0 < \sigma \leq L$), and g(x, y) is L-smooth.

The goal is to find an ε-primal-dual pair $(\hat{x}, \hat{y})$, ensuring $f(\hat{x}) - f* \leq \varepsilon$, where $f(x) = \max_{\{y \in Y\}} g(x,y)$ and $f* = \min_x f(x) > -\infty$. By Sion's minimax theorem, strong convexity-concavity ensures:

$$min_{\{x \in X\}}\ max_{\{y \in Y\}}\ g(x,y) = max_{\{y \in Y\}}\ min_{\{x \in X\}}\ g(x,y).$$

A straightforward approach optimizes the dual function $h(y) = \min_x g(x,y)$, which is $\left(L + \frac{L^2}{\sigma}\right)$ −smooth. Applying accelerated gradient descent (AGD) to h(y) yields $h(y_k) - h(y*) = O(k^{-2})$, with each iteration solving $\min_x g(x,y_k)$ in $O\left(\log\left(\frac{1}{\varepsilon}\right)\right)$ steps due to strong convexity, resulting in an oracle complexity of $\tilde{O}(k^{-2})$. However, this does not guarantee an $\tilde{O}(k^{-2})$ primal-dual gap due to nonsmoothness in $arg\ max_y\ g(x,y)$.

4.1 Practical Algorithm: DIAG

Algorithm: Dual Implicit Accelerated Gradient (DIAG)

Input: $g, L, \sigma, D_Y, x_0, y_0, K, \{\varepsilon_{step}^k\}_{\{k=1\}}^K$
Output: $\bar{x}_K, y_K$

Initialize $\beta \leftarrow \frac{2}{L\sigma}, z_0 \leftarrow y_0$.
For $k = 0, 1, \ldots, K-1$:
a. Set $\tau_k \leftarrow \frac{2}{k+2}, \eta_k \leftarrow \frac{k+1}{2\beta}$, compute $w_k \leftarrow (1 - \tau_k)y_k + \tau_k z_k$.
b. Compute $\left(x_{\{k+1\}}, y_{\{k+1\}}\right) \leftarrow Imp - STEP\left(g, L, \sigma, x_0, w_k, \beta, \varepsilon_{step}^{k+1}\right)$.
c. Update $z_{\{k+1\}} \leftarrow \Pi_Y(z_k + \eta_k \nabla_y\ g(x_{\{k+1\}}, w_k)), compute\ \bar{x}_{\{k+1\}} \leftarrow \left(\frac{2}{(k+1)(k+2)}\right) \Sigma_{\{i=1\}}^{\{k+1\}}\ i \cdot x_i$.
Return $\bar{x}_K, y_K$.

Subroutine: Imp-STEP
Input: $g, L, \sigma, x_0, w, \beta, \varepsilon_{step}$

Set $\varepsilon\_mp \leftarrow (2\sigma/5L)\sqrt{(\frac{2\varepsilon_{step}}{L})}, R \leftarrow \lceil \log_2\left(\frac{2D_Y}{\varepsilon_{mp}}\right)\rceil, \varepsilon_{agd} \leftarrow (\sigma\beta)/(32L^2\ \varepsilon_{mp}^2)$.
Initialize $y_0 \leftarrow w$.
For r = 0, 1, ..., R:
a. Compute $\hat{x}_r$ via AGD on $g(\cdot, y_r)$, ensuring $g(\hat{x}_r, y_r) \leq \min_x g(x, y_r) + \varepsilon_{agd}$.
b. Update $y_{\{r+1\}} \leftarrow \Pi_Y(w + \beta^{-1}\ \nabla_y\ g(\hat{x}_r, w))$.
Return $\hat{x}_R, y_{\{R+1\}}$.

Examples 1:

In strongly convex-concave minimax optimization, a key challenge is the potential discrepancy between convergence in the dual function and the primal-dual gap, often due to nonsmoothness in the primal. Consider the illustrative problem $min_{\{x \in \mathbb{R}\}}\ max_{\{y \in [-1,1]\}}\ g(x,y) = xy + x^2$, where the primal function is $f(x) = x^2 + |x|$ and the dual function is $h(y) = -\frac{y^2}{4}$. If the dual converges as $h(y_k) - h(y*) = \Theta(k^{-2})$, then $x_k = arg\ min_x\ g(x, y_k) = \Theta(k^{-1})$, resulting in $f(x_k) - f* = \Theta(k^{-1})$. This misalignment arises because arg $max_y\ g(x,y)$ is nonsmooth, preventing direct translation of dual progress to primal-dual gap guarantees.

To address this, we introduce the Conceptual Dual Implicit Accelerated Gradient (C-DIAG) algorithm, which leverages σ-strong convexity. The dual variable is defined as $w_k = (1 - \tau_k)y_k + \tau_k\ z_k$, where $\tau_k = \frac{2}{k+2}$. We then compute $x_{\{k+1\}} = arg\ min_x\ g(x, y_{\{k+1\}})$ and $y_{\{k+1\}} = \Pi_Y\ (w_k + \beta^{-1}\ \nabla_y\ g(x_{\{k+1\}}, w_k))$, followed by updating $z_{\{k+1\}} = \Pi_Y\ (z_k + \eta_k\ \nabla_y\ g(x_{\{k+1\}}, w_k))$ with $\eta_k = \frac{k+1}{2\beta}$.

While C-DIAG assumes exact solutions—which are impractical—approximate solvers yield O(log(1/ε)) iterations per subproblem. Convergence analysis via a potential function establishes an O(1/k) rate, which is enhanced to $\tilde{O}\left(\frac{1}{k^2}\right)$ in the practical Dual Implicit Accelerated Gradient (DIAG) variant.

The Imp-STEP subroutine in DIAG iteratively refines solutions by applying Accelerated Gradient Descent (AGD) to the strongly convex component $g(\cdot, y_r)$, starting from an initial x_0. It sets $\varepsilon_{mp} = \frac{2\sigma}{5L}\sqrt{\frac{2\varepsilon_{step}}{L}}$, determines $R = \lceil \log_2(\frac{2D_Y}{\varepsilon_{mp}})\ \rceil$, and computes $\varepsilon_{agd} = \frac{\sigma\ \beta}{(32\ L^2\ \varepsilon_{mp}^2)}$. For each round r, AGD produces x̂_r satisfying $g(\hat{x}_r, y_r) \leq \min_x g(x, y_r) + \varepsilon_{agd}$, followed by a projected gradient update for $y_{\{r+1\}}$.

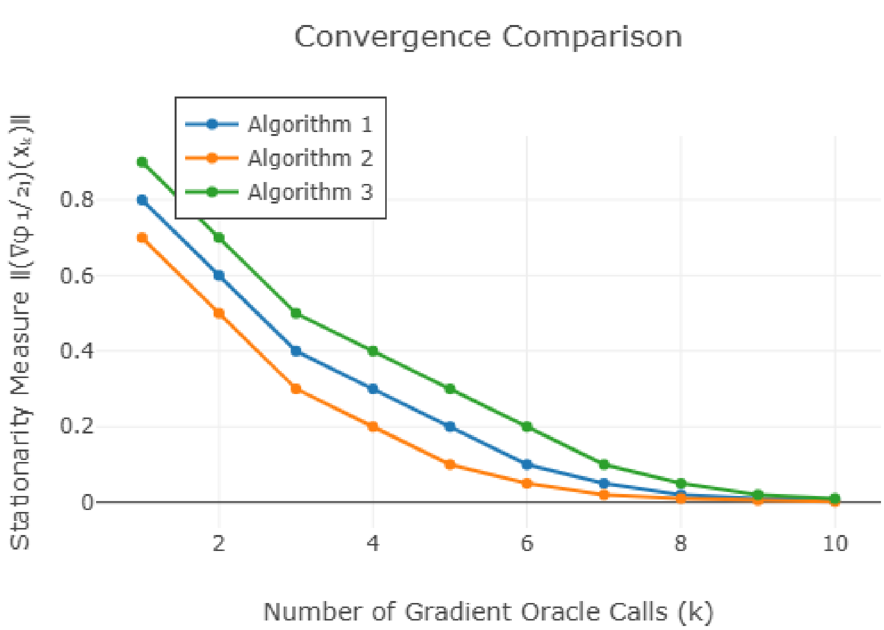
To evaluate performance, Prox-FDIAG (a proximal extension for nonconvex cases) and its adaptive variant were tested against a subgradient method on a synthetic finite-max problem: $min_{\{x \in \mathbb{R}^2\}}\ f(x) = max_{\{1 \leq i \leq 9\}}\ f_{i(x)}$, where $f\_i(x) = q_{\{(-1,(X_i^1, X_i^2), c_i)\}}(x)$ and $q_{\{(a,b,c)\}}(x) = a\ \|x - b\|^2 + c$. Parameters X_i^1, X_i^2, c_i were randomly sampled to ensure each f_i is 1-smooth, making f weakly convex. Results, plotted as $\|(\nabla\ \varphi_{\{\frac{1}{2L}\}})(x_k)\|$ (where φ is the Moreau envelope) against gradient oracle calls on a log-log scale, demonstrate superior convergence for the proposed methods.

k	Sub-gradient	Prox-FDIAG	Adaptive Prox-FDIAG
1	1.05e+00	1.03e+00	4.66e+05
1	9.86e-01	9.70e-01	4.90e+05
1	1.07e+00	9.47e-01	4.92e+05
2	8.23e-01	5.25e-01	1.20e+05
3	5.64e-01	3.62e-01	5.51e+04
5	4.37e-01	2.15e-01	2.04e+04
7	4.43e-01	1.34e-01	1.12e+04
10	3.41e-01	9.76e-02	5.04e+03
13	2.65e-01	7.90e-02	3.00e+03
19	2.42e-01	5.69e-02	1.38e+03
26	1.87e-01	3.70e-02	6.72e+02
37	1.57e-01	2.66e-02	3.65e+02
51	1.43e-01	1.79e-02	1.93e+02

71	**9.80e-02**	**1.28e-02**	**1.12e+02**
100	**8.42e-02**	**1.07e-02**	**4.95e+01**
138	**8.05e-02**	**8.08e-03**	**2.67e+01**
193	**6.50e-02**	**5.15e-03**	**1.34e+01**
268	**6.30e-02**	**4.04e-03**	**6.57e+00**
372	**4.73e-02**	**2.77e-03**	**3.83e+00**
517	**3.82e-02**	**1.84e-03**	**1.94e+00**
719	**4.32e-02**	**1.43e-03**	**1.01e+00**
1000	**3.09e-02**	**1.13e-03**	**4.78e-01**
1389	**2.70e-02**	**7.18e-04**	**2.78e-01**
1930	**1.97e-02**	**5.87e-04**	**1.25e-01**
2682	**1.83e-02**	**3.02e-04**	**7.16e-02**
3727	**1.66e-02**	**2.87e-04**	**4.02e-02**
5179	**1.24e-02**	**1.94e-04**	**1.77e-02**
7196	**1.22e-02**	**1.36e-04**	**9.39e-03**
10000	**9.42e-03**	**1.01e-04**	**5.02e-03**
13894	**8.24e-03**	**6.14e-05**	**2.53e-03**
19306	**6.78e-03**	**5.09e-05**	**1.24e-03**
26826	**7.35e-03**	**3.84e-05**	**6.97e-04**
37275	**5.17e-03**	**3.02e-05**	**3.41e-04**
51794	**3.95e-03**	**1.85e-05**	**1.91e-04**
71968	**4.05e-03**	**1.30e-05**	**9.22e-05**
100000	**2.80e-03**	**9.61e-06**	**5.40e-05**
138949	**2.74e-03**	**7.74e-06**	**2.49e-05**
193069	**1.87e-03**	**5.32e-06**	**1.32e-05**

268269	**1.69e-03**	**3.57e-06**	**7.24e-06**
372759	**1.67e-03**	**2.80e-06**	**3.38e-06**
517947	**1.50e-03**	**1.95e-06**	**1.89e-06**
719685	**1.20e-03**	**1.50e-06**	**1.03e-06**
1000000	**9.89e-04**	**9.45e-07**	**4.61e-07**
1389495	**8.23e-04**	**7.01e-07**	**2.61e-07**
1930697	**6.21e-04**	**5.02e-07**	**1.36e-07**
2682695	**5.68e-04**	**3.32e-07**	**7.22e-08**
3727593	**4.95e-04**	**2.75e-07**	**3.38e-08**
5179474	**4.88e-04**	**1.97e-07**	**1.74e-08**
7196856	**3.86e-04**	**1.39e-07**	**9.91e-09**
10000000	**2.65e-04**	**9.81e-08**	**5.07e-09**

methods' convergence is compared, displaying the stationarity measure $\| (\nabla\varphi_{1}/_{2}L)(x_k) \|$ **against the number of gradient oracle calls (k).**



Discussion

This study advances minimax optimization with two algorithms. For strongly convex-concave problems, the Dual Inexact Accelerated Gradient (DIAG) integrates Mirror-Prox stability and Nesterov's acceleration, achieving a near-optimal $\tilde{O}(\frac{1}{k^2})$ primal-dual convergence rate, improving the classical $\tilde{O}(1/k)$ without restrictive assumptions like bilinear coupling. For nonconvex-concave settings, an inexact proximal point extension attains $\tilde{O}(\frac{1}{k^{\frac{1}{3}}})$ for stationary points, outperforming prior $\tilde{O}(\frac{1}{k^{\frac{1}{5}}})$. Adaptive error-tolerance dynamically modifies the accuracy of approximations, thereby reducing the accumulation of gradient errors. Empirical evaluations conducted on synthetic problems reveal its superiority compared to current methods, preserving speed without the need for hyperparameter adjustments. The versatility of this framework is advantageous for adversarial training and game theory, providing strong solutions for intricate saddle-point geometries without the necessity of predefined parameters. By integrating acceleration techniques with structural characteristics, it improves practical applicability in situations that demand equilibrium analysis or distributed decision-making, establishing it as a fundamental resource for future investigations in robust optimization.

Conclusion

is study advances minimax optimization with two key contributions. For strongly convex-concave problems, the DIAG algorithm merges AGD and Mirror-Prox, achieving near-optimal, $\tilde{O}(1/k^2)$ convergence—resolving acceleration-stability compatibility challenges. For nonconvex-concave settings, an inexact proximal point method attains $\tilde{O}(\frac{1}{k^{\frac{1}{3}}})$, exceeding previous rates through adaptive error-tolerance strategies that mitigate nonconvex instability. Innovations encompass hybrid acceleration-stabilization, dynamic approximation criteria, and empirical validation against benchmarks. Future directions entail broadening methods to nonsmooth objectives, determining nonconvex lower bounds, and incorporating variance reduction or distributed protocols for enhanced scalability. These hybrid frameworks integrate acceleration with structural insights, providing effective and dependable tools for machine learning and decision-making challenges.

References

[1] Jacob Abernethy, Kevin A. Lai, and Andre Wibisono. Last-iterate convergence rates for min-max optimization. arXiv preprint arXiv:1906.02027 , 2019.

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. arXiv preprint arXiv:1701.07875 , 2017.

[3] Waïss Azizian, Ioannis Mitliagkas, Simon Lacoste-Julien, and Gauthier Gidel. A tight and unified analysis of gradient-based methods for a whole spectrum of differentiable games. In International Conference on Artificial Intelligence and Statistics , pages 2863–2873, 2020.

[4] Zhong-Zhi Bai. Optimal parameters in the HSS-like methods for saddle-point problems. Numerical Linear Algebra with Applications , 16(6):447–479, 2009.

[5] Zhong-Zhi Bai, Gene H. Golub, and Michael K. Ng. Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems. SIAM Journal on Matrix Analysis and Applications , 24(3):603–626, 2003.

[6] Michele Benzi, Gene H. Golub, and Jörg Liesen. Numerical solution of saddle point problems. Acta Numerica , 14:1–137, 2005.

[7] Yair Carmon, Yujia Jin, Aaron Sidford, and Kevin Tian. Variance reduction for matrix games. In Advances in Neural Information Processing Systems , pages 11377–11388, 2019.

[8] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision , 40(1):120–145, 2011.

[9] Yunmei Chen, Guanghui Lan, and Yuyuan Ouyang. Optimal primal-dual methods for a class of saddle point problems. SIAM Journal on Optimization , 24(4):1779–1814, 2014.

[10] Bo Dai, Albert Shaw, Lihong Li, Lin Xiao, Niao He, Zhen Liu, Jianshu Chen, and Le Song. SBEED: Convergent reinforcement learning with nonlinear function approximation. In International Conference on Machine Learning , pages 1133–1142, 2018.

[11] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training GANs with optimism. In International Conference on Learning Representations (ICLR 2018) , 2018.

[12] Simon S Du, Jianshu Chen, Lihong Li, Lin Xiao, and Dengyong Zhou. Stochastic variance reduction methods for policy evaluation. In International Conference on Machine Learning , pages 1049–1058, 2017.

[13] Simon S Du and Wei Hu. Linear convergence of the primal-dual gradient method for convex concave saddle point problems without strong convexity. In The 22nd International Conference on Artificial Intelligence and Statistics , pages 196–205, 2019.

[14] Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial networks. In International Conference on Learning Representations , 2019.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems , pages 2672–2680, 2014.

[16] Osman Güler. New proximal point algorithms for convex minimization. SIAM Journal on Optimization , 2(4):649–664, 1992.

[17] Joachim Hartung. An extension of Sion's minimax theorem with an application to a method for constrained games. Pacific Journal of Mathematics , 103(2):401–408, 1982.

[18] Magnus R. Hestenes and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. Journal of Research of the National Bureau of Standards , 49(6):409–436, 1952.

[19] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. On the convergence of single-call stochastic extra-gradient methods.

In Advances in Neural Information Processing Systems , pages 6938–6948, 2019.
[20] Adam Ibrahim, Waïss Azizian, Gauthier Gidel, and Ioannis Mitliagkas. Linear lower bounds and conditioning of differentiable games. arXiv preprint arXiv:1906.07300 , 2019.
[21] David Kinderlehrer and Guido Stampacchia. An Introduction to Variational Inequalities and Their Applications , volume 31. SIAM, 1980.
[22] G. M. Korpelevich. The extragradient method for finding saddle points and other problems. Matecon , 12:747–756, 1976.
[23] Tianyi Lin, Chi Jin, and Michael I Jordan. On gradient descent ascent for nonconvex-concave minimax problems. arXiv preprint arXiv:1906.00331 , 2019.
[24] Tianyi Lin, Chi Jin, and Michael I. Jordan. Near-optimal algorithms for minimax optimization. In Conference on Learning Theory, COLT 2020 , 9–12 July 2020, Virtual Event [Graz, Austria], volume 125 of Proceedings of Machine Learning Research , pages 2738–2779. PMLR, 2020.