ISSN: 2410-7727

A Survey of Primal–Dual Algorithms: From Convex–Concave Saddle Points to Nonconvex–Nonconcave Problems

مسح خوارزميات الأولي-المزدوج:
من نقاط السرج المحدبة-المقعرة إلى المسائل غير المحدبة وغير المقعرة

Iyad WALWIL*

**Department of Computer Science Apprenticeship*
Faculty of Information Technology and Artificial Intelligence
An-Najah National University — Nablus, Palestine
iyad.walwil@najah.edu

ISSN: 2410-7727

2003 ١٤٢٤

جامعة الأندلس
للعلوم والتقنية
Alandalus University For Science & Technology

ISSN: 2410-7727

A Survey of Primal–Dual Algorithms: From Convex–Concave Saddle Points to Nonconvex–Nonconcave Problems

Iyad WALWIL*

**Department of Computer Science Apprenticeship*
Faculty of Information Technology and Artificial Intelligence
An-Najah National University — Nablus, Palestine
iyad.walwil@najah.edu

ABSTRACT

Primal–dual algorithms have become a cornerstone of modern optimization, offering a unified framework for tackling structured problems that involve both primal variables and dual multipliers simultaneously. Their broad applicability stems from the observation that a wide variety of optimization models can be cast as saddle point problems, in which the interplay between primal feasibility and dual optimality is made explicit.

This survey provides a structured review of primal–dual algorithms across four progressively general problem classes: *(i)* convex–concave saddle point problems, *(ii)* nonconvex problems with affine equality constraints, *(iii)* nonconvex–concave saddle point problems, and *(iv)* the most general class of nonconvex–nonconcave problems. For each setting, we trace the historical development of key algorithms, discuss the underlying theoretical insights that enable convergence, and highlight the connections between methods.

Our exposition begins with the classical Arrow–Hurwicz–Uzawa method of the 1950s and proceeds through the proximal point method, forward–backward splitting, extra-gradient variants, ADMM, and PDHG, before moving into modern nonconvex territory. In the nonconvex regime, we survey augmented Lagrangian methods, nonconvex ADMM variants, gradient descent–ascent analysis, and the recent weak Minty variational inequality framework that underlies state-of-the-art algorithms for nonconvex–nonconcave saddle points. The goal of this survey is to offer both a historical perspective and a practical guide to researchers and practitioners working at the intersection of optimization, machine learning, and signal processing.

Keywords: primal–dual algorithms, saddle point problems, convex–concave optimization, nonconvex optimization, ADMM, PDHG, extra-gradient, variational inequalities, min–max optimization, monotone operators.

ISSN: 2410-7727

مسح خوارزميات الأولي-المزدوج

من نقاط السرج المحدبة-المقعرة إلى المسائل غير المحدبة وغير المقعرة

إياد ولويل*

*قسم علوم الحاسوب في سوق العمل
كلية تكنولوجيا المعلومات والذكاء الاصطناعي
جامعة النجاح الوطنية - نابلس، فلسطين

ملخص البحث

أصبحت خوارزميات الأولي-المزدوج ركيزةً أساسيةً في التحسين الحديث، إذ تُقدّم إطاراً موحداً لمعالجة المسائل المنظّمة التي تتضمّن في آنٍ واحد متغيرات أولية ومضاعفات مزدوجة. يعود نطاق تطبيقها الواسع إلى الملاحظة القائلة بأن طيفاً عريضاً من نماذج التحسين يمكن صياغته على شكل مسائل نقطة السرج، حيث يتجلّى التفاعل بين الجدوى الأولية والمثالية المزدوجة بصورة صريحة.

يُقدّم هذا المسح مراجعةً منهجيةً لخوارزميات الأولي-المزدوج عبر أربع فئات من المسائل تتصاعد تدريجياً في عموميّتها: (أ) مسائل نقطة السرج المحدبة-المقعرة، (ب) المسائل غير المحدبة ذات القيود الاشتراطية التآلفية، (ج) مسائل نقطة السرج غير المحدبة-المقعرة، و(د) الفئة الأعمّ والأكثر تحدياً وهي المسائل غير المحدبة وغير المقعرة. ولكل فئة، نتتبع التطور التاريخي للخوارزميات الرئيسية، ونناقش الرؤى النظرية الكامنة التي تُمكّن التقارب، ونسلّط الضوء على أوجه الترابط بين المنهجيات المختلفة.

يبدأ عرضنا بطريقة Arrow-Hurwicz-Uzawa الكلاسيكية من خمسينيات القرن الماضي، ثم ينتقل عبر طريقة النقطة القريبة، وتشعيب الأمام-الخلف، والمتغيرات ذات الأساس الإضافي، وطريقة ADMM، وطريقة PDHG، قبل الانتقال إلى مسائل التحسين الحديثة غير المحدبة. وفي النطاق غير المحدب، نستعرض طرق لاغرانج المعزّزة، ومتغيرات ADMM غير المحدبة، وتحليل نزول-صعود التدرج، وإطار متباينات Minty الضعيفة الحديثة الذي يُشكّل الأساس لخوارزميات متطورة كـEG+ وBC-SEG+ لمعالجة مسائل نقطة السرج غير المحدبة-غير المقعرة. يهدف هذا المسح إلى تقديم منظور تاريخي ودليل عملي للباحثين والمختصين العاملين في تقاطع التحسين والتعلم الآلي ومعالجة الإشارات.

الكلمات المفتاحية: خوارزميات الأولي-المزدوج، مسائل نقطة السرج، التحسين المحدب-المقعر، التحسين غير المحدب، ADMM، PDHG، التدرج الإضافي، متباينات التغيّر، تحسين الحد الأدنى-الأقصى، العوامل الرتيبة.

ISSN: 2410-7727

Introduction

Optimization problems arising in modern data science, machine learning, signal processing, and game theory are increasingly formulated as *saddle point problems (SPPs)*, also known as min–max problems. These are problems of the form

$$min_{x \in X}\ max_{y \in Y}\ L(x, y), \qquad (1)$$

where the goal is to find a pair (x*, y*) at which neither player can improve by a unilateral deviation. Such formulations arise naturally in constrained optimization (where y plays the role of a Lagrange multiplier), in robust optimization (where y indexes worst-case scenarios), and in two-player zero-sum games (where x and y are the strategies of the two players).

Primal–dual algorithms are iterative methods that simultaneously update both the primal variable x and the dual variable y, exploiting the saddle-point structure to drive the iterates toward an equilibrium. Their appeal lies in their modularity: by operating on both variables at once, they avoid the need for expensive nested optimization loops and can often be decomposed into simple, efficient steps.

Over the past seven decades, primal–dual methods have evolved dramatically. Starting from the simple gradient-based schemes of the 1950s, the field has developed rich algorithmic families — proximal point methods, splitting methods, extra-gradient methods, augmented Lagrangian methods, and coordinate descent variants — each tailored to a specific problem structure, each contributing a new idea that has influenced subsequent work. More recently, the surge of interest in machine learning applications, especially adversarial training and generative adversarial networks (GANs), has pushed the frontier toward nonconvex and nonconvex–nonconcave settings, where classical theory no longer applies and new analytical tools are required.

Scope and Organization

This survey is organized around four progressively general problem classes:

§1. Convex–Concave Saddle Point Problems (Section 2): the classical setting where L is convex in x and concave in y. We trace the development from Arrow–Hurwicz–Uzawa through proximal point, forward–backward splitting, extra-gradient, ADMM, PDHG, and coordinate descent variants.

§2. Nonconvex Problems with Affine Constraints (Section 3): the objective f is nonconvex but the feasible set is described by linear equality constraints Ax = b. We cover augmented Lagrangian methods, SQP, interior-point methods, and the rigorous nonconvex analysis of ADMM.

§3. Nonconvex–Concave Saddle Point Problems (Section 4): L is nonconvex in x but concave in y. We survey proximal-guided methods, GDA analyses, and multi-step gradient algorithms.

ISSN: 2410-7727

§4. Nonconvex–Nonconcave Saddle Point Problems (Section 5): the most general and challenging setting. We present the variational inequality perspective, local minimax concepts, and the weak Minty variational inequality framework underlying modern algorithms such as EG+ and BC-SEG+.

Throughout, we highlight the key ideas and connections across methods and provide historical context. We do not aim for encyclopedic coverage of all primal–dual algorithms; rather, we focus on the main conceptual lineages that have shaped the field.

2 Convex–Concave Saddle Point Problems

The development of primal–dual methods for convex–concave saddle point problems has been a long and rich journey. The story starts in the 1950s with the Arrow–Hurwicz–Uzawa method [1, 2], one of the first iterative algorithms designed to solve saddle point problems of the form:

$$min_{x \in X}\, max_{y \in Y}\, L(x, y), \qquad (CC - SSP)$$

where L is convex in x ∈ X and concave in y ∈ Y.

Their idea was simple: take a step downhill in the primal variable, a step uphill in the dual variable, and hope to meet at the saddle. Although conceptually simple, its convergence can be slow or even fail without strong convexity–concavity assumptions or appropriate step-size rules — a reality that quickly called for more robust techniques.

2.1 Proximal and Splitting Methods

In the 1970s, key advancements emerged along two closely related lines: implicit (proximal) updates and explicit–implicit operator splitting.

The Proximal Point Method (PPM), proposed by Martinet [3, 4] and generalized by Rockafellar [5, 6] in 1976, reformulates (CC-SPP) as finding z ∈ Z satisfying

$$0 \in T(z) \qquad (MIP)$$

for a maximal monotone mapping T, and iterates

$$z_{k+1} = (I + \alpha_k T)^{-1}(z_k). \qquad (PPM)$$

The quadratic proximity regularizer makes each saddle-point subproblem strongly convex–concave, trading computational cost per step for strong convergence guarantees.

ISSN: 2410-7727

Building directly on this operator-splitting viewpoint, the Forward–Backward Splitting (FBS) method of Lions and Mercier [7] and Passty [8] (1979) handles the decomposition T = A + B by applying A explicitly (a cheap gradient step) and B implicitly (a proximal step):

$$z_{k+1} = (I + \alpha_k B)^{-1}(I - \alpha_k A)(z_k). \qquad (FBM)$$

This forward–backward structure reduces computational burden relative to PPM while retaining its convergence properties, and it forms the template for PDHG and many later algorithms.

2.2 Extra-Gradient Methods

Around the same time, a complementary breakthrough came with Korpelevich's Extra-Gradient (EG) method [9]. Recognizing that a plain primal–dual gradient step can overshoot and cycle, EG introduces a predictor step to an intermediate point and a corrector step using the gradient evaluated there. This two-step structure avoids cycling and instability even when the operator is only monotone (not strongly so), making EG a robust foundation for subsequent algorithms.

Decades later, Paul Tseng's Forward–Backward–Forward (FBF) algorithm [10] (2000) unified both lines of work. FBF solves (MIP) for T = A + B under the weaker assumption that B is merely monotone and Lipschitz (not necessarily maximal monotone), adding a second explicit step with B at the new point to restore stability without solving harder subproblems.

Remark 2.1. FBF is the natural extra-gradient enhancement of FBS: it reduces to FBS if the second forward step is omitted, and to EG if A is the normal cone of a convex set (so the backward step becomes a projection).

2.3 ADMM and PDHG

In parallel to PPM and EG, the Alternating Direction Method of Multipliers (ADMM) [11, 12, 13] emerged in the mid-1970s from the work of Glowinski, Marroco, Gabay, and Mercier. ADMM combines dual decomposition with augmented Lagrangian techniques to handle optimization problems with separable objectives but coupled constraints, alternating between minimizing each variable block and updating Lagrange multipliers. Its practicality and flexibility have made it a cornerstone of machine learning, signal processing, and statistics.

Much later, in 2011, Chambolle and Pock introduced the Primal–Dual Hybrid Gradient (PDHG) [14] algorithm for saddle point problems of the form:

$$min_{x \in X} max_{y \in Y}\ f(x) + \langle Ax, y\rangle - g*(y)$$

where f and g* are convex and A is a linear operator. PDHG alternates a forward–backward step on the dual variable with a forward–backward step on the

ISSN: 2410-7727

primal, followed by an over-relaxed primal update to accelerate convergence. Its step-size rules ensure convergence even when the operator norm is large, making it highly practical for imaging, inverse problems, and machine learning.

Algorithm 1: Primal–Dual Hybrid Gradient (PDHG)

Input: $\gamma_x, \gamma_Y > 0, \quad \alpha \in [0, 1]$

Initialize: (x_0, y_0); set $\bar{x}_0 = x_0$

while stopping criterion not met do

$y_{k+1} = prox_{(\gamma_Y,\ g^*)}(y_k + \gamma_Y A \bar{x}_k)$

$x_{k+1} = prox_{(\gamma_x,\ f)}(x_k - \gamma_x A^T y_{k+1})$

$\bar{x}_{k+1} = x_{k+1} + \alpha(x_{k+1} - x_k)$

end while

return (x_{k+1}, y_{k+1})

2.4 Coordinate Descent Primal–Dual Methods

Once the core methods were well established, researchers began exploring coordinate-based primal–dual updates that modify only a subset of variables per iteration — an important efficiency gain at large scale.

Bianchi, Hachem, and Iutzeler introduced a Randomized Coordinate Descent Primal–Dual (RCDPD) algorithm [15] in 2014 that randomly selects coordinates to update, reducing per-step cost while maintaining convergence guarantees. In 2017 two complementary methods appeared: the Stochastic Primal–Dual Hybrid Gradient (SPDHG) [16], combining stochastic coordinate updates with primal–dual splitting, and the Stochastic Primal–Dual Coordinate (SPDC) method [17] of Zhang and Xiao, tailored to regularized empirical risk minimization. Fercoq and Bianchi's CDPD [18] (2019) extended coordinate descent to large step sizes and non-separable functions, while Latafat, Freris, and Patrinos [19] developed a block-coordinate triangularly preconditioned variant for distributed optimization. Finally, Alacaoglu, Fercoq, and Cevher's PURE-CD [20] (2022) adaptively applies large steps to dense data and sparse updates to coordinate subsets, unifying both regimes.

ISSN: 2410-7727

3 Nonconvex Optimization with Affine Equality Constraints

Many problems of practical interest are inherently nonconvex. One of the simplest such settings arises when the objective is nonconvex but the feasible region is structured by affine equality constraints:

$$min_{x \in X} \ f(x) \quad subject\ to \quad Ax = b, \quad (2)$$

where f is possibly nonconvex. The associated Lagrangian L(x, y) = f(x) + ⟨Ax − b, y⟩ fits naturally into the saddle-point framework. The dual variable enters linearly, so the main difficulty is the nonconvexity of f.

3.1 Classical Approaches: ALM and Second-Order Methods

The first major progress came in the late 1960s and early 1970s with Augmented Lagrangian Methods (ALMs), independently introduced by Hestenes [21] and Powell [22] and later formalized by Rockafellar [23]. ALM augments the classical Lagrangian with a quadratic penalty:

$$L_\rho(x,y) = f(x) + \langle Ax - b, y\rangle + (\rho/2)\|Ax - b\|^2, \quad (3)$$

stabilizing dual updates and enforcing feasibility more effectively than pure penalty methods. For nonconvex f, ALM typically converges to KKT (first-order stationary) points.

From the late 1970s onward, equality-constrained nonconvex problems were studied within the broader framework of nonlinear programming (NLP), with Sequential Quadratic Programming (SQP) [24, 25] and Interior-Point Methods (IPM) [26, 27] as dominant paradigms. Both methods offer strong convergence guarantees but require second-order (Hessian) information, making them computationally heavy and poorly suited to large-scale machine learning.

3.2 Nonconvex ADMM

The computational limitations of second-order methods motivated renewed interest in first-order primal–dual algorithms for nonconvex problems, centering on a rigorous analysis of ADMM in nonconvex settings.

Wang, Xu, and Xu introduced Bregman ADMM (BADMM) [28] in 2014, incorporating Bregman distances into the ADMM updates to extend convergence guarantees to nonconvex composite problems. Wang, Cao, and Xu then developed multi-block Bregman ADMM (MB-BADMM) [29] (2015), demonstrating convergence for a wide class of nonconvex functions with three or more variable blocks.

Hong, Luo, and Razaviyayn [30] (2016) gave one of the earliest robust convergence results for the classical ADMM in nonconvex settings, showing

ISSN: 2410-7727

convergence to stationary solutions when the penalty parameter is sufficiently large. Wang, Yin, and Zeng [31] extended these results to nonsmooth objectives with linear constraints, covering ℓ_p quasi-norms and piecewise linear penalties under fairly general assumptions.

Finally, Hong, Lee, and Razaviyayn [32] (2018) studied the Gradient Primal–Dual Algorithm (GPDA) and Gradient ADMM (GADMM), showing that with random initialization both algorithms compute second-order stationary solutions with probability one — the first result of its kind for a primal–dual algorithm using only first-order information.

4 Nonconvex–Concave Saddle Point Problems

Nonconvex–concave SPPs take the form:

$$min_{x \in X} max_{y \in Y} \; L(x, y), \quad (NC - SPP)$$

where L is possibly nonconvex in x but concave in y. This framework arises in robust learning, adversarial training, and GANs. Unlike the affine equality case, the inner maximization may be nonlinear, leading to significant technical and algorithmic difficulties.

4.1 Proximal and Weakly Convex Methods

A first systematic step was made by Rafique, Li, Lin, and Yang [33] (2018), who studied weakly-convex–concave problems where L(·, y) is weakly convex for any y. They proposed Proximally guided Stochastic sub-Gradient (PSsG) and variance-reduced methods that avoid exactly solving the inner maximization at every step, converging provably to stationary points with explicit complexity bounds. This work emphasized proximal regularization as a key tool and established a broadly applicable framework for distributionally robust learning.

4.2 Gradient Descent–Ascent and Multi-Step Variants

A rich line of work from 2019 to 2021 developed and analyzed gradient-based methods for the nonconvex–concave setting.

Motivated by GAN training dynamics, Lu et al. [34] (2019) connected min–max optimization with linear discriminator coupling to first-order algorithms, establishing sub-linear convergence to stationary solutions under mild conditions. Concurrently, Nouiehed et al. [35] proposed Multi-Step Gradient Descent–Ascent (MS-GDA), the first algorithm providing non-asymptotic convergence rates for nonconvex–concave problems without special structural assumptions. By applying Nesterov's smoothing [36] to the inner maximization, MS-GDA runs multiple steps of accelerated gradient ascent followed by gradient descent (or Frank–Wolfe) for the outer minimization. Kong et al. [37] further developed smoothing-based ideas by

ISSN: 2410-7727

replacing the inner maximization with a smooth approximation and applying an accelerated inexact proximal point method.

In 2020, Lin, Jin, and Jordan [38] analyzed two-time-scale gradient descent–ascent (GDA) and its stochastic variant (SGDA), showing provable convergence to stationary points of the outer minimization in the (strongly) concave case. Building on this, Lu, Tsaknakis, and Hong's Hybrid Block Successive Approximation (HiBSA) [39] (2021) introduced a block-coordinate framework alternating gradient descent on the nonconvex variables with gradient ascent on the concave variable, achieving global convergence to first-order stationary points through careful tuning of regularization and penalty sequences.

5 Nonconvex–Nonconcave Saddle Point Problems

The study of nonconvex–nonconcave saddle point problems (SPPs) lies beyond the well-structured convex–concave theory. Mathematically:

$$min_{x \in X} max_{y \in Y} \ L(x, y), \quad (NN - SPP)$$

where L is possibly nonconvex in x and possibly nonconcave in y.

5.1 Variational Inequality Foundations

A natural entry point is the variational inequality (VI) literature. Dang and Lan [40] (2013) extended the classical extra-gradient method to generalized monotone or Minty variational inequalities (MVI), achieving the optimal $1/\sqrt{k}$ convergence rate via Bregman mappings and establishing a foundational link between primal–dual algorithms and generalized monotone operators.

Building on this framework, Daskalakis and Panageas [41] (2018) analyzed GDA and its optimistic variant (OGDA) through the lens of dynamical systems, showing that these dynamics almost surely avoid unstable critical points. Sanjabi, Razaviyayn, and Lee [42] (2018) imposed structure via the Polyak–Łojasiewicz (PL) condition on the maximization player, obtaining near-optimal convergence rates for a multi-step GDA algorithm. Liu et al. [43] then considered weakly-convex–weakly-concave objectives and, inspired by PPM, developed an algorithm approximating the original weakly-monotone VI with a sequence of strongly monotone ones, providing the first non-asymptotic guarantees for this class. Jin, Netrapalli, and Jordan [44] (2020) further proposed local minimax points as a principled surrogate for global equilibria, showing that stable limit points of GDA coincide with these local notions.

5.2 MVI-Based Methods and the Weak-MVI Framework

Advances in MVI-related methods offered complementary perspectives. Song et al. [45] proposed Optimistic Dual Extrapolation (OptDE), requiring only one gradient evaluation per iteration and achieving linear convergence under strong MVI assumptions.

ISSN: 2410-7727

In 2021, Diakonikolas, Daskalakis, and Jordan [46] formalized the weak-MVI structure for nonconvex–nonconcave problems and designed Extra-Gradient+ (EG+), a robust method with sharp complexity guarantees in both deterministic and stochastic settings. In parallel, Daskalakis, Skoulakis, and Zampetakis [47] revealed a fundamental computational limit: computing approximate local min–max equilibria in linearly constrained settings is PPAD-complete, even under mild approximation requirements.

Pethick et al. [48] extended EG+ to CurvatureEG+ (2022), handling constrained settings suffering from limit cycles via adaptive step sizes, and obtaining the first global convergence guarantees in the presence of cyclic dynamics. Most recently, Pethick et al. [49] developed the Bias-Corrected Stochastic Extra-gradient+ (BC-SEG+), which replaces the impractical growing-batch requirement with bias-corrected updates requiring only one additional oracle call per iteration, enabling efficient convergence under weak-MVI assumptions. Together, EG+, CurvatureEG+, and BC-SEG+ mark the maturity of the weak-MVI framework as a foundation for nonconvex–nonconcave algorithms.

6 Discussion and Connections

Looking across the four settings surveyed above, several unifying themes emerge.

Proximal regularization as a universal tool. From the proximal point method in the convex–concave setting to proximally guided methods in the nonconvex–concave setting, the idea of regularizing the saddle-point problem with a quadratic proximity term appears repeatedly. This regularization trades the difficulty of the original problem for a sequence of better-conditioned subproblems and provides a principled way to handle weak monotonicity and nonconvexity.

Extra-gradient as a structural backbone. The extra-gradient idea introduced by Korpelevich has proven remarkably durable. It reappears in FBF, PDHG (via the over-relaxation step), EG+, and BC-SEG+, each time adapted to a new setting. The two-step predictor–corrector structure seems to be a fundamental device for achieving stability in the presence of non-strongly monotone operators.

Splitting and decomposition. ADMM, FBS, PDHG, and their coordinate descent variants all exploit a decomposition of the problem into simpler subproblems. This idea is not merely computational: in the nonconvex setting, careful splitting (as in GADMM and GPDA) can be used to analyze second-order properties and escape saddle points.

The role of geometry. The move from Euclidean to non-Euclidean geometry, exemplified by the Bregman distances in BADMM and the mirror maps in Mirror-Prox, reflects a broader recognition that the geometry of the feasible set or the problem structure can be exploited to obtain better convergence rates or more practical algorithms.

ISSN: 2410-7727

From global to local optimality. In the nonconvex–nonconcave setting, the impossibility of finding global solutions in polynomial time has led to a refined vocabulary of local and approximate optimality: stationary points, local minimax points, weak-MVI solutions. The development of this vocabulary has been essential for making meaningful progress.

Table 1: Summary of major primal–dual algorithms surveyed.

Algorithm	Year	Setting	Key Idea
Convex–Concave (CC)			
Arrow–Hurwicz–Uzawa	1950s	CC	Gradient descent–ascent
PPM	1970s	CC	Implicit proximal update via resolvent
FBS	1979	CC	Explicit + implicit splitting
EG	1976	CC	Predictor–corrector two-step update
ADMM	1970s	CC/NC	Augmented Lagrangian with alternating minimization
FBF	2000	CC	EG enhancement of FBS
PDHG	2011	CC	Fully split forward–backward with over-relaxation
RCDPD / SPDHG / PURE-CD	2014–2022	CC	Coordinate descent primal–dual
Nonconvex with Affine Constraints (NC-Affine)			
ALM	1969–1973	NC-Affine	Quadratic penalty stabilization
SQP / IPM	1977–1999	NC-Affine	Second-order NLP methods
BADMM / MB-BADMM	2014–2015	NC-Affine	Bregman ADMM for nonconvex
GPDA / GADMM	2018	NC-Affine	First-order, second-order stationary pts
Nonconvex–Concave (NC-C)			

ISSN: 2410-7727

Algorithm	Year	Setting	Key Idea
PSsG	2018	NC-C	Proximal regularization, weakly convex
MS-GDA	2019	NC-C	Multi-step GDA with Nesterov smoothing
GDA / SGDA	2020	NC-C	Two-time-scale gradient descent–ascent
HiBSA	2021	NC-C	Block-coordinate successive approximation
Nonconvex–Nonconcave (NN)			
N-EG	2013	NN	Non-Euclidean EG for generalized MVI
OptDE	2020	NN	Optimistic dual extrapolation
EG+	2021	NN	Weak-MVI framework, sharp complexity
CurvatureEG+	2022	NN	Adaptive EG+ for constrained cyclic problems
BC-SEG+	2023	NN	Bias-corrected stochastic EG+

CC *= Convex–Concave;* ***NC-Affine*** *= Nonconvex with affine constraints;* ***NC-C*** *= Nonconvex–Concave;* ***NN*** *= Nonconvex–Nonconcave.*

ISSN: 2410-7727

7 Conclusion

This survey has traced the development of primal–dual algorithms from their classical roots in convex–concave optimization to the modern frontier of nonconvex–nonconcave saddle point problems. Several conclusions stand out.

First, the field has shown a remarkable capacity for progressive generalization: each new setting has called forth new algorithmic ideas — often built on, or inspired by, tools from the more structured setting — rather than requiring a complete restart. Extra-gradient, proximal regularization, and splitting are not just techniques for specific problems; they are general principles that reappear in new forms at each level of generality.

Second, the interplay between algorithmic design and theoretical analysis has been essential. The rigorous convergence analysis of nonconvex ADMM, the characterization of local minimax points, and the weak-MVI framework have not only validated existing algorithms but have shaped the design of new ones.

Third, the field is still very much in motion. The nonconvex–nonconcave setting remains theoretically challenging, and many practical questions — such as the efficient handling of stochasticity, the design of adaptive step sizes, and the behavior of these algorithms in deep learning — are active areas of research.

We hope this survey serves both as a reference for practitioners seeking to understand the landscape of primal–dual methods and as an entry point for researchers looking to contribute to this rich and fast-moving field.

ISSN: 2410-7727

References

[1] K. Arrow, L. Hurwicz, and H. Uzawa. Studies in linear and non-linear programming. Stanford University Press, 1958.

[2] K. Arrow, L. Hurwicz, and H. Uzawa. Gradient methods for concave programming. In Studies in Linear and Non-Linear Programming, 1958.

[3] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. Revue Française d'Informatique et de Recherche Opérationnelle, 4:154–158, 1970.

[4] B. Martinet. Détermination approchée d'un point fixe d'une application pseudo-contractante. Comptes Rendus de l'Académie des Sciences, 274:163–165, 1972.

[5] R. T. Rockafellar. Monotone operators and the proximal point algorithm. SIAM Journal on Control and Optimization, 14(5):877–898, 1976.

[6] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Mathematics of Operations Research, 1(2):97–116, 1976.

[7] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis, 16(6):964–979, 1979.

[8] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. Journal of Mathematical Analysis and Applications, 72(2):383–390, 1979.

[9] G. M. Korpelevich. The extragradient method for finding saddle points and other problems. Matecon, 12:747–756, 1977.

[10] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. SIAM Journal on Control and Optimization, 38(2):431–446, 2000.

[11] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. ESAIM: Mathematical Modelling and Numerical Analysis, 9(R2):41–76, 1975.

[12] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Computers & Mathematics with Applications, 2(1):17–40, 1976.

[13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning, 3(1):1–122, 2011.

ISSN: 2410-7727

[14] A. Chambolle and T. 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.

[15] P. Bianchi, W. Hachem, and F. Iutzeler. A coordinate descent primal-dual algorithm and application to distributed asynchronous optimization. IEEE Transactions on Automatic Control, 61(10):2947–2957, 2016.

[16] A. Chambolle, M. Ehrhardt, P. Richtárik, and C.-B. Schönlieb. Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications. SIAM Journal on Optimization, 28(4):2783–2808, 2018.

[17] Y. Zhang and L. Xiao. Stochastic primal-dual coordinate method for regularized empirical risk minimization. Journal of Machine Learning Research, 18(1):2939–2980, 2017.

[18] O. Fercoq and P. Bianchi. A coordinate-descent primal-dual algorithm with large step size and possibly non-separable functions. SIAM Journal on Optimization, 29(1):100–134, 2019.

[19] P. Latafat, N. M. Freris, and P. Patrinos. A new randomized block-coordinate primal-dual proximal algorithm for distributed optimization. IEEE Transactions on Automatic Control, 64(10):4050–4065, 2019.

[20] A. Alacaoglu, O. Fercoq, and V. Cevher. On the complexity of a practical primal-dual coordinate method. In Proceedings of AISTATS, 2022.

[21] M. R. Hestenes. Multiplier and gradient methods. Journal of Optimization Theory and Applications, 4(5):303–320, 1969.

[22] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In Optimization, pages 283–298. Academic Press, 1969.

[23] R. T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM Journal on Control, 12(2):268–285, 1974.

[24] S.-P. Han. A globally convergent method for nonlinear programming. Journal of Optimization Theory and Applications, 22(3):297–309, 1977.

[25] M. J. D. Powell. A fast algorithm for nonlinearly constrained optimization calculations. In Numerical Analysis, pages 144–157. Springer, 1978.

[26] A. V. Fiacco and G. P. McCormick. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, 1968.

[27] R. H. Byrd, M. E. Hribar, and J. Nocedal. An interior point algorithm for large-scale nonlinear programming. SIAM Journal on Optimization, 9(4):877–900, 1999.

[28] H. Wang, A. Banerjee, Z.-Q. Luo, and B. Xu. Online alternating direction method (longer version). In ICML, 2014.

ISSN: 2410-7727

[29] F. Wang, Z. Cao, and B. Xu. Convergence of multi-block Bregman ADMM for nonconvex composite problems. Science China Information Sciences, 61:1–12, 2018.

[30] M. Hong, Z.-Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. SIAM Journal on Optimization, 26(1):337–364, 2016.

[31] Y. Wang, W. Yin, and J. Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. Journal of Scientific Computing, 78(1):29–63, 2019.

[32] M. Hong, J.-D. Lee, and M. Razaviyayn. Gradient primal-dual algorithm converges to second-order stationary solution for nonconvex distributed optimization. In ICML, 2018.

[33] H. Rafique, M. Liu, Q. Lin, and T. Yang. Weakly-convex–concave min–max optimization: provable algorithms and applications in machine learning. Optimization Methods and Software, 37(3):1087–1121, 2022.

[34] S. Lu, I. Tsaknakis, M. Hong, and Y. Chen. Hybrid block successive approximation for one-sided non-convex min-max problems: algorithms and applications. IEEE Transactions on Signal Processing, 68:3676–3691, 2020.

[35] M. Nouiehed, M. Sanjabi, T. Huang, J.-D. Lee, and M. Razaviyayn. Solving a class of non-convex min-max games using iterative first order methods. In NeurIPS, 2019.

[36] Y. Nesterov. Smooth minimization of non-smooth functions. Mathematical Programming, 103(1):127–152, 2005.

[37] W. Kong and R. D. C. Monteiro. An accelerated inexact proximal point method for solving nonconvex-concave min-max problems. SIAM Journal on Optimization, 31(4):2558–2585, 2021.

[38] T. Lin, C. Jin, and M. Jordan. Gradient descent ascent for minimax optimization with non-convexity. In ICML, 2020.

[39] S. Lu, I. Tsaknakis, and M. Hong. Block alternating optimization for non-convex min-max problems. In ICASSP, 2020.

[40] C. D. Dang and G. Lan. On the convergence properties of non-Euclidean extragradient methods for variational inequalities with generalized monotone operators. Computational Optimization and Applications, 60(2):277–310, 2015.

[41] C. Daskalakis and I. Panageas. The limit points of (optimistic) gradient descent in min-max optimization. In NeurIPS, 2018.

[42] M. Sanjabi, J. Ba, M. Razaviyayn, and J.-D. Lee. On the convergence and robustness of training GANs with regularized optimal transport. In NeurIPS, 2018.

ISSN: 2410-7727

[43] M. Liu, H. Rafique, Q. Lin, and T. Yang. First-order convergence theory for weakly-convex-weakly-concave min-max problems. Journal of Machine Learning Research, 22(169):1–34, 2021.

[44] C. Jin, P. Netrapalli, and M. I. Jordan. What is local optimality in nonconvex-nonconcave minimax optimization? In ICML, 2020.

[45] Y. Song, O. Iyengar, and O. Shamir. Optimistic dual extrapolation for coherent non-monotone variational inequalities. In NeurIPS, 2020.

[46] J. Diakonikolas, C. Daskalakis, and M. Jordan. Efficient methods for structured nonconvex-nonconcave min-max optimization. In AISTATS, 2021.

[47] C. Daskalakis, S. Skoulakis, and M. Zampetakis. The complexity of constrained min-max optimization. In STOC, 2021.

[48] T. Pethick, P. Latafat, P. Patrinos, O. Fercoq, and V. Cevher. Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems. In ICLR, 2023.

[49] T. Pethick, G. Frangella, M. Udell, and V. Cevher. Bias-corrected stochastic extragradient+ for non-monotone variational inequalities. arXiv preprint arXiv:2310.01232, 2023.

ISSN: 2410-7727

مجلــة الأندلـس للعلـوم الإنسانيـة والاجتماعيـة
مجلــة دوليـــة شهـريـــة علميـــة محكمـــة
الترقيــم الدولي الإلكتروني: ISSN:2410- 521X
الترقيــم الدولــي الورقــي: ISSN:2410- 1818
البريد الإلكتروني : journal@andalusuniv.net

2003 ١٤٢٤
جامعة الأندلس للعلوم والتقنية
Alandalus University For Science & Technology

المجلــة مفهرســة فــي المواقــع الآتيــة :

Arcif Analytics
Google scholar
YAHOO!
دار المنظومة DAR ALMANDUMAH
Crossref doi
EBSCO
Kezana IMPACT THE NEXT
شمعة shamaa
معرفة e-Marefa
معامل التأثير العربي

العام	2021	2022	2023	2024	2025
معامل أرسيف	0.2692	0.1954	0.3759	0.3068	0.5978
معامل التأثير العربي	1.60	1.73	1.25	1.55	1.59