GRAND: A Gradient-Related Ascent and Descent Algorithmic Framework for Minimax Problems

Xiaochun Niu and Ermin Wei
Department of Industrial Engineering and Management Sciences
Northwestern University
Evanston, IL 60201

Abstract—In this work, we study the minimax optimization problems, which model many distributed and centralized optimization problems. Existing works mainly focus on the design and analysis of specific methods, such as gradient-type methods, including gradient descent ascent method (GDA) and its variants such as extra-gradient (EG) and optimistic gradient descent ascent (OGDA) methods, and Newton-type methods. In this work, we propose GRAND as a gradient-related ascent and descent algorithmic framework for finding global minimax points. It allows updates within acute angles to the partial gradient directions. GRAND covers and motivates gradient-type, Newton-type, and other general descent ascent methods as special cases. It also enables flexible methods’ designs for distributed consensus optimization problems to utilize heterogeneous agents. To the best of our knowledge, GRAND is the first generalized algorithmic framework for solving minimax problems with provable convergence guarantees.

I. INTRODUCTION

Given a convex-(possibly non) concave function $L : \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}$, we consider a minimax optimization problem over $L$ of the following form,

$$\max_{y \in \mathbb{R}^p} \min_{x \in \mathbb{R}^d} L(x, y).$$

We are interested in finding the global minimax points of Problem 1. Minimax optimization problem has attracted much attention recently. This is motivated by its wide application in areas including game theory [1], [2], robust control [3], constrained optimization [4], and machine learning, especially the training of Generative Adversarial Networks (GANs) [5].

While there is growing literature on developing and analyzing optimization algorithms to solve Problem 1, most existing works focus on gradient-type methods. Such works include analysis of the gradient descent ascent method (GDA) [6], [7], and variants of GDA such as extra-gradient (EG) and optimistic gradient descent ascent (OGDA) methods [8]. However, gradient-type methods only utilize first-order information and lead to relatively slow convergence performance. On the contrary, [9] considers Newton-type descent ascent methods but only analyzes local convergence rate. Thus, in this work, we propose GRAND as a generalized gradient-related ascent and descent method for finding global minimax points. Specifically, GRAND allows update directions within uniformly bounded acute angles to the partial derivatives of $L$. GRAND covers GDA as a special case. It also motivates the design of new methods, including scaled-gradient, Newton-type, and quasi-Newton-type methods. Moreover, GRAND motivates more flexible methods’ design in distribution settings to utilize heterogeneous agents. In particular, DISH introduced in [10] can be viewed as a special case under the GRAND framework. We conduct numerical experiments to demonstrate the efficacy of GRAND. To the best of our knowledge, GRAND is the first generalized algorithmic framework for solving minimax problems with provable convergence guarantees.

II. RELATED WORKS

Our work is related to the proliferating literature on descent ascent optimization methods to solve Problem 1. Such works including linear convergence of the gradient descent ascent method (GDA) for solving strongly-convex-strongly-concave problems [11] and strongly-convex-concave problems [6], and sublinear convergence of GDA for solving strongly-convex-nonconcave and convex-nonconcave problems [7]. There are also analysis of the variants of GDA, including extra-gradient (EG) and optimistic gradient descent ascent (OGDA) methods. In particular, [8] views both of these methods as approximations of the classical proximal point method and proves a linear convergence rate of them in the strongly-convex-strongly-concave setting. [9] proposes a Newton-type method and shows a local superlinear convergence rate. However, none of them considers scaled gradient or more generalized methods. Thus, we propose GRAND as a generalized algorithmic framework and provide convergence analysis on different settings.

III. PRELIMINARIES

In this section, we introduce assumptions and definitions used throughout the paper. We first assume the following standard conditions on the objective function $L(x, y)$.

Assumption 1. The function $L$ satisfies the conditions that:

(a) The function $L(x, y)$ is twice differentiable in $(x, y)$;
(b) For any $y \in \mathbb{R}^p$, the function $L(x, y)$ is $m_x$-strongly convex with respect to $x$ with $m_x > 0$;
(c) The partial derivative $\nabla_x L(x, y)$ is $\ell_{xx}$- and $\ell_{xy}$-Lipschitz continuous with respect to $x$ and $y$, respectively. Moreover, the partial derivative $\nabla_y L(x, y)$ is
\[ \ell_{\alpha\gamma} \text{ and } \ell_{\alpha\gamma} \text{-Lipschitz continuous with respect to } x \text{ and } y, \text{ respectively. Here constants } \ell_{\alpha\gamma} > m_\alpha > 0, \text{ and } \ell_{\alpha\gamma}, \ell_{\gamma\gamma} \geq 0. \]

Assumption 1 implies that the function \( L(., y) \) is \( m_\alpha \)-strongly convex with a unique minimizer for any \( y \in \mathbb{R}^p \).

For convenience, we define the unique minimizer as \( x^*(y) \) for any \( y \in \mathbb{R}^p \), that is,

\[
x^*(y) = \arg \min_{x \in \mathbb{R}^d} L(x, y).
\]

Now define the objective function \( \psi : \mathbb{R}^p \rightarrow \mathbb{R}, \)

\[
\psi(y) = \min_x L(x, y) = L(x^*(y), y).
\]

The function \( \psi(y) \) is concave since it is the pointwise minimum of the convex function \( L(., y) \). Now we introduce the notion of global minimax point that characterizes function optimality.

**Definition 2.** A point \((x^1, y^1)\) is a global minimax point if

\[
L(x^*(y), y) \leq L(x, y) \leq L(x^1, y^1) \quad \text{holds for any } (x, y).
\]

In Section IV, we propose GRAND as a generalized algorithmic framework for finding global minimax points of Problem 1.

**IV. ALGORITHM**

To solve Problem 1 iteratively, we propose the GRAND algorithmic framework in Algorithm 1.

Algorithm 1 presents the GRAND algorithmic framework for solving Problem 1, which generalizes GDA.

**Algorithm 1** GRAND: A Gradient Related Ascent and Descent Algorithmic Framework.

Input: \( \alpha > 0, \beta > 0, x^0 \in \mathbb{R}^d, \text{ and } y^0 \in \mathbb{R}^p. \)

for \( k = 0, \cdots, K - 1 \) do

\[
x^{k+1} = x^k - \alpha s^k,
\]

\[
y^{k+1} = y^k + \beta t^k.
\]

end for

In Algorithm 1, positive constants \( \alpha \) and \( \beta \) are stepsizes, and vectors \( s^k \in \mathbb{R}^d \) and \( t^k \in \mathbb{R}^p \) are \( x \)-descent and \( y \)-ascent update directions, respectively. We assume that the update directions \( s^k \) and \( t^k \) satisfy the following conditions.

**Assumption 3** (Descent and Ascent Directions in GRAND). There are constants \( \gamma_s, \gamma_t, \Gamma_s, \Gamma_t > 0 \), such that for any iteration \( k \), the directions \( s^k \) and \( t^k \) in Algorithm 1 satisfy

\[
\|s^k\| \geq \sqrt{\gamma_s} \|\nabla_x L(x^k, y^k)\|,
\]

\[
(s^k)^T \nabla_x L(x^k, y^k) \geq \frac{1}{\Gamma_s} \|s^k\|^2,
\]

\[
\|t^k\| \geq \sqrt{\gamma_t} \|\nabla_y L(x^k, y^k)\|,
\]

\[
(t^k)^T \nabla_y L(x^k, y^k) \geq \frac{1}{\Gamma_t} \|t^k\|^2.
\]

Assumption 3 is inspired by the conditions of gradient related descent algorithms for solving minimization problems [12]. For the \( x \)-update \( s^k \), the first condition implies that \( s^k \neq 0 \) and thus \( x^{k+1} \neq x^k \) whenever \( \nabla_x L(x^k, y^k) \neq 0 \), and the second condition guarantees that \( -s^k \) is a descent direction with an acute angle to \( \nabla_x L(x^k, y^k) \). Similarly, for the \( y \)-update \( t^k \), the first condition implies that \( t^k \neq 0 \) and thus \( y^{k+1} \neq y^k \) whenever \( \nabla_y L(x^k, y^k) \neq 0 \), and the second condition guarantees that \( t^k \) is an ascent direction with an acute angle to \( \nabla_y L(x^k, y^k) \).

**Claim 4.** In strongly-convex-strongly-concave settings, GRAND converges linearly to a global minimax point.

**V. CONCLUSIONS**

In this work, we propose GRAND as a gradient-related ascent and descent algorithmic framework for solving minimax problems. Many general descent algorithms such as gradient-type, scaled-gradient-type, and Newton-type methods are special cases of GRAND. To the best of our knowledge, GRAND is the first generalized framework for minimax problems with provable convergence and rate guarantees.

**REFERENCES**

[1] T. Başar and G. J. Olsder, Dynamic noncooperative game theory. SIAM, 1998.

[2] J. Von Neumann and O. Morgenstern, “Theory of games and economic behavior,” in Theory of games and economic behavior. Princeton university press, 2007.

[3] M. Hast, K. J. Auström, B. Bernhardsson, and S. Boyd, “Pd design by convex-concave optimization,” in 2013 European Control Conference (ECC). IEEE, 2013, pp. 4460–4465.

[4] D. P. Bertsekas, Constrained optimization and Lagrange multiplier methods. Academic press, 2014.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” Advances in neural information processing systems, vol. 27, 2014.

[6] S. S. Du and W. 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. PMLR, 2019, pp. 196–205.

[7] T. Lin, C. Jin, and M. Jordan, “On gradient descent ascent for nonconvex-concave minimax problems,” in International Conference on Machine Learning. PMLR, 2020, pp. 6083–6093.

[8] A. Mokhtari, A. Ozdaglar, and S. Pattathil, “A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach,” in International Conference on Artificial Intelligence and Statistics. PMLR, 2020, pp. 1497–1507.

[9] G. Zhang, K. Wu, P. Poupart, and Y. Yu, “Newton-type methods for minimax optimization,” arXiv preprint arXiv:2006.14592, 2020.

[10] X. Niu and E. Wei, “Dish: A distributed hybrid primal-dual optimization framework to utilize system heterogeneity,” arXiv preprint arXiv:2206.03624, 2022.

[11] T. Liang and J. Stokes, “Interaction matters: A note on non-asymptotic local convergence of generative adversarial networks,” in The 22nd International Conference on Artificial Intelligence and Statistics. PMLR, 2019, pp. 907–915.

[12] D. P. Bertsekas and J. N. Tsitsiklis, Parallel and distributed computation: numerical methods. Prentice hall Englewood Cliffs, NJ, 1989, vol. 23.