

一种求解凸-PL极小极大优化问题的随机交替梯度投影算法

覃媛媛, 王福胜, 王 静

太原师范学院数学与统计学院, 山西 晋中

收稿日期: 2023年1月11日; 录用日期: 2023年2月15日; 发布日期: 2023年2月21日

摘 要

针对机器学习中常见的一类非凸极小极大优化问题, 本文提出了一种新的随机交替梯度投影算法, 该算法基于交替梯度投影算法, 在随机环境中用随机梯度进行更新, 和随机交替梯度下降-上升算法相比, 运算效率大大提高。在高斯分布数据集上的数值实验结果表明新算法是可行有效的。

关键词

机器学习, 凸-PL, GDA算法, 随机梯度

A Stochastic Alternating Gradient Projection Algorithm for Convex-PL Minimax Optimization Problems

Yuanyuan Qin, Fusheng Wang, Jing Wang

School of Mathematics and Statistics, Taiyuan Normal University, Jinzhong Shanxi

Received: Jan. 11th, 2023; accepted: Feb. 15th, 2023; published: Feb. 21st, 2023

Abstract

To solve a class of nonconvex minimax optimization problems in machine learning, we propose a new stochastic alternating gradient projection algorithm (SAGP). This algorithm is based on the alternating gradient projection algorithm (AGP), which is updated with stochastic gradients in a stochastic environment. Compared with the stochastic alternating gradient descent ascent algorithm (SAGDA), the operation efficiency is greatly improved. Numerical experiments on Gaussian distributed data sets show that the new algorithm is feasible and effective.

Keywords

Machine Learning, Convex-PL, GDA Algorithm, Stochastic Gradient

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

本文考虑以下极小极大优化问题:

$$\min_{x \in R^n} \max_{y \in R^m} f(x, y) \triangleq E[F(x, y; \xi)],$$

其中 ξ 是具有支持度 Ξ 的随机向量, $f(x, y)$ 是光滑函数, $f(x, y)$ 对于 x 是凸的, 对于 y 是非凹的。

在机器学习中, 许多问题都可表述为 $\min \max f(x, y)$ 的形式, 如生成性对抗网络(GAN)、分布式非凸优化、多域鲁棒学习、信号处理中的功率控制和收发器设计问题等。求解此类问题最简单的方法之一是梯度下降(GD)的自然泛化, 称为梯度下降 - 上升(GDA)算法, 在每次迭代时通过同步或交替的渐变更新来对 x 执行梯度下降步骤, 对 y 执行梯度上升步骤。

凸 - 凹极小极大优化问题对应的变分不等式是单调算子, 目前已有较好的成果[1] [2] [3] [4] [5]。非凸极小极大优化问题, 包括非凸 - 凹、凸 - 非凹、非凸 - 非凹极小极大优化问题, 由于其非凸性、非凹性和非光滑性的特殊结构, 一般是 NP-难的, 目前已有一些理论成果。在非凸 - 凹环境下, 求解此类问题的算法有两类, 分别是嵌套循环算法[6] [7] [8] [9] [10]和单循环算法。单循环算法多基于 GDA 算法改进, 如 two-time-scale GDA and stochastic GDA 算法[11], GDmax 算法[12], 混合块逐次逼近算法(HiBSA) [13], 交替梯度投影算法(AGP) [14], proximal alternating GDA 算法[15], Smoothed-GDA 算法[16], Stoc-AGDA 和 Stoc-Smoothed-AGDA [17]和其他改进的 GDA 算法[18] [19]。

在凸 - 非凹环境下, 对于任意给定的 x , 求解 $\max_y f(x, y)$ 是 NP-难的。几乎现有的嵌套循环算法和部分现有的单循环算法[12] [13]都会失去理论保证, 因为这些算法都需求解 $\max_y f(x, y)$ 。目前只有 AGP 算法[14]可求解目标函数 $f(x, y)$ 是光滑函数的凸 - 非凹极小极大优化问题, 并证明了其收敛性。

凸 - 非凹极小极大优化问题的一类子问题是凸-PL 极小极大优化问题, 即 $f(x, y)$ 对于 x 是凸的, 对于 y 是满足 Polyak-Lojasiewicz(PL)条件的。PL 条件最初是由 Polyak [20]引入, 并证明了梯度下降以线性速率全局收敛。目标函数的变量 y 满足 PL 条件的假设比 y 具有强凹性的假设更温和, 甚至不要求目标函数在 y 上是凹的, 这种假设在线性二次型调节器[21]和超参数化神经网络[22]中被证明成立。对于极小极大优化问题, 有的研究[17]是使变量 y 满足 PL 条件, 有的研究[23] [24]则是通过使 x 满足 PL 条件来找到全局解, 同时 PL 条件被证明适用于机器学习中某些非凸应用[21] [25], 包括与神经网络相关的问题[22] [26], 因此 PL 条件引起了广泛的关注。

非凸 - 凹环境下, 许多算法[11] [27]结合随机梯度都取得了较好的收敛结果。受启发于 Stoc-Smoothed-AGDA 算法[17], 本文考虑在凸-PL 环境下, 将随机梯度与 AGP 算法[14]相结合, 从而提出了随机交替梯度投影算法(SAGP)。

2. 交替梯度投影算法

本节首先回顾交替梯度投影算法(AGP) [14]。考虑以下极小极大优化问题:

$$\min_{x \in X} \max_{y \in Y} f(x, y),$$

其中 $f: X \times Y \rightarrow \mathbb{R}$ 是光滑函数, $X \subseteq \mathbb{R}^n, Y \subseteq \mathbb{R}^m$ 为非空闭凸集。

交替梯度投影算法是单循环算法, 每次迭代通过两个梯度投影步骤更新 x 和 y 。在第 t 次迭代中, AGP 算法使用如下辅助函数的梯度进行更新, 即:

$$\hat{f}(x, y) = f(x, y) + \frac{b_t}{2} \|x\|^2 - \frac{c_t}{2} \|y\|^2,$$

其中 $b_t \geq 0$ 和 $c_t \geq 0$ 是正则化参数。在凸 - 非凹环境下, $c_t = 0$, $\{b_t\}$ 是非负单调递减序列, 以下给出交替梯度投影算法的框架:

算法 1 交替梯度投影算法(AGP)

步骤 1 输入固定步长 $\tau_1 > 0$, $\tau_2 > 0$, 初始点 (x_0, y_0) 及参数 $b_0 > 0$,

步骤 2 计算 b_t 并更新 x_t :

$$x_{t+1} = P_X(x_t - \tau_1(\nabla_x f(x_t, y_t) + b_t x_t)),$$

步骤 3 更新 y_t :

$$y_{t+1} = P_Y(y_t + \tau_2 \nabla_y f(x_{t+1}, y_t)),$$

步骤 4 当算法满足终止准则时, 停止; 否则, 令 $t = t + 1$, 返回步骤 2。

3. PL 条件

目标函数 $f(x, y)$ 在 y 上的 PL 条件: 对于任意固定的 x , $\max_y f(x, y)$ 有非空解集和有限最优值, 存在 $\mu > 0$ 使得 $\|\nabla_y f(x, y)\|^2 \geq 2\mu \left[\max_y f(x, y) - f(x, y) \right], \forall x, y$ 。

4. 随机交替梯度投影算法

根据文献[17]所述, 在随机环境中求解 NC-PL 极小极大优化问题, 基于交替梯度下降 - 上升算法 (AGDA) 和 Smoothed-AGDA 算法 [16], 采用随机梯度 $G_x(x, y, \xi)$ 和 $G_y(x, y, \xi)$ 进行更新, 其中 $G_x(x, y, \xi)$ 和 $G_y(x, y, \xi)$ 分别是 $\nabla_x f(x, y)$ 和 $\nabla_y f(x, y)$ 的无偏随机估计且方差有界 $\sigma^2 > 0$, 构造了新算法 Stoc-AGDA 和 Stoc-Smoothed-AGDA。由于上述算法具有良好的数值表现, 对于凸-PL 极小极大优化问题, 本文考虑将随机梯度结合到交替梯度投影算法(AGP)中, 以下给出随机交替梯度投影算法的框架:

算法 2 随机交替梯度投影算法(SAGP)

步骤 1 输入固定步长 $\tau_1 > 0$, $\tau_2 > 0$, 初始点 (x_0, y_0) 及参数 $b_0 > 0$,

步骤 2 for all $t = 0, 1, 2, \dots, T - 1$ do,

步骤 3 抽取两个独立同分布样本 ξ_1^t, ξ_2^t ,

步骤 4 $x_{t+1} = x_t - \tau_1 [G_x(x_t, y_t, \xi_1^t) + b_t x_t]$,

步骤 5 $y_{t+1} = y_t + \tau_2 G_y(x_{t+1}, y_t, \xi_2^t)$,

步骤 6 end for,

步骤 7 随机从 $\{(x_t, y_t)\}_{t=0}^{T-1}$ 中均匀抽取 (\tilde{x}, \tilde{y}) 。

5. 数值实验

本节将 SAGP 算法应用于求解具有线性生成器的 Toy WGAN 模型, 本文设置与[17]相同, 并给出实

验结果验证了新算法的有效性，模型表示如下：

$$\min_{\mu, \sigma} \max_{\phi_1, \phi_2} F(\mu, \sigma, \phi_1, \phi_2) \triangleq \mathbb{E}_{(x^{real}, z) \sim D} D_{\phi}(x^{real}) - D_{\phi}(G_{\mu, \sigma}(z)) - \lambda \|\phi\|^2,$$

其中生成器为 $G_{\mu, \sigma}(z) = \mu + \sigma z$ ，判别器为 $D_{\phi}(x) = \phi_1 x + \phi_2 x^2$ ， x 是来自生成器的真实数据或假数据， D 为真实数据和潜在变量的分布。真实数据 x^{real} 和潜在变量 z 的高斯分布数据集来自平均值为 0、方差为 1 的正态分布， x^{real} 产生于 $\hat{\mu} = 0, \hat{\sigma} = 0.1$ 。批量大小固定为 100，实验重复 3 次。实验结果如图 1 所示。

图 1 对比了 SAGP 算法和 SAGDA 算法的收敛速度，其中 x 轴代表迭代次数， y 轴代表随机梯度范数取值范围的变化过程，其中 RMSprop 算法步长为 $5e-4$ ，参数为 0.9；Adam 算法步长为 $5e-4$ ，参数为 0.5 和 0.9；SAGDA 算法步长 $\tau_1 = 1e-1$ ， $\tau_2 = 5e-1$ ；SAGP 算法步长 $\tau_1 = 2e-1$ ， $\tau_2 = 8e-1$ 。图 1 和图 2 分别对应于四种算法在 Gaussian 分布数据集上的实验结果对比。从图中可以看出，本文提出的新算法 SAGP 的收敛速度整体上比 SAGDA 算法快，有效地验证了新算法的可行性。

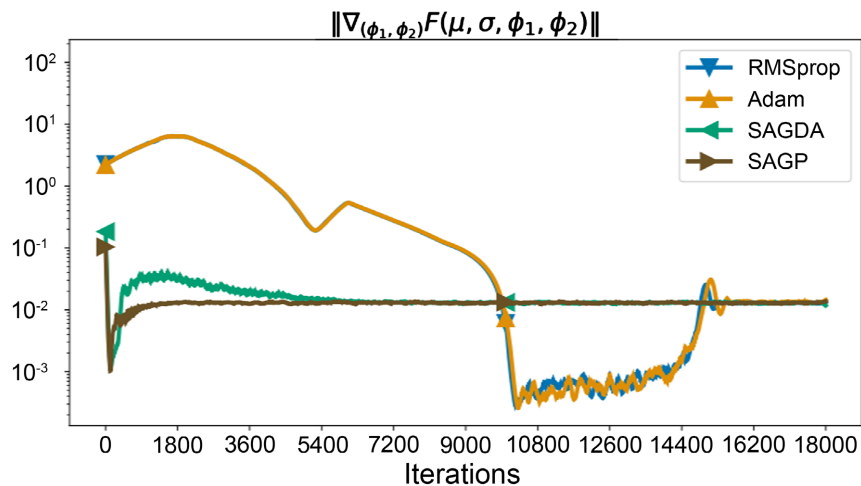


Figure 1. The change process of the range of the random gradient norm of the discriminator on the Gaussian distribution data set

图 1. 在 Gaussian 分布数据集上判别器的随机梯度范数取值范围的变化过程

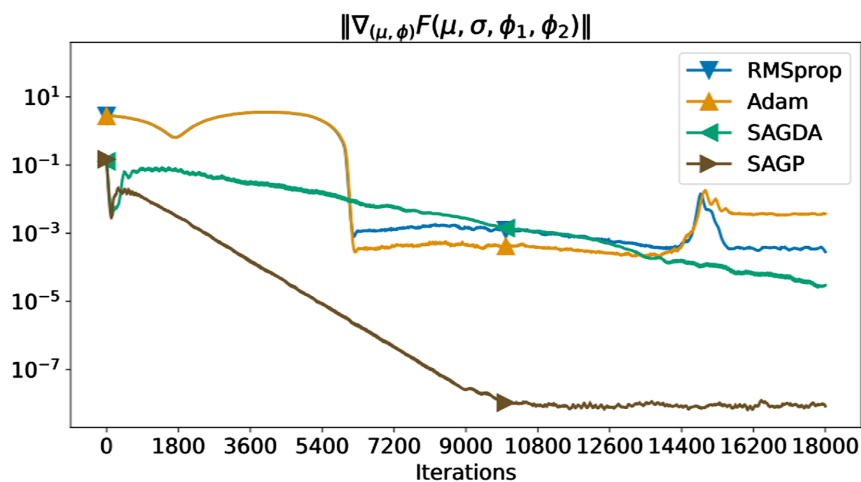


Figure 2. The change process of the value range of the random gradient norm of the generator on the Gaussian distribution data set

图 2. 在 Gaussian 分布数据集上生成器的随机梯度范数取值范围的变化过程

6. 小结

本文在凸-PL 环境下将交替梯度投影算法(AGP)与随机梯度相结合,提出了一种新的随机交替梯度投影算法(SAGP)。从实验结果分析来看,新算法在高斯分布数据集上优于 SAGDA 算法,从而验证了新算法的有效性和可行性。

基金项目

山西省基础研究计划(自由探索类)面上项目(202103021224303); 太原师范学院研究生教育创新项目(SYYJSYC-2287)。

参考文献

- [1] Nemirovski, A. (2004) Prox-Method with Rate of Convergence $O(1/t)$ for Variational Inequalities with Lipschitz Continuous Monotone Operators and Smooth Convex-Concave Saddle Point Problems. *SIAM Journal on Optimization*, **15**, 229-251. <https://doi.org/10.1137/S1052623403425629>
- [2] Nesterov, Y. (2007) Dual Extrapolation and Its Applications to Solving Variational Inequalities and Related Problems. *Mathematical Programming*, **109**, 319-344. <https://doi.org/10.1007/s10107-006-0034-z>
- [3] Monteiro, R.D.C. and Svaiter, B.F. (2010) On the Complexity of the Hybrid Proximal Extragradient Method for the Iterates and the Ergodic Mean. *SIAM Journal on Optimization*, **20**, 2755-2787. <https://doi.org/10.1137/090753127>
- [4] Monteiro, R.D.C. and Svaiter, B.F. (2011) Complexity of Variants of Tseng's Modified F-B Splitting and Korpelevich's Methods for Hemivariational Inequalities with Applications to Saddle-Point and Convex Optimization Problems. *SIAM Journal on Optimization*, **21**, 1688-1720. <https://doi.org/10.1137/100801652>
- [5] Abernethy, J., Lai, K.A. and Wibisono, A. (2019) Last-Iterate Convergence Rates for Min-Max Optimization. ArXiv Preprint arxiv: 1906.02027.
- [6] Rafique, H., Liu, M., Lin, Q. and Yang, T. (2018) Non-Convex Min-Max Optimization: Provable Algorithms and Applications in Machine Learning. ArXiv Preprint arXiv: 1810.02060.
- [7] Nouiehed, M., Sanjabi, M., Huang, T., Lee, J.D. and Razaviyayn, M. (2019) Solving a Class of Non-Convex Min-Max Games Using Iterative First Order Methods. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, 8-14 December 2019, 14934-14942.
- [8] Thekumparampil, K.K., Jain, P., Netrapalli, P. and Oh, S. (2019) Efficient Algorithms for Smooth Minimax Optimization. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, 8-14 December 2019, 12659-12670.
- [9] Kong, W. and Monteiro, R.D.C. (2021) An Accelerated Inexact Proximal Point Method for Solving Nonconvex Concave Min-Max Problems. *SIAM Journal on Optimization*, **31**, 2558-2585. <https://doi.org/10.1137/20M1313222>
- [10] Lin, T., Jin, C. and Jordan, M.I. (2020) Near-Optimal Algorithms for Minimax Optimization. In: Abernethy, J. and Agarwal, S., Eds., *Proceedings of Thirty Third Conference on Learning Theory*, PMLR 125, ML Research Press, Maastricht, 2738-2779.
- [11] Lin, T., Jin, C. and Jordan, M.I. (2020) On Gradient Descent Ascent for Nonconvex-Concave Minimax Problems. In: Daumé III, H. and Singh, S., Eds., *Proceedings of the 37th International Conference on Machine Learning*, PMLR 119, ML Research Press, Maastricht, 6083-6093.
- [12] Jin, C., Netrapalli, P. and Jordan, M.I. (2019) Minmax Optimization: Stable Limit Points of Gradient Descent Ascent Are Locally Optimal. ArXiv Preprint arXiv: 1902.00618.
- [13] Lu, S., Tsaknakis, I., Hong, M. and Chen, Y. (2020) Hybrid Block Successive Approximation for One-Sided Non-Convex Min-Max Problems: Algorithms and Applications. *IEEE Transactions on Signal Processing*, **68**, 3676-3691. <https://doi.org/10.1109/TSP.2020.2986363>
- [14] Xu, Z., Zhang, H., Xu, Y. and Lan, G. (2020) A Unified Single-loop Alternating Gradient Projection Algorithm for Nonconvex-Concave and Convex-Nonconcave Minimax Problems. ArXiv Preprint arXiv: 2006.02032.
- [15] Boş, R.I. and Böhm, A. (2020) Alternating Proximal-Gradient Steps for (Stochastic) Nonconvex-Concave Minimax Problems. ArXiv Preprint arXiv: 2007.13605.
- [16] Zhang, J., Xiao, P., Sun, R. and Luo, Z.-Q. (2020) A Single-Loop Smoothed Gradient Descent-Ascent Algorithm for Nonconvex-Concave Min-Max Problems. *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, 6-12 December 2020.

-
- [17] Yang, J., Orvieto, A., Lucchi, A. and He, N. (2022) Faster Single-Loop Algorithms for Minimax Optimization without Strong Concavity. In: Camps-Valls, G., Ruiz, F.J.R. and Valera, I., Eds., *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, PMLR 151, ML Research Press, Maastricht, 5485-5517.
- [18] Chambolle, A. and Pock, T. (2016) On the Ergodic Convergence Rates of a First-Order Primaldual Algorithm. *Mathematical Programming*, **159**, 253-287. <https://doi.org/10.1007/s10107-015-0957-3>
- [19] Daskalakis, C. and Panageas, I. (2018) The Limit Points of (Optimistic) Gradient Descent in Min-Max Optimization. *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montréal, 2-8 December 2018, 9236-9246.
- [20] Polyak, B.T. (1963) Gradient Methods for Minimizing Functionals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, **3**, 643-653.
- [21] Fazel, M., Ge, R., Kakade, S. and Mesbahi, M. (2018) Global Convergence of Policy Gradient Methods for the Linear Quadratic Regulator. In: Dy, J. and Krause, A., Eds., *Proceedings of the 35th International Conference on Machine Learning*, PMLR 80, ML Research Press, Maastricht, 1467-1476.
- [22] Liu, C., Zhu, L. and Belkin, M. (2020) Loss Landscapes and Optimization in Over-Parameterized Non-Linear Systems and Neural Networks. ArXiv Preprint arXiv: 2003.00307.
- [23] Yang, J., Kiyavash, N. and He, N. (2020) Global Convergence and Variance Reduction for a Class of Nonconvex-Nonconcave Minimax Problems. *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, 6-12 December 2020, 1153-1165.
- [24] Guo, Z., Liu, M., Yuan, Z., Shen, L., Liu, W. and Yang, T. (2020) Communication-Efficient Distributed Stochastic AUC Maximization with Deep Neural Networks. In: Daumé III, H. and Singh, A., Eds., *Proceedings of the 37th International Conference on Machine Learning*, PMLR 119, ML Research Press, Maastricht, 3864-3874.
- [25] Cai, Q., Hong, M., Chen, Y. and Wang, Z. (2019) On the Global Convergence of Imitation Learning: A Case for Linear Quadratic Regulator. ArXiv Preprint arXiv: 1901.03674.
- [26] Du, S., Lee, J., Li, H., Wang, L. and Zhai, X. (2019) Gradient Descent Finds Global Minima of Deep Neural Networks. In: Chaudhuri, K. and Salakhutdinov, R., Eds., *Proceedings of the 36th International Conference on Machine Learning*, PMLR 97, ML Research Press, Maastricht, 1675-1685.
- [27] Rafique, H., Liu, M., Lin, Q. and Yang, T. (2022) Weakly-Convex-Concave Min-Max Optimization Provable Algorithms and Applications in Machine Learning. *Optimization Methods and Software*, **37**, 1087-1121. <https://doi.org/10.1080/10556788.2021.1895152>