ISSN: 2664-9640

A Stochastic Multi-Objective Proximal Gradient Algorithm for Accuracy-Fairness Trade-Off

Si Shi *, Yangdong Xu

School of Science, Chongqing University of Posts and Telecommunications, Chongqing, China.

* Corresponding Author

Abstract

This Paper proposes a stochastic multi-objective proximal gradient algorithm designed to tackle non-smooth stochastic multi-objective optimization problems. The algorithm builds upon the single-objective stochastic proximal gradient method and multi-gradient descent method. The algorithm is further investigated to generate the approximated Pareto front. Finally, the proposed algorithm is applied to a binary classification problem containing prediction accuracy and disparate impact to evaluate its effectiveness. The results demonstrate that the Pareto front generated by the proposed algorithm is helpful for decision-makers to balance the prediction accuracy in binary classification.

Keywords

Stochastic multi-objective optimization, Proximal gradient algorithm, Binary classification.

1. Introduction

Multi-objective optimization (MOO) problems are prevalent in real-life scenarios, in which multiple conflicting objectives need to be addressed simultaneously. In contrast to single-objective optimization, improving one objective may lead to a decline in the performance of other objectives. To tackle this issue, the concept of the Pareto optimal solution is introduced, which reveals the trade-offs between the different objectives. The objective function value corresponding to the Pareto optimal solution is referred to as the Pareto front. Analyzing the Pareto front assists decision-makers in balancing each objective to make informed decisions that align with their interests.

According to the certainty of the parameters, MOO can be divided into deterministic multi-objective optimization (DMOO) and stochastic multi-objective optimization (SMOO). The former has garnered considerable attention from scholars, leading to the development of various solving algorithms. Scalarization algorithms and heuristic algorithms are the primary approaches for solving DMOO problems, and further details on these methods can be found in [1-3]. However, both of them have drawbacks. The former requires prior information and increases the decision-making burden. The latter is difficult to obtain convergence analysis theoretically. Based on this, in recent years, there has been a growing interest in directly solving DMOO problems by drawing on single-objective optimization algorithms. This approach has emerged as a popular research direction in the field. The primary characteristic of these methods is attempting to move in a direction that simultaneously decreases all objective functions. In the DMOO, we can site steepest descent methods [4,5], Newton methods [6], proximal gradient methods [7-10], and proximal point methods [11] for unconstrained problems; while projected gradient methods [12], augmented Lagrangian methods [13] for constrained problems.

While many multi-objective problems in real life also involve uncertainties, such as transportation networks [14] and agricultural production [15]. Therefore, for SMOO, it is necessary to develop an optimization model and corresponding algorithms with both multi-objective properties and stochastic properties. The primary strategies for solving SMOO problems are the multi-objective methods and stochastic methods. The former method can refer to [16] which simplifies the SMOO problem into a deterministic MOO problem and subsequently employs techniques in DMOO. Meanwhile, the latter strategy begins by reducing the SMOO problem to a single objective stochastic problem and then applying single-objective stochastic optimization techniques. It should be noted that stochastic methods inherit certain limitations of scalarization techniques employed in the DMOO. Therefore, to simplify the problem, it is assumed that the stochastic variables in the individual objectives are mutually independent.

Recently, researchers have sought to address stochastic multi-objective optimization (SMOO) problems by drawing on strategies from the extension of classical scalar optimization methods in DMOO. These approaches often assume that stochastic parameters are independent of each other and then adopt multi-objective methods. As a result, both theoretical and practical achievements have been made in this field. [17] extended the single-objective stochastic gradient algorithm to multi-objective optimization, utilizing the common descent defined in the DMOP. Subsequently, [18] analyzed the stochastic multi-gradient method under strong assumptions but failed to calculate the approximation of the entire Pareto front. Later, [19] addressed the aforementioned limitation to study the Pareto front stochastic multi-gradient methods (PF-SMG) and further applied it to evaluate fairness in binary problems. Following this, [20] further applied PF-SMG to the analysis of accuracy-fairness with multiple sensitive attributes or different fairness measures. Additionally, [21] considered a stochastic alternating algorithm for conflicting bi-objective optimization, where each function is not necessarily smooth.

The work described above primarily focuses on scenarios where the objective functions are smooth. It is noted that the above algorithms no longer work well when the objective functions involve non-smooth terms. Examples of this type can be found in machine learning, such as loss functions with non-smooth regularization terms, such as l_1-norm. Therefore, it is essential to consider solving non-smooth composite SMOO problems using splitting algorithms like proximal gradient methods. Building upon the previous research on multi-objective proximity gradient algorithms (see for [7-10] details) and the exploration of stochastic single-objective proximity gradient algorithms (see [22] for details), this paper investigates a class of proximal gradient methods for SMOO and conducts a numerical experiment to evaluate the efficacy of the proposed methods.

This study provides a stochastic extension of the proximal gradient algorithm in DMOO. Indeed, as discussed below, the primary motivation for this research is to handle non-smooth composite SMOO problems. An experimental result on the binary classification problem shows that the proposed algorithm is feasible.

This paper is structured as follows. Section 2 introduces the problem under investigation and proposes a stochastic multi-objective proximal gradient algorithm. Section 3 reports the results of numerical experiments conducted for a binary classification problem. Section 4 finally presents a summary and discusses future research directions.

2. Algorithm Framework

This paper focuses on a stochastic multi-objective optimization problem which is characterized by the following structure:

$$\begin{aligned} \min \quad & F(x) + H(x) \\ \text{s.t.} \quad & x \in \mathbb{R}^n, \end{aligned} \tag{1}$$

where $F \coloneqq (f_1(x), \cdots, f_m(x))^\top$ and $f_i(x) = \mathbb{E}[f_i(x, w]$ is a proper, continuous differentiable convex function; $H \coloneqq (h_1(x), \cdots, h_m(x))^\top$ and $h_i(x)$ is a proper, convex function (not necessarily differentiable); $\mathbb{R}^n$ represents a feasible region without uncertainty.

Now we draw on the stochastic multi-gradient method documented in and construct its corresponding expression for the differentiable part $F(x)$ of problem (1), that is

$$\begin{aligned} \min_{\lambda \in \mathbb{R}^n} \quad & \left\| \sum_{i=1}^{m} \lambda_i g_i(x_k, w_k) \right\|^2 \\ \text{s.t.} \quad & \lambda \in \Delta^m, \end{aligned} \tag{2}$$

where $g_i(x_k, w_k)$ is the approximate estimate of $\nabla f_i(x_k)$. The optimal value of (2) is defined as $\lambda^g(x_k, w_k)$, and the convex combination coefficients depend on x_k and w_k. Then the common descent direction of $F(x)$ is defined by

$$g(x_k, w_k) = \sum_{i=1}^{m} \lambda_i^g(x_k, w_k) g_i(x_k, w_k)\,. \tag{3}$$

Subsequently, we introduce the stochastic multi-objective proximal gradient (SMPG) algorithm for addressing problem (1), which is presented in the following framework:

Stochastic Multi-Objective Proximal Gradient (SMPG) Algorithm

1. Input an initial point $x_0 \in \mathbb{R}^n$ and a step size sequence $\{t_k\}_{k\in\mathbb{N}} > 0$.
2. **for** $k = 0,1, \ldots,$ **do**
3. Compute the approximate estimates $g_i(x_k, w_k)$ of $\nabla f_i(x_k)$. Solve problem (2) and (3) to obtain the common descent direction $g(x_k, w_k)$ for $F(x)$ with $\lambda^g(x_k, w_k)$.
4. Solve the following subproblem:
$$x^{k+1} \in \operatorname{argmin}\left\{H(x) + \frac{1}{2t_k}\|x - x_k - t_k g(x_k, w_k)\|^2 e\right\}.$$
5. Set $x^k = x^{k+1}$.
6. **end for**

The framework of SMPG is borrowed from the single-objective proximal gradient algorithm, which includes both gradient-like steps and proximal-like steps. The first step utilizes the stochastic multi-objective multi-gradient method in [19] to obtain the common descent direction for the differentiable part.

Further, we draw on the Pareto Front stochastic multi-gradient algorithm developed by [19] The algorithm framework corresponding to SMPG to generate the entire Pareto front is

Pareto-Front Stochastic Multi-Objective Proximal Gradient (PF-SMPG) Algorithm

1. Given a list of starting points L_0, Choose parameters $r, p, q \in \mathbb{N}$.
2. **for** $k = 0,1, \ldots,$ **do**
3. Set $L_{k+1} = L_k$.
4. **for** each x in the list L_{k+1} **do**
5. Add r perturbed points to the list L_{k+1} from a neighborhood of x.
6. **end for**
7. **for** each point x in the list L_{k+1} **do**
8. **for** $t = 1, \ldots, p$ **do**
9. Apply q iterations of the **SMPG** algorithm starting from x.

ISSN: 2664-9640

Add the final output point x_t to the new list L_{k+1}.
end for
end for
Remove all the dominated points from L_{k+1}.
end for

Remark 1 Since stochastic optimization presents a challenge in determining a suitable stopping criterion. Thus PF-SMPG algorithm employs either the maximum number of iterations or the maximum number of points on the Pareto frontier as the termination criterion.

3. Numerical Experiment

3.1. Stochastic Bi-Objective problem

In the field of machine learning, achieving fairness and prediction accuracy can be a challenging task. Typically, the goals of maximizing prediction accuracy and maximizing fairness conflict with each other. This is because the prediction model may heavily rely on sensitive attributes like gender, race, and age, which could lead to biased predictions for certain groups. Therefore, fair machine learning needs to strike a delicate balance between maximizing prediction accuracy and ensuring fairness for all groups. The Pareto front generated by solving this optimization problem represents the possible trade-off between accuracy and fairness. Ultimately, this approach enables us to design fair machine-learning models that reduce bias while maintaining high levels of prediction accuracy. Referring to [23], this paper selects disparate impact as fairness criteria.

In our research context, the training dataset comprises the non-sensitive feature vector Z, a binary sensitive attribute A and binary labels Y. Given access to N samples $\{z_i, a_i, y_i\}_{i=1}^{N}$. Let a binary predictor $\hat{Y} = \hat{Y}(Z; x) \in \{-1, +1\}$ be a function of the parameters and solely based on the non-sensitive features Z. As mentioned above, the trade-off between predictive accuracy and fairness may be mathematically formulated as a stochastic bi-objective optimization problem, which can be expressed as follows:

$$\min \left(\mathbb{E}\left[\ell\left(\hat{Y}(Z;x), Y\right)\right], \mathrm{CV}\left(\hat{Y}(Z;x)\right)\right)^{\top}. \tag{4}$$

To simplify the experiment, one can approximate the first objective using an empirical logistic regression loss, which is expressed as

$$f_1(c, b) = \frac{1}{N}\sum_{i=1}^{N} \log\left(1 + \exp\left(-y_i(c^T z_i + b)\right)\right). \tag{5}$$

Actually, (5) serves as an indicator of prediction accuracy. Here we add a regularization term $\lambda_1/2\,\|c\|_1$ to avoid over-fitting. The second objective in (4) is the CV score which measures disparate impact. CV score can be approximated by the

$$f_2^{DI}(c; b) = \left(\frac{1}{N}\sum_{i=1}^{N}(a_i - \bar{a})(c^{\top} z_i + b)\right)^2, \tag{6}$$

where a_i is the expected value of i-th sensitive attribute and $\bar{a}$ is an approximated value of using N samples. $\lambda_2/2\,\|c\|_1$ can be added to (6) to decrease the estimation error. Thus, we actually solve a stochastic bi-objective optimization problem described below.

$$\min \left(f_1(c; b) + \frac{\lambda_1}{2}\|x\|_1, f_2^{DI}(c; b) + \frac{\lambda_2}{2}\|x\|_1 \right)^{\top}. \tag{7}$$

3.2. Numerical Results

Some details of the numerical experiment for solving the problem (7) are as follows:

(a) The dataset consisted of two non-sensitive features and one sensitive feature, where gender was designated as the sensitive feature. A value of 0.0 for the sensitive feature denoted membership in the female group, while a value of 1.0 indicated membership in the male group.
(b) The initial step size is set as 2, followed by step size decay every 200 iterations. λ_1 and λ_2 are set to 2×10^{-2}. Additionally, the maximum number of iterations and non-dominated points in the list are limited to 1000 and 1500, respectively.

The experimental results are shown in Figure 1. Figure 1(a) displays the Pareto front obtained by the PF-SMPG algorithm, revealing the inherent trade-off between accuracy and fairness. Figure 1(b) illustrates that with an increase in f_2^{DI}, the percentage of high-income females declines, indicating that predictors with high accuracy are discriminatory against females. Based on the findings in Figure 1(c), we can conclude that there is a positive correlation between the value of f_2^{DI} and the CV score for this dataset. In Figure 1(d), the X-axis range is in the vicinity of 5% which reveals that reducing the variable by approximately 5% can eliminate differential effects on accuracy.

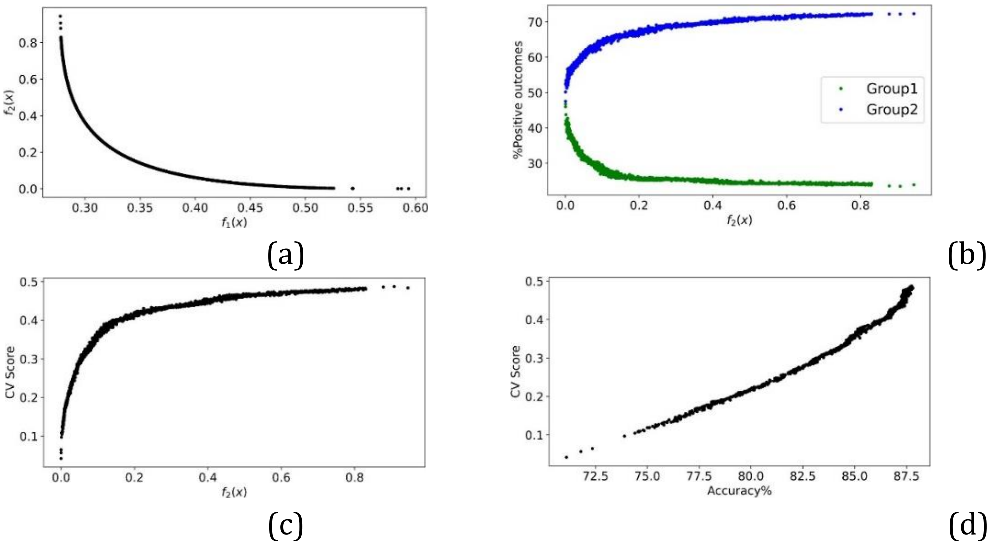

Figure 1: Trade-off results for synthetic dataset

4. Conclusion

This paper aims to study a stochastic multi-objective proximal gradient algorithm designed to tackle non-smooth stochastic multi-objective optimization problems. On this basis, PF-SMPG is constructed to generate the whole Pareto front. The proposed algorithm is then applied to a binary classification problem, and experimental results demonstrate its feasibility. Notably, given the prevalence of non-convex objective functions in deep learning neural networks, further exploration of related versions of SMPG algorithms to address these issues is crucial for future research.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant number: 11801051) and the Natural Science Foundation of Chongqing (Grant number: cstc2019jcyjmsxmX0075).

References

[1] G. Eichfelder: Scalarizations for adaptively solving multi-objective optimization problems, Computational Optimization and Applications, Vol. 44 (2009), p.249-273.

[2] K. Deb, A. Pratap, S. Agarwal, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, Vol. 6 (2002) No.2, p.182-197.

[3] S. Bandyopadhyay, S. Saha, U. Maulik, et al. A simulated annealing-based multiobjective optimization algorithm: AMOSA, IEEE Transactions on Evolutionary Computation, Vol. 12 (2008) No.3, p.269-283.

[4] J. Fliege, B.F. Svaiter: Steepest descent methods for multicriteria optimization, Mathematical Methods of Operations Research, Vol. 51 (2000), p.479-494.

[5] M. El Moudden, A. El Mouatasim: Accelerated diagonal steepest descent method for unconstrained multiobjective optimization, Journal of Optimization Theory and Applications, Vol. 188 (2021) No.1, p.220-242.

[6] J. Fliege, L.G Drummond and B.F. Svaiter: Newton's method for multiobjective optimization, SIAM Journal on Optimization, Vol. 20 (2009) No.2, p. 602-626.

[7] H. Tanabe, E.H. Fukuda and N. Yamashita: Proximal gradient methods for multiobjective optimization and their applications, Computational Optimization and Applications, Vol. 72 (2019), p.339-361.

[8] R.I. Boţ, S.M. Grad . Inertial forward–backward methods for solving vector optimization problems, Optimization, Vol. 67 (2018) No.7, p.959-974.

[9] Y. Bello-Cruz, J.G. Melo and R.V. Serra: A proximal gradient splitting method for solving convex vector optimization problems, Optimization, Vol. 71 (2022) No.1, p.33-53.

[10] H. Tanabe, E.H. Fukuda and N. Yamashita: Convergence rates analysis of a multiobjective proximal gradient method, Optimization Letters, Vol. 17 (2023) No.2, p.333-350.

[11] H. Bonnel, A.N. Iusem and B.F. Svaiter: Proximal methods in vector optimization, SIAM Journal on Optimization, Vol. 15 (2005) No.4, p.953-970.

[12] L.G. Drummond, A.N. Iusem: A projected gradient method for vector optimization problems, Computational Optimization and applications, Vol. 28 (2004), p.5-29.

[13] G. Cocchi, M. Lapucci: An augmented Lagrangian algorithm for multi-objective optimization, Computational Optimization and Applications, Vol. 77 (2020) No.1, p.29-56.

[14] W.J. Gutjahr, A. Pichler: Stochastic multi-objective optimization: a survey on non-scalarizing methods, Annals of Operations Research, Vol. 236 (2016), p.475-499.

[15] A. Chen, J. Kim, S. Lee, et al. Stochastic multi-objective models for network design problem, Expert Systems with Applications, Vol. 37 (2010) No.2, p.1608-1619.

[16] M. Li, Q. Fu, P. Guo, et al. Stochastic multi-objective decision making for sustainable irrigation in a changing environment, Journal of Cleaner Production, Vol. 223 (2019), p.928-945.

[17] F. Poirion, Q. Mercier and J.A. Désidéri: Descent algorithm for nonsmooth stochastic multiobjective optimization, Computational Optimization and Applications, Vol. 68 (2017), p.317-331.

[18] Q. Mercier, F. Poirion and J.A. Désidéri: A stochastic multiple gradient descent algorithm, European Journal of Operational Research, Vol. 271 (2018) No.3, p.808-817.

[19] S. Liu, L.N. Vicente: The stochastic multi-gradient algorithm for multi-objective optimization and its application to supervised machine learning, Annals of Operations Research, 2021, doi:10. 1007/s10479-021-04033-z.

[20] S. Liu, L.N. Vicente: Accuracy and fairness trade-offs in machine learning: A stochastic multi-objective approach, Computational Management Science, Vol. 19 (2022) No.3, p.513-537.

[21] S. Liu, L.N. Vicente: Convergence rates of the stochastic alternating algorithm for bi-objective optimization, arXiv:2203.10605, 2022.

[22] L. Rosasco, S. Villa, B.C. Vũ. Convergence of stochastic proximal gradient algorithm, Applied Mathematics and Optimization, Vol. 83 (2020), p.891-917.

[23] S. Barocas, A.D. Selbst: Big data's disparate impact, California Law Review, Vol. 104 (2016), p.671-732.