DISCLAIMER

Final Technical Report

Project Title: Randomized Federated Learning
for Nonsmooth, Nonconvex, and Hierarchical Optimization

Principal Investigator: Farzad Yousefian
Recipient Organization: Rutgers University – New Brunswick

Co-Principal Investigator: Uday V. Shanbhag
Co-Recipient Organization: University of Michigan –Ann Arbor

Award Number: DE-SC0023303

Total Project Costs: $400,000

Project Period: Start Date 9/1/2022 – End Date 8/31/2024

No-cost extension: Start Date 9/1/2024 – End Date 8/31/2025

Sponsoring Office: U.S. Department of Energy, Office of Science
Program Office: Advanced Scientific Computing Research (ASCR)

Project Manager: Dr. Hal Finkel

Report Date: December 30, 2025

Disclaimer

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research (ASCR), under Award Number DE-SC0023303.

Executive Summary

This project delivers new mathematical models, algorithms, and theoretical guarantees for federated scientific machine learning (FL) in the presence of nonsmoothness, nonconvexity, and hierarchical structures. The research addresses three foundational optimization settings that arise in DOE mission-relevant applications, including: (i) nonsmooth nonconvex stochastic optimization, such as for deep ReLU networks and robust learning; (ii) bilevel optimization, as encountered in federated hyperparameter tuning and hierarchical decision-making problems; and (iii) minimax optimization, which underpins distributionally robust FL. To resolve these challenges, the project develops a unified framework of randomized, inexact, zeroth-order federated optimization methods that achieve provable convergence, communication efficiency, and sample-complexity guarantees.

Major outcomes include the development and analysis of smoothing-enabled zeroth-order FL algorithms and their inexact implicit extensions for decentralized bilevel optimization, minimax problems, and stochastic mathematical programs with equilibrium constraints. The project introduces randomized convolution-based smoothing to construct stochastic generalized gradients for nonsmooth objectives, applies Clarke subdifferential calculus to establish convergence theory, and designs inexact implicit schemes to control bias propagation in hierarchical formulations. Empirical evaluations on federated neural networks, federated hyperparameter optimization, and fair classification problems in FL validate the practical effectiveness of the proposed approaches.

Collectively, the project contributes to the theoretical and computational foundations of federated scientific machine learning and to workforce development through graduate training and the dissemination of results via publications and seminars.

Contents

1 Introduction **5**

2 Relevance to ASCR **6**

3 Technical Approach **6**

3.1 Algorithm Design 7

3.2 Theoretical Analysis 7

3.3 Results 8

4 Accomplishments **8**

5 Publications and Products **9**

6 Inventions and Patents **11**

7 Impact **11**

8 Conclusions **12**

1 Introduction

Federated learning (FL) has emerged as a foundational paradigm for scientific machine learning (SciML) in settings, where data are distributed, heterogeneous, and subject to privacy constraints. These characteristics are central to many DOE mission-relevant applications, including scientific discovery workflows, large-scale simulations, and data-driven modeling across geographically distributed facilities. While FL enables collaborative model training without centralized data aggregation, its deployment in SciML settings introduces significant mathematical and computational challenges that are not adequately addressed by existing methods.

The objectives of this project are to advance mathematical models, tools, and algorithms for federated scientific machine learning by addressing decentralized nonsmooth, nonconvex, and hierarchical optimization problems arising in a wide range of DOE mission-relevant applications. The research is organized around three interconnected technical thrusts that pursue the following objectives:

1. Devise a class of randomized, smoothing-enabled zeroth-order federated learning methods with provable communication, iteration, and sample-complexity guarantees for nonsmooth nonconvex FL;

2. Devise inexact randomized implicit variants of the methods in (1) with provable guarantees for federated bilevel optimization problems and their two-stage game-theoretic generalizations;

3. Extend the implicit zeroth-order algorithmic framework in (2) to address federated minimax optimization problems.

A defining feature of many SciML problems is the presence of *nonsmoothness*, *nonconvexity*, and *hierarchy* in the underlying optimization formulations. Representative examples include training deep ReLU neural networks, robust learning with composite objectives, hyperparameter optimization formulated as bilevel programs, and distributionally robust learning and fair classification problems in SciML posed as minimax optimization. For such problems, standard gradient-based FL algorithms, largely developed under smoothness, convexity, or single-level assumptions, lack rigorous performance guarantees and often fail to scale under realistic data and system heterogeneity. These limitations are further exacerbated in federated environments by restricted communication, partial client participation, and limited access to gradient or subgradient information.

This project addresses these challenges by offering a unified mathematical and algorithmic framework for federated optimization that accommodates nonsmooth, nonconvex, and hierarchical structures. The research focuses on the development and analysis of a class of randomized zeroth-order FL methods that rely on local function evaluations rather than explicit derivatives, addressing emerging SciML applications in FL under realistic assumptions. By integrating tools from nonsmooth analysis, randomized smoothing, implicit optimization, and inexact computation, the project establishes provable convergence and complexity guarantees for three core problem classes: nonsmooth nonconvex federated optimization, federated bilevel optimization, and federated minimax optimization. These advances significantly broaden the scope of FL problems that can be addressed with mathematical rigor and practical scalability in DOE-relevant SciML applications.

2 Relevance to ASCR

The outcomes of this project directly align with the mission of the DOE Office of Advanced Scientific Computing Research (ASCR) by advancing the mathematical foundations, algorithms, and computational methodologies needed to enable next-generation scientific machine learning at scale. ASCR places strong emphasis on developing rigorous, scalable, and communication-efficient methods that can operate effectively on heterogeneous and distributed computing systems. The federated optimization frameworks developed in this project address these priorities by providing provably convergent algorithms that are robust to data heterogeneity, limited communication, and decentralized computation, that are among the key challenges in extreme-scale and distributed scientific computing environments.

From a mathematical perspective, the project contributes new theory for nonsmooth, nonconvex, bilevel, and minimax optimization in federated settings, expanding the scope of problems that can be treated within SciML beyond traditional smooth and single-level formulations. The use of Clarke subdifferential calculus, randomized smoothing, and implicit inexact methods provides foundational tools that are broadly applicable to ASCR-supported research in optimization, uncertainty quantification, and data-driven modeling. From an algorithmic and computational standpoint, the proposed zeroth-order and communication-efficient methods allow for scalability across large numbers of heterogeneous clients and computing platforms, consistent with ASCR's focus on performance portability and resilience in emerging computing architectures.

Finally, the project contributes to ASCR's workforce development and community-building goals through the training of graduate students and postdoctoral researchers at the intersection of optimization, machine learning, and high-performance computing. The dissemination of results through publications, curricular development, and seminars strengthens the ASCR community engaged in distributed learning for scientific applications. Collectively, the project advances the ASCR mission in applied mathematics and computational science that support DOE's scientific and engineering endeavors.

3 Technical Approach

This project develops a unified federated learning framework for addressing optimization problems that are nonsmooth, nonconvex, constrained, and hierarchical. Such problem structures arise naturally in scientific machine learning, including robust learning, hyperparameter tuning, adversarial training, and multi-stage decision-making under uncertainty. Standard FL methods are typically designed for addressing problems with smooth objectives and are not equipped to handle nondifferentiability, implicit objectives, hierarchy, or lower-level constraints. The core technical approach of this project is to overcome these limitations by combining a randomized smoothing technique with zeroth-order (gradient-free) optimization, resulting in communication-efficient algorithms suitable for decentralized and heterogeneous computing environments.

A key approach is to formulate a wide range of hierarchical problems, including bilevel optimization, minimax formulations, and single-stage and two-stage stochastic mathematical programs with equilibrium constraints, as instances of a single nonsmooth, nonconvex federated objective, defined in an implicit fashion. Because the implicit objective is often unavailable in closed form and may lack gradients, our methodology avoids the reliance on first-order information and instead, we consider a smoothed approximate problem amenable to analysis and computation. This approximate problem preserves the structure of the original problem while enabling the use of efficient randomized algorithms in federated settings.

3.1 Algorithm Design

The algorithmic framework is built around a smoothed zeroth-order federated optimization scheme. Nonsmoothness and nonconvexity are handled through randomized smoothing of the objective, while client-specific constraints are incorporated via Moreau smooth penalty functions.

For hierarchical problems in FL, such as bilevel (and minimax) FL problems, the framework incorporates inexact solutions of the lower(inner)-level problem computed in a federated manner. Rather than assuming exact resolution of these lower-level problems, our algorithms explicitly account for such approximation errors by adaptively controlling the inexactness. This approach allows for preserving communication efficiency through limited lower(inner)-level computations.

Overall, the algorithms are distributed, communication-efficient, and privacy-aware with respect to local constraints, and naturally compatible with decentralized clients.

3.2 Theoretical Analysis

The theoretical analysis establishes convergence guarantees for the proposed algorithms under mild assumptions, notably requiring only Lipschitz continuity of the objective rather than smoothness. The analysis shows that stationary points of the smoothed problem correspond to meaningful approximate stationary solutions of the original nonsmooth problem, providing a suitable notion of solution quality in regimes where exact stationarity is computationally intractable.

For both single-level and hierarchical FL problems, the project derives iteration and communication complexity bounds that match or improve upon existing results for smooth nonconvex FL. The analysis also quantifies the effect of inexact lower-level solutions in hierarchical problems and shows how carefully controlling approximation errors prevents bias accumulation. These results provide the first known complexity guarantees for federated learning methods applied to general nonsmooth, nonconvex, and constrained hierarchical problems.

3.3 Results

The project's results demonstrate that it is possible to design provably convergent, communication-efficient FL algorithms without relying on gradients, smoothness, or unconstrained formulations. The proposed framework significantly broadens the class of problems that can be addressed in federated scientific machine learning and provides a unifying perspective for single-level, bilevel, minimax, and equilibrium-constrained optimization in decentralized regimes.

Analytically, the project establishes convergence to approximate stationary solutions for nonsmooth nonconvex federated objectives and extends these guarantees to hierarchical settings with implicit objectives. The results rigorously justify the use of smoothing and zeroth-order methods in FL and show that approximation errors arising from inexact lower-level solutions can be systematically controlled. These contributions close a major theoretical gap in the FL literature, where prior methods required strong smoothness assumptions or excluded constrained and hierarchical formulations.

4 Accomplishments

The key accomplishments of this project are as follows.

- Algorithmic and theoretical advances. Among the key achievements, this project has led to the design and analysis of several new zeroth-order methods with explicit performance guarantees for addressing challenging optimization problem classes. Such problem classes include nonsmooth and nonconvex stochastic optimization with a Lipschitz continuous objective function, federated nonsmooth and nonconvex stochastic optimization, bilevel programs with distributed stochastic objective functions in the both upper and lower levels, minimax problems with distributed stochastic objective functions in the both inner and outer levels, distributed stochastic mathematical programs with equilibrium constraints in both single-stage and two-stage settings, among others. Notably, this research project allowed for the development of amongst the first non-asymptotic rate and overall complexity analysis for a class of single-stage and two-stage stochastic mathematical programs with equilibrium constraints. This effort in turn paved the way for developing potentially one of the first convergent schemes, equipped with rate and complexity guarantees, for computing equilibria to multi-leader multi-follower games afflicted by uncertainty [1–5].

- Training of graduate and undergraduate students. This project provided extensive training opportunities for both graduate and undergraduate students in advanced computational methods for federated learning. Several doctoral students at Rutgers University, Penn State University, and the University of Michigan were actively involved in the research, contributing to the design, analysis, and implementation of the new computational algorithms developed in this project. These doctoral students were paired with undergraduate students through programs such as the Aresty Research Program at Rutgers, providing mentorship in research methods and hands-on experience with algorithm development. Together, the students and PIs guided the undergraduates in preparing multiple research posters that showcased the novel algorithmic federated schemes, fostering a collaborative and interdisciplinary learning environment.

- Curricular development. PI Yousefian offered a doctoral course titled "Distributed and Stochastic Optimization" in Spring 2023 and 2024 at Rutgers. The course comprised of nine modules. He incorporated some of the research outcomes of the project in two of those modules. These included a module on randomized zeroth-order stochastic methods and another module on federated learning. PI Shanbhag has offered several graduate level courses at his home institution, including "Stochastic Optimization".

- SIAM conference. The PIs organized a minisymposium at the SIAM Conference on Optimization (OP23) held from May 31 to June 3, 2023 in Seattle, WA. The minisymposium included 12 talks and was titled "On Addressing Nonsmoothness, Hierarchy, and Uncertainty in Optimization and Games". In one of the talks, the doctoral student presented the outcome of the research.

- Outreach. PI Yousefian served as a co-lead for a Special Issue on Federated, Distributed Learning and Analytics (FDLA) for the IISE Transactions journal.

5 Publications and Products

1. Qiu, Yuyang, Uday Shanbhag, and Farzad Yousefian. "Zeroth-order methods for nondifferentiable, nonconvex, and hierarchical federated optimization." *Advances in Neural Information Processing Systems* 36 (2023): 3425-3438.

2. Qiu, Yuyang, Uday V. Shanbhag, and Farzad Yousefian. "Zeroth-Order Federated Methods for Stochastic MPECs and Nondifferentiable Nonconvex Hierarchical Optimization." *arXiv*, 22 Sept. 2023, arXiv:2309.13024, https://arxiv.org/abs/2309.13024

3. Marrinan, Luke, Uday V. Shanbhag, and Farzad Yousefian. "Zeroth-order Gradient and Quasi-Newton Methods for Nonsmooth Nonconvex Stochastic Optimization." *SIAM Journal on Optimization* (to appear), *arXiv*, arXiv:2401.08665, Oct. 17 2025, https://arxiv.org/abs/2401.08665

4. Ebrahimi, Mohammadjavad, Uday V. Shanbhag, and Farzad Yousefian. "Distributed gradient tracking methods with guarantees for computing a solution to stochastic MPECs." 2024 *American Control Conference (ACC)* (pp. 2182-2187). IEEE, 2024.

5. Samadi, Sepideh, Daniel Burbano, and Farzad Yousefian. "Achieving optimal complexity guarantees for a class of bilevel convex optimization problems." 2024 *American Control Conference (ACC)* (pp. 2206-2211). IEEE, 2024.

6. Kaushik, Harshal D., Sepideh Samadi, and Farzad Yousefian. "An Incremental Gradient Method for Optimization Problems With Variational Inequality Constraints." *IEEE*

Transactions on Automatic Control, vol. 68, no. 12, Dec. 2023, pp. 7879–7886. IEEE, https://doi.org/10.1109/TAC.2023.3251851

7. Burbano, Daniel, and Farzad Yousefian. “A Fish Rheotaxis Mechanism as a Zero-Order Optimization Strategy.” *IEEE Access*, vol. 11, 2023, pp. 102781–102795, https://doi.org/10.1109/ACCESS.2023.10250765

8. Qiu, Yuyang, Farzad Yousefian, and Brian Zhang. “Iteratively regularized gradient tracking methods for optimal equilibrium seeking.” *arXiv preprint* arXiv:2411.18883 (2024).

9. Samadi, Sepideh, and Farzad Yousefian. “Improved guarantees for optimal Nash equilibrium seeking and bilevel variational inequalities.” *SIAM Journal on Optimization* 35.1 (2025): 369-399.

10. Ebrahimi, Mohammadjavad, Yuyang Qiu, Shisheng Cui, and Farzad Yousefian. “Regularized Federated Methods with Universal Guarantees for Simple Bilevel Optimization.” *arXiv preprint* arXiv:2503.08634 (2025).

11. Samadi, Sepideh, Daniel Burbano, and Farzad Yousefian. “On iteratively regularized first-order methods for simple bilevel optimization.” *arXiv preprint* arXiv:2504.08079 (2025).

12. Ebrahimi, Mohammadjavad, Uday V. Shanbhag, and Farzad Yousefian. “On the Resolution of Stochastic MPECs over Networks: Distributed Implicit Zeroth-Order Gradient Tracking Methods.” *arXiv preprint* arXiv:2505.22916 (2025).

13. Marrinan, Luke, Uday V. Shanbhag, and Farzad Yousefian. “On the Sampling-Based Computation of Nash Equilibria Under Uncertainty via the Nikaido–Isoda Function.” *Vietnam Journal of Mathematics* (2025): 1-25.

14. Zhang, Peixuan, Uday V. Shanbhag, and Ethan X. Fang. “A Smoothed Augmented Lagrangian Framework for Convex Optimization with Nonsmooth Constraints.” *Journal of Scientific Computing*, vol. 104, no. 2, 2025, p. 46. Springer.

15. Yang, Shuoguang, Ethan X. Fang, and Uday V. Shanbhag. “Data-Driven Compositional Optimization in Misspecified Regimes.” *Operations Research*, vol. 73, no. 3, 2025, pp. 1395–1411. INFORMS.

16. Xiao, Zhuoyu, and Uday V. Shanbhag. “Computing Equilibria in Stochastic Nonconvex and Non-monotone Games via Gradient-Response Schemes.” *arXiv*, 2025, arXiv:2504.14056, https://arxiv.org/abs/2504.14056.

17. Xiao, Zhuoyu, and Uday V. Shanbhag. “Computing Quasi-Nash Equilibria via Gradient-Response Schemes.” *Proceedings of Machine Learning Research*, vol. 283, 2025, pp. 1–13.

18. Zhang, Peixuan, and Uday V. Shanbhag. "A Smoothed Augmented Lagrangian Framework for Convex Optimization with Nonsmooth Stochastic Constraints." *Proceedings of the 2024 Winter Simulation Conference*, IEEE, 2024, pp. 3181–3192.

19. Cui, Shisheng, Uday V. Shanbhag, and Mathias Staudigl. "A Regularized Variance-Reduced Modified Extragradient Method for Stochastic Hierarchical Games." *Journal of Optimization Theory and Applications*, vol. 206, no. 1, 2025, p. 11, https://doi.org/10.1007/s10957-025-02683-8.

20. Lei, Jinlong, and Uday V. Shanbhag. "Variance-Reduced Accelerated First-Order Methods: Central Limit Theorems and Confidence Statements." *Mathematics of Operations Research*, vol. 50, no. 2, 2025, pp. 1364–1397, https://doi.org/10.1287/moor.2021.0068.

21. He, Linyun, Uday V. Shanbhag, and Eunhye Song. "Stochastic Approximation for Multi-period Simulation Optimization with Streaming Input Data." *ACM Transactions on Modeling and Computer Simulation*, vol. 34, no. 2, 2024, pp. 6:1–6:27, https://doi.org/10.1145/3617595.

22. Cui, Shisheng, and Uday V. Shanbhag. "Variance-Reduced Splitting Schemes for Monotone Stochastic Generalized Equations." *IEEE Transactions on Automatic Control*, vol. 68, no. 11, 2023, pp. 6636–6648, https://doi.org/10.1109/TAC.2023.3290121.

6 Inventions and Patents

- No inventions resulted from this project.

7 Impact

This project contributes to advancing scientific machine learning (SciML) capabilities relevant to DOE mission challenges. Research insights from the development of robust federated learning (FL) methods facilitate scientific benefits from the extreme heterogeneity of emerging computing technologies and promote fairness in model training. The project establishes new theoretical guarantees for ReLU neural networks and robust SciML, improving the quality of privacy-sensitive client models in applications where FL is increasingly deployed. Performance guarantees for hierarchical and two-stage formulations enable new research directions for federated SciML in game-theoretic settings.

Project outcomes are incorporated into graduate-level courses, including *Advanced Topics in Optimization (Distributed and Stochastic Optimization)* at Rutgers University and *Stochastic Optimization* at Penn State University and the University of Michigan, and are disseminated to the broader research community through publications and seminars. Collaborative activities support the training of graduate students and postdoctoral researchers by engaging them in foundational research in data science and computational methods. Doctoral students also mentor undergraduates through programs such as the Aresty Research Program at Rutgers, guiding them in algorithm development and research poster preparation and fostering a collaborative and interdisciplinary learning environment.

8 Conclusions

This project has contributed to the advancement of theoretical, algorithmic, and computational foundations of federated scientific machine learning for nonsmooth, nonconvex, and hierarchical optimization problems. By developing randomized zeroth-order and inexact implicit optimization methods, we have established provable convergence, iteration, and communication guarantees for a broad class of FL problems, including single-level, bilevel, minimax, and equilibrium-constrained formulations. The integration of randomized smoothing, Clarke subdifferential calculus, and adaptive inexact computation has enabled scalable and communication-efficient algorithms that operate effectively in heterogeneous and decentralized environments, addressing key challenges in DOE-relevant SciML applications.

The research outcomes have both foundational and practical significance. On the theoretical side, this project provides the first known convergence and complexity guarantees for federated learning methods applied to nonsmooth, nonconvex, and hierarchical problems, significantly extending the scope of tractable FL formulations. On the practical side, the developed algorithms were validated empirically on federated neural networks, hyperparameter tuning, and fair classification tasks, demonstrating robust performance under realistic client heterogeneity and partial participation. Beyond algorithmic advances, the project contributed substantially to workforce development, training graduate and undergraduate students in cutting-edge computational methods and disseminating results through publications, software artifacts, and seminars. Collectively, these efforts advance the capabilities of federated SciML, provide a strong foundation for future research, and support DOE's mission in data-intensive scientific computing.

References

[1] M. Wang, P. Chakraborty, and U. V. Shanbhag, "Improving dimension dependence in complexity guarantees for zeroth-order methods via exponentially-shifted gaussian smoothing," in *2024 Winter Simulation Conference (WSC)*. IEEE, 2024, pp. 3193–3204.

[2] I. E. Bardakci, A. Jalilzadeh, C. M. Lagoa, and U. V. Shanbhag, "Probability maximization via minkowski functionals: convex representations and tractable resolution," *Math. Program.*, vol. 199, no. 1, pp. 595–637, 2023. [Online]. Available: https://doi.org/10.1007/s10107-022-01859-8

[3] S. Cui, U. V. Shanbhag, and F. Yousefian, "Complexity guarantees for an implicit smoothing-enabled method for stochastic mpecs," *Math. Program.*, vol. 198, no. 2, pp. 1153–1225, 2023. [Online]. Available: https://doi.org/10.1007/s10107-022-01893-6

[4] S. Cui and U. V. Shanbhag, "On the computation of equilibria in monotone and potential stochastic hierarchical games," *Math. Program.*, vol. 198, no. 2, pp. 1227–1285, 2023. [Online]. Available: https://doi.org/10.1007/s10107-022-01897-2

[5] P. Zhang, U. V. Shanbhag, C. M. Lagoa, and I. E. Bardakci, "Global resolution of chance-constrained optimization problems: Minkowski functionals and monotone inclusions," in *62nd IEEE Conference on Decision and Control, CDC 2023, Singapore, December 13-15, 2023*. IEEE, 2023, pp. 6301–6306. [Online]. Available: https://doi.org/10.1109/CDC49753.2023.10383862