# Quasi-Newton methods for machine learning: forget the past, just sample

| | |
|---|---|
| **Pitcher's Names** | Andrea Landini & Ali Yaghoubi \| FoR category: Machine Learning \| **Date Completed**: October 10th, 2025 |
| **(A) Working Title** | Quasi-Newton methods for machine learning: forget the past, just sample |
| **(B) Basic Research Question** | This research investigates how sampling fresh local curvature information at each iteration, rather than reusing past data, can make quasi-Newton methods more efficient, parallelizable, and effective for large-scale machine learning optimization. |
| **(C) Key paper(s)** | 1. **Byrd, R. H., Hansen, S. L., Nocedal, J., & Singer, Y. (2016). "A stochastic quasi-Newton method for large-scale optimization."** *SIAM Journal on Optimization, 26*(**2), 1008–1031.**<br>– A leading modern reference connecting quasi-Newton updates with stochastic, large-scale machine learning settings.<br>2. **Schraudolph, N. N., Yu, J., & Günter, S. (2007). "A stochastic quasi-Newton method for online convex optimization."** *Artificial Intelligence and Statistics (AISTATS)*.<br>– Early, influential work introducing stochastic quasi-Newton ideas tailored to online and machine learning contexts. |
| **(D) Motivation/Puzzle** | Classical quasi-Newton methods rely on curvature information accumulated from past iterations, which can become stale and unreliable in large-scale, nonconvex machine learning problems. This research is motivated by the puzzle of how to retain the efficiency and curvature awareness of quasi-Newton methods while overcoming their dependence on outdated information. By "forgetting the past" and instead sampling fresh, local curvature data at each iteration, the study aims to improve convergence, scalability, and robustness, especially in distributed and deep learning settings where traditional methods struggle. |
| **THREE** | **Three** core aspects of any empirical research project i.e. the "**ID**io**T**s" guide |
| **(E) Idea?** | 1. **Core idea:** Replace historical curvature updates with freshly sampled local curvature information each iteration to make quasi-Newton methods more accurate, scalable, and parallelizable for machine learning.<br>2. **Hypothesis and variables:** The sampled quasi-Newton approach (independent variable) improves convergence speed and efficiency (dependent variable) compared to classical methods, due to more reliable, locally relevant curvature data.<br>3. **Identification/tension:** No endogeneity issues arise—results follow from algorithmic design. The key theoretical tension contrasts "learning from history" versus "sampling the present" in constructing curvature information. |
| **(F) Data?** | 1. **Data/setting:** Computational experiments using benchmark machine learning datasets such as MNIST, CIFAR-10, and ImageNet, chosen for scalability and comparability. The unit of analysis is the optimization run for each algorithm–dataset pair.<br>2. **Sample size:** Dozens of runs per algorithm–dataset combination with thousands of gradient or Hessian evaluations per run, providing rich iteration-level data.<br>3. **Panel dataset:** Yes, the structure is panel-like with multiple algorithms across datasets and iterations.<br>4. **Data sources:** Publicly available open-source datasets, no hand collection, data generated by code, no survey or funding required.<br>5. **Missing data:** None expected since simulations are deterministic, no merge or data cleaning issues anticipated.<br>6. **Variation/power:** High, as performance metrics vary significantly across algorithms and scales, providing strong statistical power.<br>7. **Other obstacles:** Limited external validity beyond computational settings, and construct validity depends on how well experimental metrics reflect real-world training efficiency. |
| **(G) Tools?** | The empirical framework is an experimental computational design comparing algorithm performance rather than a regression model. The research uses controlled benchmarking across datasets to assess convergence, accuracy, and efficiency. No survey or interview instruments are involved. Python with PyTorch, TensorFlow, NumPy, and SciPy will be used, all readily accessible. Standard statistical tests such as t-tests and ANOVA will assess differences. The data are numeric and fully compatible with this framework. Statistical validity is high due to replication, controlled settings, and consistent experimental procedures. |
| **TWO** | **Two** key questions |
| **(H) What's New?** | The novelty of this research lies primarily in the conceptual idea rather than in the data or tools. The core contribution is the development of sampled quasi-Newton methods that replace historical curvature information with freshly sampled local curvature at each iteration. The data and computational tools are standard and serve a supporting role. The intellectual driver is therefore methodological innovation, while the empirical implementation and data processing |

| | |
|---|---|
| | function as necessary complements to demonstrate and validate the idea. In a conceptual Venn diagram, the idea occupies the central position with data and tools intersecting as supportive components. |
| **(I) So What?** | Understanding the effectiveness of sampled quasi-Newton methods is important because it advances how large-scale machine learning models are optimized. If these methods consistently outperform classical approaches, they could reduce training time, improve convergence stability, and lower computational costs in both academic and industrial applications. The outcomes could influence major decisions in algorithm selection for deep learning, guide the design of scalable optimization software, and shape future research on combining first- and second-order optimization. Ultimately, this contributes to more efficient and reliable deployment of machine learning systems across data-intensive industries. |
| **ONE** | **One** bottom line |
| **(J) Contribution?** | The primary source of contribution to the research literature is **methodological innovation** in optimization for machine learning. This study introduces a new framework for constructing curvature information in quasi-Newton methods based on random sampling rather than historical updates. It extends the theoretical foundations of second-order optimization by proving convergence properties under this sampling scheme and demonstrates empirical superiority on benchmark tasks. The contribution therefore advances both the **theory** and **practical implementation** of large-scale, curvature-based optimization in modern machine learning. |
| (K) Other Considerations | **Collaboration:** Collaboration is desirable, particularly for access to high-performance computing resources and expertise in large-scale machine learning optimization. External collaboration with computational mathematics or computer science groups could strengthen the implementation and evaluation phases.<br>**Target journals:** *Optimization Methods and Software*, *Journal of Machine Learning Research (JMLR)*, or *SIAM Journal on Optimization (SIOPT)* are realistic and appropriately ambitious targets.<br>**Risk assessment:** Overall risk is **moderate**. The “no result” risk is low because the algorithms can always be benchmarked empirically. Competitor risk is moderate, given active research in stochastic and quasi-Newton methods. Obsolescence risk is low, as second-order optimization remains relevant. Main challenges concern algorithm tuning and computational cost rather than the conceptual idea. There are no ethical issues or need for ethics clearance.<br>**Scope:** The scope is well balanced, focusing clearly on algorithmic innovation without being overly narrow or diffuse. |