Master's Thesis

Bridging Optimization and Sampling: The Role of Warmstarts and Mass Matrix Adaptation in SG-MCMC

Department of Statistics
Ludwig-Maximilians-Universität München

Moritz Schlager

Munich, Month Day$^{\text{th}}$, Year

Submitted in partial fulfillment of the requirements for the degree of M. Sc.
Supervised by Prof. Dr. David Rügamer and Emanuel Sommer

Abstract

This is my new abstract.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

- in high dimensional, complex posteriors we are often confronted with ill-conditioned modes
- this might lead to slow convergence when the sampler is not able to adapt to the local curvature of the posterior
- we can show this effect on some simple gaussian (mixture) examples

Note: the visualizations show the true density in the background, while compute the steps of the MCMC sampler on samples of the log density.

1.2 Research Objective & Contributions

- preconditioning the MCMC sampler to accord to the local curvature of the posterior
- investigate whether this leads to better uncertainty quantification and convergence properties
- investigate whether this leads to better exploration of the posterior (detect multi-modality)

This contributes to the fields in the way, that overall less samples/shorter chains are needed to achieve resonably good results.

1.3 Outline

2 Background

2.1 Notation

2.2 Sampling from posteriors

2.2.1 Markov Chains

2.2.2 MCMC samplers

- Metropolis Hastings (Proposal Distribution)
- Sampling using gradient information
- General Recipe

2.2.3 Stochastic Samplers

2.3 Bayesian Deep Learning

- How it works & relation to classical Deep Learning
- Posterior Approximation Prorblem
- Bayesian Deep Ensembles

3 Method

3.1 Background & General Feasibility

- best case would be to compute the inverse Fisher Information Matrix in each step as preconditioning to apply to local geometry of the current position
- cannot be done in our MCMC preconditioning context since it is much to expensive to estimate the FIM in every step from scratch
- we cannot do a Hessian approximation like RmsProp or Adam, since they utilize values depending which depend on previous steps. This does then not meet meet the markov property (MCMC theory breaks)
- this is why we optimize first obtaining a high likelihood soultion and a rough estimation of the local geometry (via Adam like Hessia approximation or IVON)
- using this information might help shorten the warmup/burnin phase without hurting the performance
- during sampling we then want to efficiently sample the a local subspace of the high dimensional posterior where the optimizer landed
- multimodality is tackled through ensembling and lengthy markov chains

3.2 Metrics

- explain how impact of preconditioning is measured and what we aim for
- also acknowledge the limitations of metrics when going from a a simple landscape to a overparameterized model (see appendix B in Sommer et. al. 2024)
- -¿ in high dimensions we rather look at convergence in downstream

3.3 Approach

- formalize the preconditioning step

4 Experiments

4.1 Feasibility Study

Low-dimensional example with a tractable and known posterior distribution.

4.1.1 Experimental Setup

4.1.2 Results

Also discuss the results when showing them.

4.2 Main Experiments

Larger newworks evaluated on real world datasets.

4.2.1 Experimental Setup

Grid of expeiments/configurations to evaluate.

4.2.2 Results

Also discuss the results when showing them.

5 Conclusion

- Summary of the main findings
- Implications of the findings
- Limitations of the current work
- Future research directions

A Appendix

B Experiment Details

B.1 Datasets

B.2 Implementation of DEI-MCMC

used for main experiments

B.3 Computational Resources

C Additional Experiment Results

D Electronic appendix

Data, code and figures are provided in electronic form.

Declaration of authorship

I hereby declare that the report submitted is my own unaided work. All direct or indirect sources used are acknowledged as references. I am aware that the Thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the report as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future Theses submitted. Further rights of reproduction and usage, however, are not granted here. This paper was not previously presented to another examination board and has not been published.

Location, date

Name