Global Convergence Newton

Raffael Colonnello
University of Basel
Raffael.Colonnello@unibas.ch

Fynn Gohlke
University of Basel
Fynn.Gohlke@stud.unibas.ch

Benedikt Heuser
University of Basel
ben.heuser@unibas.ch

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

Newton's method is a popular optimization algorithm that is commonly used to solve optimization problems. It is a second-order optimization algorithm since it uses second-order information of the objective function. Newton's method is known to have fast local convergence guarantees for convex functions. However, the global convergence properties of Newton's method are still an active area of research. The purpose of this project is to survey and analyze various strategies to achieve global convergence.

This sentence will be cited by the sources so i can test if bibtex is working properly [1]

2 Background

In this paper we consider the problem

$$\min_{x \in \mathbb{R}^d} f(\mathbf{x}) \tag{1}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a twice-differentiable function whose Hessian is L-Lipschitz continuous. First-order optimization methods are widely used for such problems due to their low per-iteration computational cost and their suitability for parallelization. However, in the presence of ill-conditioned objective landscapes, these methods often suffer from slow convergence.

In contrast, second-order methods such as Newton's method typically exhibit much faster convergence in these settings. In particular, Newton's method enjoys local quadratic convergence under suitable regularity conditions. Nevertheless, outside the region where these conditions hold its performance may degrade to only sublinear convergence.

In this paper, we explore the theoretical foundations of several Newton-type methods, focusing on line search strategies, regularization techniques, and trust region approaches.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

2.1 Classic Newton's Method

The classical origin of Newton's method is as an algorithm for finding the roots of functions. In this paper it is used to find a local minimum in a function. It uses the update rule:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - eta(\nabla^2 f(\mathbf{x}_k))^{-1} \nabla f(\mathbf{x}_k) \tag{2}$$

The inverse Hessian can be interpreted as transforming the gradient landscape to be more isotropic, thereby improving the conditioning of the problem.

2.2 Cubic Newton

The cubic Newton method was one of the first to achieve a good complexity guarantee globally. It is based on cubic regularization. It uses the update rule:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\nabla^2 f(\mathbf{x}_k) + H||\mathbf{x}_{k+1} - \mathbf{x}_k||\mathbf{I})^{-1} \nabla f(\mathbf{x}_k) \tag{3}$$

2.3 Levenberg and Marquardt method

The Levenberg and Marquardt method is parameterized by a sequence $\{\lambda_k\}_k^\infty$ where usually $\lambda_k \equiv \lambda > 0$. The update rule is:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\nabla^2 f(\mathbf{x}_k) + \lambda_k \mathbf{I})^{-1} \nabla f(\mathbf{x}_k) \tag{4}$$

2.4 Regularized Newton

In their 2023 article Michenko presents a variation of Newton's method that uses the update rule [2]:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\nabla^2 f(\mathbf{x}_k) + \sqrt{H||\nabla f(\mathbf{x}_k)||}\mathbf{I})^{-1} \nabla f(\mathbf{x}_k) \tag{5}$$

where $H > 0$ is a constant. The convergence rate of this algorithm is $\mathcal{O}(\frac{1}{k^2})$. This method uses an adaptive variant of the Levenberg-Marquardt regularization.

References

[1] Slavomír Hanzely, Dmitry Kamzolov, Dmitry Pasechnyuk, Alexander Gasnikov, Peter Richtárik, and Martin Takác. A damped newton method achieves global $(o)(\frac{1}{k^2})$ and local quadratic convergence rate. *Advances in Neural Information Processing Systems*, 35:25320–25334, 2022.

[2] Konstantin Mishchenko. Regularized newton method with global convergence. *SIAM Journal on Optimization*, 33(3):1440–1462, 2023.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? **[TODO]**
   (b) Did you describe the limitations of your work? **[TODO]**
   (c) Did you discuss any potential negative societal impacts of your work? **[TODO]**
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? **[TODO]**
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? **[TODO]**
   (b) Did you include complete proofs of all theoretical results? **[TODO]**
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **[TODO]**
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **[TODO]**
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **[TODO]**
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **[TODO]**
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? **[TODO]**
   (b) Did you mention the license of the assets? **[TODO]**
   (c) Did you include any new assets either in the supplemental material or as a URL? **[TODO]**
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? **[TODO]**
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **[TODO]**
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? **[TODO]**
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? **[TODO]**
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? **[TODO]**

A Appendix

Optionally include extra information (complete proofs, additional experiments and plots) in the appendix. This section will often be part of the supplemental material.