Efficient implementation of limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm

Luca Formaggia

September 30, 2023

An important problem in several applications, notably machine learning [1], is finding the minimum of a given function $f : \mathbb{R}^n \to \mathbb{R}$:

$$?\mathbf{x} : \ \mathbf{x} = \operatorname{argmin} f(\mathbf{x})$$

An overview of classic iterative methods for the problem is found in [2], another important reference is the book og J. Nocedal and S. Wright [6], which is freely downloadable for PoliMi students from the Springer site.

Assuming f be continuous up to its second partial derivatives, a method that can provide a second order convergence is Newton Method, whose generic iterate can be written as

$$\begin{cases} \boldsymbol{H}_k \boldsymbol{\Delta}_k = -\nabla_k f, \\ \mathbf{x}_{k+1} = \mathbf{x}_k + \boldsymbol{\Delta}_k, \end{cases}$$

where $\mathbf{x}_k$ is the iterate at step k, $\boldsymbol{H}_k$ and ∇_k are the Hessian and the Jacobian of f in $\mathbf{x}_k$, respectively. In practice, often the Newton step $\boldsymbol{\Delta}_k$ is limited (at least at the initial iterations) using so-called line-search or trust-region techniques, in order to improve convergence properties. In fact, convergence depends, in general, also on the choice of the first iterate $\mathbf{x}_0$. Details may be found in the cited references.

The algorithm caN become quite expensive also for moderately large n because of the need of computing at each iteration the n^2 partial derivatives forming the Hessian. Therefore, alternatives have been proposed to better the situation while mantaining a convergence higher than the linear one provided by the gradient method.

One of the most successful attempt is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, where the Hessian is replaced by a suitable approximation. You may find a description in the cited literature, as well as on Wikipedia [7]. Also in this case the algorithm normally includes techniques to enlarge the set of initial iterates that lead to convergence by line-search or trust-region techniques.

For large scale problems the use of a full matrix to store the approximation of the Hessian can be too burdensome and a limited memory version, called L-BFGS [8, 5] has been developed.

Objectives

Basic objectives

- Develop a class for the classic BFGS algorithm using the Eigen, eigen.tuxfamily.org, library for the linear algebra part.
- Derive a second class where the BFGS algorithm is replaced by L-BFGS;
- Apply the code to a set of examples (one can handcraft an example)

You are free to look around on the web for solutions and suggestions. However, try to program things by yourself, the objective here is to learn to program a scientific computing code.

More advanced objectives.

- If the cost function f is of the form $f(\mathbf{x}) = \sum_j \|r_j(\mathbf{x})\|^2$, as in least-square problems or deep learning applications, a stochastic variant is possible. One can extend the project to consider this case as well. Some information may be extracted from [4].
- A parallel version implementable on GPU has been proposed in the literature, see [3]. A possible extension is to implement the GPU-accelerated version.

References

[1] L. Batou, F.E. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 2018.

[2] Kelley C.T. *Iterative Methods for Optimization.* SIAM, 1999.

[3] Yung Fei, Guodong Rong, Bing Wang, and Weping Wang. Parallel L-BFGS algorithm on GPU. *Computer and Graphics*, pages 1–9, 2014.

[4] Tian De Guo, Yan Liu, and Cong-Ying Han. An overview of stochastic quasi-Newton methods for large-scale machine learning. *Journal of the Operations Research Society of China*, 2023.

[5] Dong C Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

[6] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2016.

[7] Wikipedia. Broyden–Fletcher–Goldfarb–Shanno algorithm, 2023.

[8] Wikipedia. Limited-memory BFGS, 2023.