

Achieving geometric convergence for distributed optimization with Barzilai-Borwein step sizes

Juan GAO¹, Xin-Wei LIU^{2*}, Yu-Hong DAI^{3,4}, Yakui HUANG² & Peng YANG¹

¹*School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China;*

²*Institute of Mathematics, Hebei University of Technology, Tianjin 300401, China;*

³*LSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China;*

⁴*School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

Received 25 November 2020/Revised 8 February 2021/Accepted 1 April 2021/Published online 9 December 2021

Citation Gao J, Liu X-W, Dai Y-H, et al. Achieving geometric convergence for distributed optimization with Barzilai-Borwein step sizes. *Sci China Inf Sci*, 2022, 65(4): 149204, <https://doi.org/10.1007/s11432-020-3256-x>

Dear editor,

We consider distributed optimization, which can be formulated to minimize the average of all local objective functions:

$$\min_{x \in \mathbb{R}^p} f(x) = \frac{1}{n} \sum_{i=1}^n f_i(x), \quad (1)$$

where each local function $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$ is L_i -smooth and μ_i -strongly convex and known only by agent i . All agents communicate over a time-invariant undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set of agents, \mathcal{E} is the collection of pairs $(i, j) (i, j \in \mathcal{V})$ such that agents i and j can exchange information with each other. Problem (1) has been extensively studied in machine learning and smart grids, e.g., [1, 2].

Distributed gradient methods have been a great success in solving the problem (1), primarily because of their low computational cost and easy implementation. To ensure convergence to the exact solution, distributed gradient descent (DGD) [3] should use a diminishing step size, which may result in a slow convergence rate. With a constant step size, DGD can be fast; however, it only converges to a neighborhood of the exact solution. Recently, distributed gradient methods have achieved significant improvements, which provide us new variants that geometrically converge to the exact solution for smooth and strongly convex functions [4–7]. Ref. [4] proposed an improved DGD, called EXTRA, which exploits the difference of two consecutive DGD iterates with different weight matrices to cancel the steady-state error. A variant of DGD [5], named NEAR-DGD⁺, employs a multi-consensus inner loop strategy to DGD. NEAR-DGD⁺ requires that the number of consensus steps is increasing at an appropriate rate, which leads to additional communications. Moreover, another variant of DGD [6, 7] is based on the dynamic average consensus approach, which replaces local gradients in DGD with tracking gradients. Using an adapt-then-combine strategy, these methods in [6, 7] are capable of using uncoordinated constant step sizes. Note that the aforementioned methods neither

address how to design an appropriate step size for each agent using its local information nor do they consider automatically computing the step sizes.

The centralized Barzilai-Borwein (BB) method is a simple and effective technique for selecting the step size and requires fewer storages and inexpensive computations [8, 9]. Moreover, the BB step size is automatically computed using gradient information. Recent years have witnessed the successful applications of the centralized BB method in image processing and machine learning [9].

Inspired by wonderful features and successful applications of the centralized BB method, we propose a distributed gradient method with Barzilai-Borwein step sizes (DGM-BB-C), which combines an adapt-then-combine variation of the dynamic average consensus approach [6, 7] with multi-consensus inner loops [5]. The primary contributions of our work are highlighted below. (i) Each agent can automatically compute the step size using its local gradient information. The step sizes of DGM-BB-C are not less than $\frac{1}{L_i}$, which provides a selection for a larger step size than previous studies. For example, Refs. [5–7] required that the step sizes are not greater than $\frac{1}{L}$ with $L = \max\{L_i\}$. (ii) By simultaneously using the dynamic average consensus approach and multi-consensus inner loops, DGM-BB-C can seek the exact optimum when the number of consensus steps stays constant, which results in fewer communications than NEAR-DGD⁺ [5]. (iii) In contrast with existing methods [5–7], DGM-BB-C uses the Barzilai-Borwein step sizes and finite consensus steps, which leads to faster convergence. Under suitable conditions, we confirm that DGM-BB-C has geometric convergence to the optimal solution. We numerically show the superiority of DGM-BB-C compared with certain advanced methods and validate our theoretical discoveries.

Distributed Barzilai-Borwein step sizes. We now apply the centralized BB method [8] (described in Appendix A) to the distributed optimization. Note that the step size cannot be straightly computed using the centralized BB method because distributed optimization methods never compute the

* Corresponding author (email: mathlxw@hebut.edu.cn)