Journal of Computational Mathematics
Vol.37, No.5, 2019, 689–703.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1808-m2017-0214

CONVERGENCE RATE OF GRADIENT DESCENT METHOD FOR MULTI-OBJECTIVE OPTIMIZATION*

Liaoyuan Zeng and Yuhong Dai[1)]

Institute of Computational Mathematics and Scientific/Engineering Computing, State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

Email: zengly@lsec.cc.ac.cn, dyh@lsec.cc.ac.cn

Yakui Huang

School of Science, Hebei University of Technology, Tianjin 300401, China

Email: huangyakui2006@gmail.com

Abstract

The convergence rate of the gradient descent method is considered for unconstrained multi-objective optimization problems (MOP). Under standard assumptions, we prove that the gradient descent method with constant stepsizes converges sublinearly when the objective functions are convex and the convergence rate can be strengthened to be linear if the objective functions are strongly convex. The results are also extended to the gradient descent method with the Armijo line search. Hence, we see that the gradient descent method for MOP enjoys the same convergence properties as those for scalar optimization.

Mathematics subject classification: 90C29, 90C52.

Key words: Multi-objective optimization, Gradient descent, Convergence rate.

1. Introduction

Consider the multi-objective optimization problem (MOP)

$$\min_{x\in\mathbb{R}^n} \ F(x) := [F_1(x), \dots, F_m(x)], \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a vector function and the components $F_i(x)$, $i = 1, \dots, m$, are continuously differentiable functions. The multi-objective optimization problem has lots of applications in different fields such as engineering design [19], economic modeling [20], and financial risk management [17]. See [8] for more applications.

In general, it is not expected to find a solution point that minimizes all objective functions at the same time, but a Pareto optimal solution instead. A Pareto optimum means that it is not possible to find any other feasible point whose objective function values are all smaller than those at the Pareto optimum [13,16]. There are many approaches for finding a Pareto optimal solution of MOP. One popular approach in industry is to reformulate MOP as a scalar optimization problem whose objective function is a weighted combination of $F_i(x)$, $i = 1, \dots, m$ [11, 12]. However, this approach may yield an unbounded objective if the weights are not properly chosen. Another popular approach is the evolutionary algorithm which searches for a Pareto optimum in a set of candidate solutions with some genetic operator [5]. This method is able to find an

* Received September 6, 2017 / Revised version received May 7, 2018 / Accepted August 5, 2018 / Published online February 11, 2019 /

1) Corresbonding author

approximate Pareto optimum instead of an exact Pareto optimum. Another kind of approach is descent methods, which extend the traditional descent methods for scalar optimization to solve MOP, such as the gradient descent method [10], the Newton method [9] and the projected gradient method [6].

In this paper, we concern the gradient descent method for MOP which was proposed by Fliege and Svaiter in 2000 [10]. They proposed a subproblem for computing a descent direction for all objective functions and adopted the Armijo line search with backtracking to compute stepsizes. A remarkable property of this method is that it is a parameter free approach. This is quite different from the weighted approach for MOP. With this method, there is no need to analyze the prior information including the relationships and conflictions between the objectives (such information is very important for choosing the weights for the weighted method). Specifically, an example is shown in [9] that the weighted method fails for a large range of weights which lead to unbounded weighted objective, but a descent method works with any initial point.

For scalar optimization problems, there are many works on the convergence rate of the gradient descent method with different stepsize strategies (e.g., [2, 4, 15]). The gradient descent method with constant stepsizes enjoys sublinear convergence for unconstrained scalar optimization problems with convex smooth objective functions and it enjoys linear convergence for the case with strongly convex objective functions (Chapter 2 in [14]). In the literature of MOP, it is proved that the gradient descent method converges to a weak Pareto optimum under standard assumptions [7, 10]. Nevertheless, there is little research on the convergence rate of the gradient descent method for MOP. In [3], the authors pointed out that the convergence rate of the gradient descent method for MOP with strongly convex quadratic objectives is related to the ratio of the largest and the smallest eigenvalues of all Hessian matrices. However, there is no further analysis for the general case with convex objectives.

In this paper, we firstly analyze the convergence rate of the gradient descent method with constant stepsizes for MOP. More precisely, we establish the sublinear convergence of the method for unconstrained MOP with convex and smooth objective functions. The rate is further strengthened to be linear convergence if the objective functions are strongly convex. Then we generalize the convergence results for the case of the Armijo line search.

The paper is organized as follows. Section 2 gives some definitions, assumptions and propositions related to MOP. Specifically, the Pareto optimal condition for MOP and the convexity for objective functions are included. In Section 3, the key subproblem for the descent direction in [10] and the framework of the gradient descent method for MOP are introduced. Our main convergence results are provided in Section 4. Section 5 addresses the generalization of the convergence results to the Armijo line search. Conclusions and discussions are made in the last section.

2. Preliminaries

For two vectors x and y, denote the vector inequality $x \preceq y$ as a partial order which is defined as a componentwise relationship, i.e., $x_i \leq y_i$, $\forall i = 1, \ldots, m$. Denote $x \prec y$ as the strict inequality which is componentwise too. For a vector function $F(x) : \mathbb{R}^n \to \mathbb{R}^m$, $JF(x) \in \mathbb{R}^{m\times n}$ stands for the Jacobian matrix at x, i.e., $JF(x)_{i,j} = \frac{\partial F_i(x)}{\partial x_j}$. We use $\|\cdot\|$ to denote the standard Euclidean norm in the real vector space.

Definition 2.1. *A continuously differentiable function $f(x)$ is called convex on $\mathbb{R}^n$ if for any*