Journal of Computational and Applied Mathematics 255 (2014) 765–777

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Quasi-Newton's method for multiobjective optimization

Žiga Povalej *

Turboinštitut d.d., Rovšnikova 7, 1210 Ljubljana, Slovenia
FMF, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 26 October 2011
Received in revised form 17 May 2013

Keywords:
Multiobjective optimization
Newton's method
Quasi-Newton's method
Pareto optimum

ABSTRACT

In this paper we present a quasi-Newton's method for unconstrained multiobjective optimization of strongly convex objective functions. Hence, we can approximate the Hessian matrices by using the well known BFGS method. The approximation of the Hessian matrices is usually faster than their exact evaluation, as used in, e.g., recently proposed Newton's method for multiobjective optimization. We propose and analyze a new algorithm and prove that its convergence is superlinear.

1. Introduction

In multiobjective optimization, several conflicting objectives have to be minimized simultaneously. Generally, no unique solution exists but a set of mathematically equally good solutions can be identified, by using the concept of *Pareto optimality*. Many applications of multiobjective optimization can be found in engineering [1–6], economics and finance [7,8], medicine [9–11], management and planning [12,13], etc.

There exist many solution strategies to solve the multiobjective optimization problems. One of the basic approaches is the weighting method (see [14]), where one single-objective optimization problem is formed by weighting several objective functions. Similar problem has the ε-constraint method, introduced in [15]. Here, we minimize only the chosen objective functions and we bound the others. Some algorithms require the preordering of the functions due to their importance [16]. First, the most important function is optimized, then the second, etc. The disadvantage of these approaches is that the choice of weights, constraints, or the importance of the functions respectively, is not known in advance and has to be prespecified.

Some other techniques for multiobjective optimization that do not need any a priori information were developed in recent years, e.g., the steepest descent algorithm, studied in [17–19] with at most linear convergence. Other methods depend on heuristic, especially evolutionary approaches. Efficient evolutionary algorithms can be found in [20]. Unfortunately, there exist no convergence proofs and the empirical convergence is quite slow. For a survey on other multiobjective approaches see, e.g., [21,22].

Recently, a parameter-free optimization method for unconstrained multiobjective optimization was developed [23]. It borrows the idea of the Newton's method for single-objective optimization. The necessary assumption is that the objective functions are twice continuously differentiable but no other parameters or ordering of the functions are needed. The authors show that the rate of convergence is at least superlinear and it is quadratic if the second derivatives are Lipschitz continuous.

In this paper, we present a similar multiobjective optimization method, which approximates the second derivative matrices instead of evaluating them. Therefore, the time for one step reduces. The rate of convergence is proven to be superlinear. This concept is analogous to quasi-Newton's method for single-objective optimization.

* Correspondence to: Turboinštitut d.d., Rovšnikova 7, 1210 Ljubljana, Slovenia. Tel.: +386 1 5820 154; fax: +386 1 5820 112.
E-mail address: ziga.povalej@turboinstitut.si.

0377-0427/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cam.2013.06.045

The rest of the paper is as follows. Section 2 introduces the problem, notation and preliminaries. Section 3 derives a direction search program, solved by Karush–Kuhn–Tucker multipliers. Next, we present the algorithm considered. Section 4 establishes some theoretical results. They are needed in Section 5, which contains the main convergence result. The outcomes of numerical simulations and commentaries are presented in Section 6, which is followed by conclusions in Section 7.

2. Notation and preliminaries

Let us start with some notation. Let $\mathbb{R}$ be the set of real numbers, $\mathbb{R}_+(\mathbb{R}_-)$ the set of strictly positive (negative) real numbers and $\mathbb{N}$ the set of positive integers. Assume that $U \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$, is an open set. Furthermore, let an objective function

$$\boldsymbol{F} : U \longrightarrow \mathbb{R}^m,$$

where $m \geq 2$, be given. Function $\boldsymbol{F} = (f_1, f_2, \ldots, f_m)$ is a vector function which components are scalar functions

$$f_j : U \longrightarrow \mathbb{R}, \quad j = 1, 2, \ldots, m.$$

Note that n and m are independent in general.

Throughout the paper we will assume that $\boldsymbol{F} \in \mathcal{C}^2(U)$, i.e., $\boldsymbol{F}$ is twice continuously differentiable on U. For $\boldsymbol{x} \in U$, let $\nabla f_j(\boldsymbol{x}) \in \mathbb{R}^n$ denote the gradient of f_j at $\boldsymbol{x}$ for all $j = 1, 2, \ldots, m$. The matrix $\mathcal{J}\boldsymbol{F}(\boldsymbol{x}) \in \mathbb{R}^{m\times n}$ is the Jacobian matrix of $\boldsymbol{F}$ at $\boldsymbol{x}$, i.e., the j-th row of $\mathcal{J}\boldsymbol{F}(\boldsymbol{x})$ is $\nabla f_j(\boldsymbol{x})^T$ for all $j = 1, 2, \ldots, m$. Let $\nabla^2 f_j(\boldsymbol{x})$ be the Hessian matrix of f_j at $\boldsymbol{x}$, again for all $j = 1, 2, \ldots, m$. By $\mathrm{Im}(M)$ we denote the range of a matrix $M \in \mathbb{R}^{m\times n}$ and the identity matrix is $I \in \mathbb{R}^{n\times n}$. For matrices $M, N \in \mathbb{R}^{n\times n}$ we write $M > N$ if $M - N$ is positive definite.

The Euclidean norm in $\mathbb{R}^n$ will be denoted by $\|\cdot\|$. We use the same notation for the induced operator norms on the corresponding matrix spaces. Let $B(\boldsymbol{x}, r)$ be a closed ball with a center $\boldsymbol{x} \in \mathbb{R}^n$ and radius $r \in \mathbb{R}_+$.

We shall assume strong convexity for f_j on U for all $j = 1, 2, \ldots, m$. Function

$$f : U \longrightarrow \mathbb{R},$$

$f \in \mathcal{C}^2(U)$, is *strongly convex* if for all $\boldsymbol{x}, \boldsymbol{y} \in U$

$$(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^T (\boldsymbol{x} - \boldsymbol{y}) \geq a\|\boldsymbol{x} - \boldsymbol{y}\|^2, \tag{1}$$

for some $a > 0$ (see [24]). It is easy to see that (1) is equivalent to

$$\nabla^2 f(\boldsymbol{x}) \geq aI, \quad \text{for all } \boldsymbol{x} \in U.$$

Strong convexity implies strict and usual convexity. Hence, if f_j are strongly convex, Hessian matrices $\nabla^2 f_j(\boldsymbol{x})$ are positive definite for all $\boldsymbol{x} \in U$ and for all $j = 1, 2, \ldots, m$.

Our task is to find a Pareto optimum of the objective function $\boldsymbol{F}$.

Definition 1. A point $\boldsymbol{x}^* \in U$ is a *Pareto optimum*, if there is no $\boldsymbol{y} \in U$ for which

$$\boldsymbol{F}(\boldsymbol{y}) \leq \boldsymbol{F}(\boldsymbol{x}^*) \quad \text{and} \quad \boldsymbol{F}(\boldsymbol{y}) \neq \boldsymbol{F}(\boldsymbol{x}^*)$$

holds. The point $\boldsymbol{x}^* \in U$ is a *weak Pareto optimum* if there is no $\boldsymbol{y} \in U$ such that

$$\boldsymbol{F}(\boldsymbol{y}) < \boldsymbol{F}(\boldsymbol{x}^*).$$

The inequality signs $\leq$ and $<$ are understood in a componentwise sense. Clearly every Pareto optimum is also weak Pareto optimum.

Sometimes there exist points which are Pareto optima only on a restricted subset of U.

Definition 2. A point $\boldsymbol{x}^* \in U$ is a *local Pareto optimum* if there exists a neighborhood $V \subseteq U$ of $\boldsymbol{x}^*$ such that the point $\boldsymbol{x}^*$ is a Pareto optimum for $\boldsymbol{F}$ restricted on V. Similarly, a point $\boldsymbol{x}^* \in U$ is a *local weak Pareto optimum* if there exists a neighborhood $V \subseteq U$ of $\boldsymbol{x}^*$ such that the point $\boldsymbol{x}^*$ is a weak Pareto optimum for $\boldsymbol{F}$ restricted on V.

Let us introduce a necessary condition for Pareto optimality first. A point $\boldsymbol{x}^* \in U$ is *stationary point* of $\boldsymbol{F}$ if

$$\mathrm{Im}(\mathcal{J}\boldsymbol{F}(\boldsymbol{x}^*)) \cap \mathbb{R}^m_- = \emptyset. \tag{2}$$

This definition was first used in [17]. Note that for $m = 1$, (2) reduces to the classical condition $\nabla f(\boldsymbol{x}^*) = 0$. If $\boldsymbol{x}^*$ is a stationary point of $\boldsymbol{F}$, then from (2) follows that for all $\boldsymbol{p} \in \mathbb{R}^n$ there exists $j_0 \in \{1, 2, \ldots, m\}$ such that $\nabla f_{j_0}(\boldsymbol{x}^*)^T\boldsymbol{p} \geq 0$.

If $\boldsymbol{x} \in U$ is not stationary, then there exists $\boldsymbol{p} \in \mathbb{R}^n$ such that $\nabla f_j(\boldsymbol{x})^T\boldsymbol{p} < 0$ for all $j = 1, 2, \ldots, m$. Function $\boldsymbol{F}$ is continuously differentiable, therefore,

$$\lim_{\alpha\to 0} \frac{f_j(\boldsymbol{x} + \alpha\boldsymbol{p}) - f_j(\boldsymbol{x})}{\alpha} = \nabla f_j(\boldsymbol{x})^T\boldsymbol{p} < 0, \quad j = 1, 2, \ldots, m.$$