Iranian Journal of Operations Research
Vol. 5, No. 1, 2014, pp. 47-54

Quasi-Newton Methods for Nonconvex Constrained Multiobjective Optimization

N. Hoseini Monjezi[1]

Here, a quasi-Newton algorithm for constrained multiobjective optimization is proposed. Under suitable assumptions, global convergence of the algorithm is established.

***Keywords**: Constrained multiobjective program, Quasi-Newton method, Critical point, Global convergence.*

Manuscript was received on 26/10/2013, revised on 03/02/2014 and accepted for publication on 19/09/2014.

1. Introduction

In many areas in engineering, economics and science new developments are made possible by application of modern optimization methods. Optimization problems arising nowadays in applications are mostly multiobjective that is, several competing objectives are aspired simultaneously [8, 9]. However, a single solution may not generally minimize every objective function simultaneously. A concept of optimality which is useful in multiobjective framework is that of Pareto optimality.

We present a numerical algorithm for the following non-convex constrained multiobjective problem,

$$\begin{array}{llll} \text{(P)} & \min & F(x) & \\ & \text{s.t.} & g_i(x) \le 0 & i \in L = \{1, \dots, l\}, \\ & & x \in \mathbb{R}^n, & \end{array}$$

where the objective function $F = (F_1, \dots, F_m)^T : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable and the constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$, for $i \in L$, are continuously differentiable.

We denote the set of feasible region by S:

$$S = \left\{ x \in \mathbb{R}^n \;\middle|\; g_i(x) \le 0,\; i \in L \right\}.$$

Here, we present a quasi-Newton method for computing the critical points of (P) without any convexity assumption. Our aim is to extend the results of [5] for constrained multiobjective problems. Moreover, we reduce the assumptions used in [5]. The advantage of our work are: (1) there is no need to compute the Hessian, (2) convexity assumptions of the functions are not needed. Instead, we use the quasi-Newton method to approximate the second order information of the objective functions.

The organization of the remainder of our work is as follows. In Section 2, we present some preliminaries and give notations. The new algorithm and its properties are described in Section 3. In Section 4, we focus on the analysis of global convergence under some suitable assumptions.

[1] Department of Mathematics, Univetsity of Isfahan, Isfahan, Iran. Email: hoseini_najmeh@yahoo.com

2. Preliminaries

We begin this section by introducing some notations. Suppose that $\mathbb{R}$ is the set of all real numbers, $\mathbb{R}_+$ denotes the set of non-negative real numbers and $\mathbb{R}_{++}$ is the set of positive real numbers.
For any $u, v \in \mathbb{R}^n$, denote

$$u \leq v \iff v - u \in \mathbb{R}^m_+ \iff v_j - u_j \geq 0, \qquad j = 1, \ldots, m,$$

$$u < v \iff v - u \in \mathbb{R}^m_{++} \iff v_j - u_j > 0, \qquad j = 1, \ldots, m.$$

Generally, in multiobjective problems, it is not possible to find a joint optimality solution for all objective functions. So, we need to provide a concept of optimality in the multiobjective framework such as Pareto optimality or efficiency, as follows.

Definition 2.1. A point $x^* \in S$ is said to be a local Pareto optimum of (P) if and only if there exists a neighborhood V of x^* such that there does not exist $x \in V \cap S$ satisfying

$$F(x) \leq F(x*) \quad \text{and} \quad F(x) \neq F(x*). \tag{1}$$

Definition 2.2. A point $x^* \in S$ is said to be a global Pareto optimum of (P) if and only if there does not exist $x \in S$ such that

$$F(x) \leq F(x*) \quad \text{and} \quad F(x) \neq F(x*).$$

These definitions have been extensively used; e.g., see [5, 6]. Obviously, any globally Pareto optimal solution is locally Pareto optimal. The converse is true if S is a convex set and F is $R^m - convex$ (i.e., if F is componentwise convex). We use the following defnition to define a descent type algorithm for constrained multiobjective optimization.

Defnition 2.3. A point $x^* \in S$ is a critical or stationary point for (P) if

$$\Re\big(\nabla F(x^*)\big) \cap (-\mathbb{R}^m_{++}) = \emptyset \ \text{ and } \ \Re\big(\nabla g(x^*)\big) \cap \big(-\mathbb{R}^l_{++}\big) = \emptyset,$$

where $\Re(B)$ denotes the image or range space of B.

Definition 2.4. A direction $d \in \mathbb{R}^n$ is a descent direction for (P) at x if d satisfies the following conditions:

$$F_j'(x; d) < 0, \qquad j \in \{1, \ldots, m\} \tag{2}$$

$$g_i'(x; d) \leq 0, \qquad i \in L, \tag{3}$$

where the directional derivative at x in the direction d is defined as:

$$F_j'(x; d) = \lim_{\alpha \to 0} \frac{F_j(x + \alpha d) - F_j(x)}{\alpha}. \tag{4}$$

It follows from (4) that $F_j'(x; d) = \nabla F_j(x)^T d$, for $j \in \{1, \cdots, m\}$, and $g_i'(x; d) = \nabla g_i(x)^T d$, for $i \in L$. The descent direction d will reduce every objective function value and will not increase any constraint function value if it is used to update the design x. It is easy to see that (2) and (3) imply that $d \in \mathbb{R}^n$ is a descent direction for (P) at x if and only if

$$\begin{aligned} \nabla F_j(x)^T d &< 0, \qquad \forall\, j \in \{1,\dots,m\},\\ \nabla g_i(x)^{\mathrm{T}} d &\le 0, \qquad \forall\, i \in L. \end{aligned}$$

3. Algorithm

Suppose that x is a feasible point for (P), that is, $x \in S$. Given sufficiently small $\epsilon > 0$, we define $d_\epsilon(x)$ as the optimal solution of

$$SP_\epsilon(x) \begin{cases} \min & t \\ \text{s.t.} & \nabla F_j(x)^t d + \dfrac{1}{2}\, d^T B_j(x) d \le t, \quad j \in \{1,\dots,m\} \\ & \nabla g_i(x)^{\mathrm{T}} d \le t, \qquad\qquad i \in L \\ & \| d \| \le 1, \qquad t \le -\epsilon, \qquad d \in \mathbb{R}^n, \end{cases}$$

where $B_j(x)$ is the Hessian of F_j at x or its approximation obtained by a quasi-Newton method. We use the approximation obtained by quasi-Newton methods, and impose $\| d \| \le 1$ to improve the performance as $\| d \| \le 1$ eliminates the possible case $\| d \| \to \infty$.

Lemma 3.1. Let $x \in S$ be given.
(i) If $\epsilon = 0$, then the feasible set of $SP_0(x)$ is nonempty.
(ii) Let $B_j(x)$, for $j \in \{1, \dots, m\}$, be positive semidefinite. If the feasible set of $SP_\epsilon(x)$ is nonempty, then x is noncritical and any feasible point $d_\epsilon(x)$ is a descent direction for (P); otherwise, x is a good estimate of the critical point for (P).

Proof. (i). If $\epsilon = 0$, then $d = 0$ is always a feasible point of $SP_\epsilon(x)$.
(ii). Suppose $d_\epsilon(x)$ is a feasible point of $SP_\epsilon(x)$. Then, for $j \in \{1,\cdots,m\}$,

$$\nabla F_j(x)^T d_\epsilon(x) \le t - \frac{1}{2}\, d_\epsilon(x)^T B_j(x) d_\epsilon(x) < -\epsilon < 0,$$

and

$$\nabla g_i(x)^T d_\epsilon(x) \le t \le -\epsilon < 0, \qquad \forall\, i \in L.$$

Hence, $d_\epsilon(x)$ is a descent direction for (P) and

$$\begin{aligned} \Re\big(\nabla F(x^*)\big) \cap (-\,\mathbb{R}^m_{++}) &\ne \emptyset,\\ \Re\big(\nabla g(x^*)\big) \cap \big(-\,\mathbb{R}^l_{++}\big) &\ne \emptyset. \end{aligned}$$

Therefore, x is noncritical.

Now, we show that if the feasible set of $SP_\epsilon(x)$ is empty for a given ϵ, then there does not exist any descent direction at x. Suppose, by contradiction that $\bar{d} \in \mathbb{R}^n$ is a descent direction at x for (P). We have

$$\begin{aligned} \nabla F_j(x)^T \bar{d} &< 0, \qquad j \in \{1,\dots,m\}\\ \nabla g_i(x)^{\mathrm{T}} \bar{d} &\le 0, \qquad i \in L. \end{aligned}$$

Therefore, there exists $0 < \bar{\alpha} < 1$ such that for all $\alpha \in (0, \bar{\alpha}]$,

$$\alpha \nabla F_j(x)^T \bar{d} + \frac{1}{2}\,\alpha^2 \bar{d}^T B_j(x)\bar{d} \le 0, \qquad j \in \{1,\dots,m\},$$
$$\alpha \nabla g_i(x)^T \bar{d} \le 0, \qquad i \in L.$$

If we set

$$-\epsilon = \left\{\alpha\ \nabla F_j(x)^t \bar{d} + \frac{1}{2}\,\alpha^2\,\bar{d}^T\,B_j(x)\bar{d}, \alpha\,\nabla g_i(x)^T \bar{d}\ \right\}.$$

for any $\alpha \in (0, \bar{\alpha}]$, then $\alpha\bar{d}$ is feasible for $SP_\epsilon(x)$. Thus, we arrive at a contradiction which complete the proof.

Lemma 3.2. Suppose that $x_k \in S$ and a sufficiently small $\epsilon > 0$ is given. Let $\bar{d}_k$ be the solution of $SP_\epsilon(x)$ and $\alpha_k \in (0, 1)$ be the step-length. Then,

$$x_{k+1} = x_k + \alpha_k \bar{d}_k$$

is a feasible point.

Proof. Since $x_k \in S$, it follow that

$$g_i(x_k) \le 0, \qquad \forall i \in L.$$

Since $\bar{d}_k$ is the solution of $SP_\epsilon(x)$, then it is feasible and

$$\nabla g_i(x_k)^T \bar{d}_k \le t < 0, \qquad \forall i \in L.$$

By using Taylor series, we obtain:

$$g_i(x_{k+1}) = g_i\big(x_k + \alpha_k \bar{d}_k\big) = g_i(x_k) + \alpha_k \nabla g_i(x_k)^T \bar{d}_k + \alpha_k \parallel \bar{d}_k \parallel \phi(x_{k+1}, x_k),$$

where $\lim_{\parallel x_{k+1} - x_k \parallel \to 0} \phi(x_{k+1}, x_k) = 0.$

Hence,

$$\frac{g_i(x_{k+1}) - g_i(x_k)}{\alpha_k} = \underbrace{\nabla g_i(x)^T \bar{d}_k}_{<0} + \parallel \bar{d}_k \parallel \phi(x_{k+1}, x_k) \le 0 .$$

Therefore, $x_{k+1} \in S$ and the proof is complete.

Based on the above results, we now state the following algorithm.

Algorithm 1. A quasi-Newton Algorithm.

Step 0: Let $x_0 \in S$ be the initial point. Let a sufficiently small positive scalar $\epsilon > 0$ and a positive semidefinte initial matrix $B_j(x_0)$, for $j \in \{1, \cdots, m\}$, be given. Set $k = 0$.
Step 1: Construct $SP_\epsilon(x)$. If it is infeasible then stop, else solve $SP_\epsilon(x_k)$ and compute $d_\epsilon(x_k)$.
Step 2: Choose the step size α_k by Armijo-like rule such that $x_{k+1} = x_k + \alpha_k d_\epsilon(x_k) \in S$.
Step 3: Update $B_j(x_{k+1})$ with the BFGS method for $j \in \{1, \cdots, m\}$. Set $k = k + 1$. Go to Step 1.

Remarks:

(1) In Step1, if x_k is noncritical, then by solving $SP_\epsilon(x_k)$, we can find $d_\epsilon(x_k)$ for (P) at x_k such that

$$\nabla F_j(x_k)^T d_\epsilon(x_k) < 0, \qquad j \in \{1, \dots, m\},$$
$$\nabla g_i(x_k)^T d_\epsilon(x_k) \leq 0, \qquad i \in L.$$

Hence, there exists $\bar{\alpha}_j^1 > 0$ such that

$$F_j\big(x_k + \alpha d_\epsilon(x_k)\big) < F_j(x_k), \qquad \forall\, \alpha \in \big(0, \bar{\alpha}_j^1\big),$$

and there exists $\bar{\alpha}_i^2 > 0$ such that

$$g_i\big(x_k + \alpha d_\epsilon(x_k)\big) < g_i(x_k), \qquad \forall\, \alpha \in \big(0, \bar{\alpha}_i^2\big).$$

Now, let $\bar{\alpha}_1 = \min_{j=1,\dots,m} \{\bar{\alpha}_j^1\}$, and $\bar{\alpha}_2 = \min_{i \in L}\{\bar{\alpha}_i^2\}$ and

$$\bar{\alpha} := \min\{\bar{\alpha}_1, \bar{\alpha}_2\}.$$

Therefore α can be chosen in Step 2.

(2) In Step 3, $B_j(x_{k+1})$ needs to be updated. We use the BFGS update formula to update $B_j(x_{k+1})$ similar to the one given in [4]. The quasi-Newton matrix $B_j(x_{k+1})$, for $j \in \{1, \cdots, m\}$, is updated as follows:

$$B_j(x_{k+1}) = B_j(x_k) - \frac{B_j(x_k) s_k s_k^T B_j(x_k)}{s_k^T B_j(x_k) s_k} + \frac{r_{jk} r_{jk}^T}{s_k^T r_{jk}},$$

where $r_{jk} = \theta_{jk}\, y_{jk} + (1 - \theta_{jk}) B_j(x_k) s_k$, $s_k = x_{k+1} - x_k, y_{jk} = \nabla F_j(x_{k+1}) - \nabla F_j(x_k) + \big(A(x_{k+1}) - A(x_k)\big)^T \lambda_{k+1}$ and

$$\theta_{jk} = \begin{cases} 1, & \text{if } s_k^T y_{jk} \geq 0.2\, s_k^T B_j(x_k) s_k \\ \dfrac{0.8 s_k^T B_j(x_k) s_k}{s_k^T B_j(x_k) s_k - s_k^T y_{jk}}, & \text{if } s_k^T y_{jk} < 0.2\, s_k^T B_j(x_k) s_k, \end{cases}$$

$A(x) = [\nabla g(x)^T]^T$ is the Jacobian of the constraints. The approximate Hessian matrix B_k generated by this formula is positive semidefinite.

4. Global Convergence

In this section, we prove the global convergence of Algorithm 1. First, we make an assumption. Note that here the assumptions used in [5] are reduced:

Assumption (A): Let the level set $L_0 = \{x \in S \mid F(x) \leq F(x_0)\}$ be bounded.

Assumption (A) has been used in proving global convergence of Newton type methods in solving standard scalar optimization problems and also in proving global convergence for solving multiobjective optimization problems.

Now, we state our main theorem result. Define

$$\pi(x) = \sup_{\|d\| \le 1} \min_{\substack{j \in \{1,..,m\} \\ i \in L}} \left\{ -\nabla F_j(x)^T d \, , -\nabla g_i(x)^T d \right\}.$$

Lemma 4.1. A point x^* is critical for (P) if and only if $\pi(x^*) = 0$.

Proof. Suppose x^* is a critical point for (P). By Definition 2.3 we obtain:

$$\Re\big(\nabla F(x^*)\big) \cap (-\mathbb{R}^m_{++}) = \emptyset \quad \text{and} \quad \Re\big(\nabla g(x^*)\big) \cap \big(-\mathbb{R}^l_{++}\big) = \emptyset.$$

It follows that the following system

$$\begin{aligned} &\nabla F_j(x^*)^T d < 0, && j \in \{1, \dots, m\} \\ &\nabla g_i(x^*)^T d < 0, && i \in L. \end{aligned}$$

has no solution. Hence,

$$\min_{\substack{j \in \{1,..,m\} \\ i \in L}} \left\{ -\nabla F_j(x)^T d \, , -\nabla g_i(x)^T d \right\} \le 0.$$

Therefore,

$$\sup \min_{\substack{j \in \{1,..,m\} \\ i \in L}} \left\{ -\nabla F_j(x)^T d \, , -\nabla g_i(x)^T d \right\} = 0,$$

and

$$\pi(x^*) = 0.$$

It is easy to verify that $\pi(x^*) = 0$ implies that x^* is a critical point.

Theorem 4.2. Suppose that Assumption (A) holds, and for sufficiently large k, the step-length α_k, $\frac{1}{3} \le \alpha_k \le \frac{1}{2}$, is accepted. Then, every limit point of the sequence $\{x_k\}$ is critical for (P).

Proof. By Lemma 4.1, it is sufficient to show that every limit point is a solution of $\pi(.) = 0$.

Now, suppose that x^* is a limit point of $\{x_k\}$. Without loss of generality, we may assume that the subsequence $\{x_k\}_{k \in \kappa}$ converges to x^*. Let d_k be a solution of $SP_\epsilon(x_k)$. We have

$$F_j(x_{k+1}) - F_j(x_k) = \alpha_k \nabla F_j(x_k)^T d_k + \frac{1}{2} \alpha_k{}^2 d_k{}^T B_j(x) d_k + O(\| d_k \|^2). \tag{5}$$

Then,

$$\nabla F_j(x_k)^T d_k + \frac{1}{2} d_k{}^T B_j(x) d_k \le t \le -\epsilon < 0.$$

Therefore,

$$\frac{1}{2} {d_k}^T B_j(x) d_k < -\nabla F_j(x_k)^T d_k. \tag{6}$$

From (5), we deduce

$$-\alpha_k \nabla F_j(x_k)^T d_k = -F_j(x_{k+1}) + F_j(x_k) + \frac{1}{2} {\alpha_k}^2 {d_k}^T B_j(x) d_k + O(\| d_k \|^2).$$

From (6), we get

$$-\alpha_k \nabla F_j(x_k)^T d_k \leq -F_j(x_{k+1}) + F_j(x_k) - {\alpha_k}^2 \nabla F_j(x_k)^T d_k.$$

Hence,

$${\alpha_k}^2 \nabla F_j(x_k)^T d_k - \alpha_k \nabla F_j(x_k)^T d_k \leq -F_j(x_{k+1}) + F_j(x_k)$$

$$(\alpha_k - {\alpha_k}^2)(-\nabla F_j(x_k)^T d_k) \leq -F_j(x_{k+1}) + F_j(x_k)$$

$$(\alpha_k - {\alpha_k}^2) \min_{j \in \{1,..,m\}} \left(-\nabla F_j(x_k)^T d_k\right) \leq \min_{j \in \{1,..,m\}} \left(F_j(x_k) - F_j(x_{k+1})\right)$$

We know

$$(\alpha_k - {\alpha_k}^2) \min_{\substack{j \in \{1,..,m\} \\ i \in L}} (-\nabla F_j(x_k)^T d_k, -\nabla g_i(x)^T d) \leq (\alpha_k - {\alpha_k}^2) \min_{j \in \{1,..,m\}} (-\nabla F_j(x_k)^T d_k).$$

So, we obtain:

$$(\alpha_k - {\alpha_k}^2)\, \pi(x_k) \leq \max_{j \in \{1,..,m\}} (F_j(x_k) - F_j(x_{k+1})).$$

Hence,

$$\begin{aligned} \sum_{\substack{k \geq N \\ k \in \kappa}} (\alpha_k - {\alpha_k}^2) \pi(x_k) &\leq \sum_{\substack{k \geq N \\ k \in \kappa}} \max_{j \in \{1,..,m\}} \left(F_j(x_k) - F_j(x_{k+1})\right) \\ &\leq \max_{j \in \{1,..,m\}} (F_j(x_0) - F_j(x^*)) = M. \end{aligned} \tag{7}$$

Now, we prove $\pi(x^*) = 0$, by contradiction. Assume that $\pi(x^*) > 0$. It follows that there exist $\mu > 0$ and $\epsilon_0 > 0$ such that

$$\forall\, 0 < \epsilon < \epsilon_0 \qquad \| x_k - x^* \| \leq \epsilon \quad \text{s.t.} \quad \pi(x_k) \geq \mu > 0.$$

On the other hand, from $\alpha_k \in (0,1)$, it follows

$$0 < {\alpha_k}^2 < \alpha_k.$$

Hence,

$$(\alpha_k - {\alpha_k}^2) > 0.$$

Then, from the assumption, there exists $\overline{K}$ and the step-length $\frac{1}{3} \le \alpha_k \le \frac{1}{2}$ is accepted, for any $k \ge \overline{K}$ and $k \in \kappa$. So, there exists $\delta > 0$ such that $(\alpha_k - {\alpha_k}^2) \ge \delta > 0$ and we have

$$\pi(x_k)\,(\alpha_k - {\alpha_k}^2) \ge \mu\,\delta.$$

This means

$$\sum_{\substack{k \ge N \\ k \in \kappa}} (\alpha_k - {\alpha_k}^2)\,\pi(x_k) \le \sum_{\substack{k \ge N \\ k \in \kappa}} \mu\,\delta = \infty,$$

which contradicts (7). Therefore, $\pi(x^*) = 0$ and x^* is a critical point.

References

[1] Eichfelder, G. (2008), Adaptive Scalarization Methods in Multiobjective Optimization, Springer-Verlag, Berlin, Heidelberg.

[2] Fliege, J. and Grana Drummond, L.M. and Svaiter, B.F. (2009), Newton's method for multiobjective optimization, *SIAM J. Optim.*, 20(2), 602-626.

[3] Geoffrion, A.M. (1968), Proper efficiency and the theory of vector maximization, *J. Optim. Theory Appl.*, 22, 618-630.

[4] Nocedal, J. and Wright, S.J. (2006), Numerical Optimization, 2nd edition, Springer, New York.

[5] Qu, S., Goh, M. and Chan, F.T.S. (2011), Quasi-Newton methods for solving multiobjective optimization, *Operations Research Letters*, 39, 397-399.

[6] Qu, S. and Goh, M. and Liang, B. (2013), Trust region methods for solving multiobjective optimisation, *Optim. Meth & Soft.*, 28(4), 796-811.

[7] Sawaragi, Y. and Nakayama, H. and Tanino, T. (1985), Theory of Multiobjective Optimization, Vol. 176 of Mathematics in Science and Engineering, Academic Press, Orlando, San Diego, New York, London.

[8] Shan, S. and Wang, G.G. (2005), An efficient Pareto set identification approach for multiobjective optimization on black-box functions, *J. Mechanical Design*, 127, 866-874.

[9] Stadtler, W. and Dauer, J. (1992), Multicriteria Optimization in Engineering: A Tutorial and Survey, American Institute of Aeronautics and Astronautics.