J. Nonlinear Var. Anal. 11 (2027), No. 1, pp. 115-132
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.11.2027.1.06

NEWTON'S METHODS USING BREGMAN FUNCTIONS FOR CONVEX MULTIOBJECTIVE OPTIMIZATION

M.A.T. ANSARY[1], JOYDEEP DUTTA[2,*]

[1]Department of Mathematics, Indian Institute of Technology Jodhpur, Jodhpur, India
[2]Department of Economic Sciences, Indian Institute of Technology Kanpur, Kanpur, India

Dedicated to Professor Christiane Tammer on occasion of her 70th birthday

Abstract. In this paper, we focus on a multiobjective convex optimization problem constrained by a closed convex set. We develop a Newton type method for such a problem. The key step here is to solve subproblems involving Bregman distances along with the quadratic approximations of the objective functions at the given iterate. These subproblems allow us to generate descent directions and an Armijo type rule is used to furnish a step length. We also show that, under mild assumptions, the sequence of iterates generated by the algorithm converges to a critical point, providing us with a weak Pareto minimizer.

Keywords. Bregman functions; Convex multiobjective optimization; Newton method; Pareto minimizer.

1. INTRODUCTION

Multiobjective optimization problems play a major role in optimization theory and applications; see, e.g., [9, 15, 16, 18, 22] and the references therein. Computing Pareto and weak Pareto minimizers traditionally uses scalarization technique where the multiobjective problem is replaced by a parametrized single level problem. These methods are typically slow and may not always lead to the required solutions as in the non-convex case finding the global minimizers of scalar problem becomes very difficult. This has been mitigated by the development of evolutionary algorithms which can generate an approximation of the Pareto front pretty quickly at least for problems with two or three objectives; see, e.g., [9, 10]. In the last two decades there has been a great interest in developing descent methods for multiobjective optimization. Attempts were made to develop analogues of the various important descent methods for multiobjective optimization. To the best of our knowledge, Fliege and Svaiter [13] introduced the steepest descent method for multiobjective optimization. This was followed by Fliege, Drummond and Svaiter [12] which presented the analogue of the celebrated Newton's method for multiobjective optimization. Consequently, quasi-Newton methods were also developed for multiobjective optimization; see, e.g., [3, 20, 21] and the references therein. Further sequential

*Corresponding author.
E-mail address: jdutta@iitk.ac.in (J. Dutta)
Received 1 April 2025; Accepted 23 April 2026; Published online 10 November 2026.
©2027 Journal of Nonlinear and Variational Analysis

quadratic programming (SQP for short) based descent methods for multiobjective optimization were developed by Fliege and Vaz [14] and Ansary and Panda [4] in different ways. A variant of the sequential quadratically constrained quadratic programming (SQCQP for short) was developed by Ansary and Panda [1, 2] for multiobjective programming. Bregman distances played a key role in convex programming; see Chen and Teboulle [8]. Bregman distances have distance like properties but are non-Euclidean in general and thus also take care of the curvature of the associated space. In scalar convex optimization, when the proximal point method or the proximal gradient method is used to solve a problem, instead of the squared Euclidean distance which is used in the proximal step, one uses the Bregman distance. That improves the scope, applicability, and flexibility of these methods. For more details on these methods, we refer to Beck [6]. In [12], Fliege, Drummond, and Svaiter developed the Newton method for computing the Pareto minimizers of an unconstrained convex multiobjective optimization problem. They also assumed that for all the points the Hessian matrices of the functions are always positive definite. This tells us that the component functions were thus strictly convex and hence the set of weak Pareto minimizers coincide with the Pareto minimizers. Keeping up the same approach as that of the scalar case at a given iterate, they considered the quadratic approximation of each functions. Then they constructed a max function from these quadratic approximations and then minimized these max function to compute a direction of descent. Once that is done, simple line search approach is used to move to the next improved point. However it is interesting that they also tried to numerically experiment with their algorithm for even non-convex case though convergence results are for strictly convex functions. Our approach aligns closely with that of [12] even though we consider a constrained problem. In order to handle the constraints in a more effective way, we add the Bregman distance to the quadratic approximation that was used in [12]. The use of Bregman distance forces the iterates that are generated to remain in the interior of the constraint set. This provides advantage both from the point of view of theory and also computation. In the last section of this article, we consider this approach to add the Bregman function for multiobjective problems with only convex objective and all of them need not be strongly convex. In fact, such a situation also cover linear multiobjective problems to be minimized over closed convex sets. The reader can see that the constraint set used here has a special structure. It is the closure of the so-called zone of a Bregman function, i.e. a convex function which generates a Bregman distance. However, we demonstrate that some important class of the problems can be formulated in that way and thus making our approach viable.

2. Preliminaries, Newton Method, and Bregman Functions

In this study, we mainly focus on the problem

$$\min_{x\in\bar{S}}(f_1(x),\ldots,f_m(x)), \tag{MOP}$$

where each of the functions $f_j(x)$ are twice continuously differentiable either all convex or all strongly convex, and $\bar{S}$ is the closure of the zone S of a Bregman function, which we define below. By a Pareto minimizer of (MOP), we mean a point $\bar{x}\in\bar{S}$ such that there exists no other point $x\in\bar{S}$ such that $f_j(x)\leq f_j(\bar{x})$, for all $j=1,2,\ldots,m$ and $f_j(x)<f_j(\bar{x})$, for some j. For simplicity of the exposition, we write $\Lambda_m=\{1,2,\ldots,m\}$. Further, if all f_j's, $j\in\Lambda_m$ are strongly convex, we write (S-MOP), instead of (MOP). A point $\bar{x}\in\bar{S}$ is called a weak Pareto minimizer of (MOP) if there exists no $x\in\bar{S}$ such that $f_j(x)<f_j(\bar{x})$, for all $j\in\Lambda_m$. In fact, if

we consider (S-MOP), then the set of Pareto minimizers coincide with the set of weak Pareto minimizers. If we write $F(x) = (f_1(x), \ldots, f_m(x))$, then a point $\bar{x}$ is a weak Pareto minimizer if $F(x) - F(\bar{x}) \in W$, for all $x \in \bar{S}$, and where $W = \mathbb{R}^m \setminus (-\text{int}\mathbb{R}^m_+)$. Note that W is a closed cone but not a convex cone. If $\bar{x}$ is a weak Pareto minimizer of (MOP), then a necessary and sufficient condition for optimality states that there exist $\tau_1 \geq 0, \ldots, \tau_m \geq 0$, and $\tau = (\tau_1, \ldots, \tau_m) \neq 0$ such that

$$0 \in \sum_{j \in \Lambda_m} \tau_j \nabla f_j(\bar{x}) + N_{\bar{S}}(\bar{x}).$$

Note that the zone S of a Bregman function is an open convex set. If $\bar{x} \in S$, then the above optimality condition will read as $0 = \sum_{j \in \Lambda_m} \tau_j \nabla f_j(\bar{x})$ since $N_{\bar{S}}(\bar{x}) = \{0\}$ if $\bar{x} \in S$. Let us also state that solving problem (S-MOP) is itself a major step towards solving problem (MOP). Consider the problem (MOP), where all the functions f_j are convex and need not be strongly convex. For example, if all f_j are linear, or affine. The link between (MOP) and (S-MOP) is made in Section 3.

Let us now introduce the notion of Newton's method for a strongly convex function $f : \mathbb{R}^n \to \mathbb{R}$ over the whole space $\mathbb{R}^n$. Newton's method demands that f is to be a twice continuously differentiable function. Then Newton's method demands that, at the k-th iteration, the direction d^k should be computed as a minimizer of the quadratic approximation of f at the current iterate x^k, i.e.,

$$d^k = \underset{d}{\text{argmin}} \left\{ \langle \nabla f(x^k), d \rangle + \frac{1}{2} \langle d, \nabla^2 f(x^k) d \rangle \right\}.$$

This shows that $\nabla f(x^k) + \nabla^2 f(x^k) d^k = 0$. Thus $d^k = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k)$. This is what is known as the Newton direction and one can easily show that it is a descent direction. Thus we can define the descent step as $x^{k+1} = x^k - \alpha_k (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$, where $\alpha_k > 0$ is a step-size or step length such that $f(x^{k+1}) < f(x^k)$. How can one move from the single objective case to the multi objective one? Let us now consider the problem (S-MOP), i.e., $F(x) = \min(f_1(x), \ldots, f_m(x))$. In this particular case of (S-MOP), we consider it to be unconstrained.

Note that, in Section 3, we consider (S-MOP), with additional constraints. In order to develop the Newton's method for multiobjective optimization, Fliege, Drummond, and Svaiter proceed by taking a clue from the single-objective case but by using the Chebyshev scalarization scheme. Consider $\hat{q}_j(x,d) = \langle \nabla f_j(x), d \rangle + \frac{1}{2} \langle d, \nabla^2 f_j(x) d \rangle$ and $\hat{q}(x,d) = \max_{j \in \Lambda_m} \hat{q}_j(x,d)$. This is the step where the Chebyshev type scalarization is used. Thus, in the multiobjective case, we compute the direction of descent, $d(x)$, by solving the following minimization problem; i.e.,

$$d(x) = \underset{d \in \mathbb{R}^n}{\text{argmin}} \hat{q}(x,d).$$

Since $\hat{q}_j(x,d)$ is strongly convex in d, for each $j \in \Lambda_m$, we see that $\hat{q}(x,d)$ is also strongly convex in d for each x. Thus $d(x)$ is uniquely determined. In fact, the descent step is now followed by finding a step-length $\alpha_k > 0$, using an Armijo type criterion. Fliege, Drummond and Svaiter [12] posed the above minimization problem as the following equivalent form (R-MOP)

$$\begin{array}{ll} \min_{d,t} & t \\ \text{subject to} & \hat{q}_j(x,d) \leq t, \quad \text{for all } j \in \Lambda_m, \end{array}$$

where $(t,d) \in \mathbb{R} \times \mathbb{R}^n$. In fact, $d(x)$ is a solution of $\min_{d \in \mathbb{R}^n} \hat{q}(x,d)$. It can be demonstrated that $d(x) = \underset{d \in \mathbb{R}^n}{\text{argmin}}\, \hat{q}(x,d)$, if and only if there exists a $t = t(x)$ such that $(t(x), d(x))$ is a solution to

the reformulated problem (R-MOP). Note that

$$d(x) = -\left[\sum_{j\in\Lambda_m}\lambda_j(x)\nabla^2 f_j(x)\right]^{-1}\sum_{j\in\Lambda_m}\lambda_j(x)\nabla f_j(x),$$

where $\lambda_j(x)$ is the Lagrange multiplier associated with $(t(x),d(x))$.

As we mentioned in Section 1, one key motivation for this article is from Doikov and Nesterov [11], where they minimized a scalar function by using the Newton method regularized by a Bregman distance. We attempt to do a similar thing for the multiobjective case. We begin by introducing the notion of the Bregman distance.

Consider the open convex set S and let $\overline{S}$ denote its closure. Let $h:\overline{S}\to\mathbb{R}$ be a strictly convex function. The Bregman distance induced by the function h is a function $D_h:\overline{S}\times S\to\mathbb{R}$, given as $D_h(x,y)=h(x)-h(y)-\langle\nabla h(y),x-y\rangle$. Note that $D_h(x,y)\geq 0$ and $D_h(x,y)=0$ if and only if $x=y$. However D_h is not symmetric and the triangular inequality does not hold. D_h is thus a non-Euclidean distance. If h is a twice continuously differentiable strongly convex function, with $\mu>0$ as the modulus of strong convexity, one can show that $D_h(x,y)\geq\frac{\mu}{2}\|x-y\|^2$. The function h is termed as the Bregman function and has the following properties

i): h is continuously differentiable on S;
ii): h is strictly convex and continuous on $\overline{S}$;
iii): for any $\alpha\in\mathbb{R}$, $y\in S$ and $x\in\overline{S}$, $L_1(y,\alpha)=\{z\in\overline{S}:D_h(z,y)\leq\alpha\}$ and $L_2(x,\alpha)=\{z\in S:D_h(x,z)\leq\alpha\}$ are bounded;
iv): if $\{y^k\}$ in S converges to $y^*\in\overline{S}$, then $D_h(y^*,y^k)\to 0$ as $k\to\infty$;
v): let $\{x^k\}$ and $\{y^k\}$ be sequences in S such that y^k converges to $y^*\in\overline{S}$ and $\{x^k\}$ is bounded. If $D_h(x^k,y^k)\to 0$, then $x^k\to y^*$.

The Bregman function h is said to be *boundary coercive* if, for any sequence $\{y^k\}$ which converges to a point y in the boundary of S, $\lim_{k\to\infty}\langle\nabla h(y^k),x-y^k\rangle=-\infty$, for all $x\in S$. Given a proper convex function $\phi:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$, the subdifferential of ϕ at a point x in its domain is given as $\partial\phi(x)=\{v\in\mathbb{R}^n:\phi(y)-\phi(x)\geq\langle v,y-x\rangle,\forall y\in\mathbb{R}^n\}$. If x is not in the domain of ϕ, we set $\partial\phi(x)=\emptyset$. Let $C\subset\mathbb{R}^n$ be a convex set and let $x\in C$. Then the normal cone to C at x is defined as $N_C(x)=\{v\in\mathbb{R}^n:\langle v,y-x\rangle\leq 0,\forall y\in C\}$. If x is the global minimizer of the real-valued convex function ϕ over the convex set C, then the necessary and sufficient condition for optimality is expressed through the inclusion $0\in\partial\phi(x)+N_C(x)$. If we consider the strongly convex function f on $\mathbb{R}^n$ and consider the approach of Doikov and Nesterov [11], then the direction of descent is now a unique minimizer of the problem

$$d(x)=\operatorname*{argmin}_d\left\{\langle\nabla f(x),d\rangle+\frac{1}{2}\langle d,\nabla^2 f(x)d\rangle+\mu D_h(x+d,x)\right\},$$

where $\mu>0$ is the regularization parameter. Consider $h(x)=\frac{1}{2}\langle x,Qx\rangle$, where Q is a symmetric positive definite matrix. Thus

$$D_h(x+d,x)=\frac{1}{2}\langle x+d,Q(x+d)\rangle-\frac{1}{2}\langle x,Qx\rangle-\langle Qx,d\rangle=\frac{1}{2}\langle d,Qd\rangle.$$

Thus $D_h(x+d,x)=\frac{1}{2}\langle d,Qd\rangle$. Hence we can write

$$d(x)=\operatorname*{argmin}_d\left\{\langle\nabla f(x),d\rangle+\frac{1}{2}\langle d,\nabla^2 f(x)d\rangle+\frac{\mu}{2}\langle d,Qd\rangle\right\},$$

which yields $\nabla f(x) + \nabla^2 f(x)d(x) + \mu Q d(x) = 0$. Hence, $d(x) = -\left[\nabla^2 f(x) + \mu Q\right]^{-1} \nabla f(x)$.

3. Strong Convexity, Newton Direction, and Bregman Distances

In this section, our key aim is to focus on the problem (S-MOP) given as

$$\min F(x) = (f_1(x), \ldots, f_m(x)) \quad \text{subject to} \quad x \in \overline{S},$$

where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is a ρ_i-strongly convex function for each $i = 1, \ldots, m$ and S is an open convex set which is the zone of a Bregman function. Let us recall from Section 2 that if f_i is also differentiable, then, for any $x, y \in \mathbb{R}^n$,

$$f_i(y) \geq f_i(x) + \langle \nabla f_i(x), y - x \rangle + \frac{\rho_i}{2}\|y - x\|^2.$$

In our setting, we assume that each f_i is twice continuously differentiable and hence $\nabla^2 f_j(x)$ is positive definite for each $x \in \mathbb{R}^n$. The key idea is to study (S-MOP) rather than (MOP), which is driven by the idea that the solution to (MOP) can be obtained by solving a sequence of multiobjective problems with strongly convex objective. Consider the problem (S-MOP$_k$) given by

$$\min_{x \in \overline{S}} (f_1(x) + \frac{\varepsilon_k}{2}\|x\|^2, \ldots, f_m(x) + \frac{\varepsilon_k}{2}\|x\|^2),$$

where ε_k is a decreasing sequence of positive numbers decreasing to zero. If f_i is convex, for each k , then problem (S-MOP$_k$) is of the form (S-MOP). Since $\overline{S}$ is closed, it is known that, for each k, problem (S-MOP$_k$) has a weak Pareto minimizer which in this case is also a Pareto minimizer. The following theorem links (MOP) with (S-MOP$_k$).

Theorem 3.1. *Let x^k be a weak Pareto minimizer of (S-MOP$_k$). Assume that x^* is a limit point of $\{x^k\}$. Then x^* is a weak Pareto minimizer of (MOP).*

Proof. Since x^k is a weak Pareto minimizer of (S-MOP$_k$), then, for any $x \in \overline{S}$ writing in the vector form, we have, for each k,

$$(F(x) + \frac{\varepsilon_k}{2}\|x\|^2 e) - (F(x^k) + \frac{\varepsilon_k}{2}\|x^k\|^2 e) \in W,$$

where $e \in \mathbb{R}^m$ is a vector whose all components are 1. Since W is closed as $k \to \infty$, then $F(x) - F(x^*) \in W$. As $x \in \overline{S}$ was arbitrarily chosen, we conclude that x^* is the weak Pareto minimizer of (MOP). □

To develop an algorithm for (MOP), we consider the following auxiliary functions q_j, given as

$$q_j(x, d) = \langle \nabla f_j(x), d \rangle + \frac{1}{2}\langle d, \nabla^2 f_j(x) d \rangle.$$

We will consider such functions at each iteration. Observe that, for each given x, the map $d \mapsto q_j(x, d)$ is ρ_j-strongly convex in d with $\rho_j = \lambda_{\min}(\nabla^2 f_j(x))$, for $j = 1, 2, \ldots, m$. These auxiliary functions play an important role in computing the descent direction. However, in order to do that, we need to take the following step which is akin to a scalarization step. For each given x, we construct the function $q(x, .)$ given as $q(x, d) = \max_{j \in \Lambda_m} q_j(x, d)$, where we recall that $\Lambda_m = \{1, \ldots, m\}$. We can conclude that $d \to q(x, d)$ is also a strongly convex function. If we follow the approach of Fliege, Drummond and Svaiter [12], then we need to minimize

$d \mapsto q(x,d)$, whose solutions are descent directions. Instead of just minimizing $q(x,d)$, we add a Bregman distance to it. Consider the following minimization problem which we mark as (P(x))

$$\min_{d\in D(x)} q(x,d)+\frac{1}{2}D_h(x+d,x),$$

where $D(x)$ is given by $D(x)=\{d\in\mathbb{R}^n : x+d\in\bar{S}\}$. Observe that $D(x)$ is a closed convex set for any given $x\in\mathbb{R}^n$. One major reason to augment $q(x,d)$ with $D_h(x+d,x)$, where h is a Bregman function is the following. Since the above optimization problem is strongly convex in d for each x, it has a unique minimizer for all x, which we mark by $d(x)$. Note that $x+d(x)\in S$. With the aid of the Bregman function, we can enforce feasibility. Indeed, we can write $d(x)$ as follows

$$d(x)=\text{argmin}_{d\in D(x)}q(x,d)+\frac{1}{2}D_h(x+d,x).$$

Lemma 3.1. *Let D_h be induced by a boundary coercive function h. Then, for every $x\in S$, $x+d(x)\in S$ and $d(x)\in intD(x)$.*

Proof. The vector $d(x)\in\mathbb{R}^n$ is a minimizer of $q(x,d)+\frac{1}{2}\left(h(x+d)-h(x)-\langle\nabla h(x),d\rangle\right)$ over $D(x)$. Note that

$$0\in\partial\left(q(x,\cdot)+\frac{1}{2}h(x+\cdot)\right)(d(x))-\frac{1}{2}\nabla h(x)+N_{D(x)}(d(x)).$$

Thus there exists $v\in N_{D(x)}(d(x))$ such that $-v+\frac{1}{2}\nabla h(x)\in\partial\left(q(x,\cdot)+\frac{1}{2}h(x+\cdot)\right)(d(x))$. It follows that $\partial\left(q(x,\cdot)+\frac{1}{2}h(x+\cdot)\right)(d(x))\neq\emptyset$. On the contrary, if $x+d(x)\notin S$, then we will show that $\partial\left(q(x,\cdot)+\frac{1}{2}h(x+\cdot)\right)(d(x))=\emptyset$. It is clear that $x+d(x)\in\bar{S}$. So, if $x+d(x)$ is in the boundary of S, which we mark by bd$S$, then $\partial\left(q(x,\cdot)+\frac{1}{2}h(x+\cdot)\right)(d(x))=\emptyset$ if h is boundary coercive. Let us assume that $x+d(x)\in$ bd$S$ and $\hat{\xi}\in\partial\left(q(x,\cdot)+\frac{1}{2}h(x+\cdot)\right)(d(x))$. As $x\in S$ and S is an open set, there exists a vector $\hat{d}\in\mathbb{R}^n$, with $\|\hat{d}\|$ sufficiently small such that $x+\hat{d}\in S$. Consider a sequence of real numbers $\{\varepsilon_k\}_k$ such that $0<\varepsilon_k<1$ and $\varepsilon_k\to 0$ as $k\to\infty$. Construct the following sequence of vectors $\{d^k\}_k$ in $\mathbb{R}^n$ given by $d^k=\varepsilon_k\hat{d}+(1-\varepsilon_k)d(x)$. We have $x+d^k\in S$. Observe that

$$x+d^k=x+\varepsilon_k\hat{d}+(1-\varepsilon_k)d(x)=\varepsilon_k(x+\hat{d})+(1-\varepsilon_k)(x+d(x)).$$

Since $x+\hat{d}\in S$ and $x+d(x)\in$ bd$S$, we find from basic properties of convex sets that $x+d^k\in S$, for each $k\in\mathbb{N}$ (see [5, Chapter 2]). Further, $d^k\to d(x)$, as $k\to\infty$. From the convexity of h, we have $h(x+d^k)-h(x+d(x))\leq-\left\langle\nabla h(x+d^k),d(x)-d^k\right\rangle$. Since $\hat{\xi}\in\partial\left(q(x,\cdot)+\frac{1}{2}h(x+\cdot)\right)(d(x))$, we have $\left[q(x,d^k)+\frac{1}{2}h(x+d^k)\right]-\left[q(x,d(x))+\frac{1}{2}h(x+d(x))\right]\geq\langle\hat{\xi},d^k-d(x)\rangle$. It follows that

$$\varepsilon_k\left\langle\hat{\xi},\hat{d}-d(x)\right\rangle\leq q(x,d^k)-q(x,d(x))+\frac{1}{2}\left\langle\nabla h(x+d^k),d^k-d(x)\right\rangle.$$

From the convexity of q, we see that

$$q(x,d^k)-q(x,d(x))\leq\varepsilon_k q(x,\hat{d})+(1-\varepsilon_k)q(x,d(x))-q(x,d(x))\leq\varepsilon_k\left(q(x,\hat{d})-q(x,d(x))\right).$$

Note that

$$d^k-d(x)=d(x)+\varepsilon_k(\hat{d}-d(x))-d(x),\ d^k-d(x)=\varepsilon_k(\hat{d}-d(x)).$$

Further, one has $\hat{d}-d^k=\hat{d}-(d(x)+\varepsilon_k(\hat{d}-d(x)))=(1-\varepsilon_k)(\hat{d}-d(x))$. Thus $d^k-d(x)=\frac{\varepsilon_k}{(1-\varepsilon_k)}(\hat{d}-d^k)$. Combining the last two inequalities and dividing by $\varepsilon_k>0$, we obtain

$$\begin{aligned}\left\langle\hat{\xi},\hat{d}-d(x)\right\rangle &\leq \left(q(x,\hat{d})-q(x,d(x))\right)+\frac{1}{2\varepsilon_k}\left\langle\nabla h(x+d^k),d^k-d(x)\right\rangle\\ &= \left(q(x,\hat{d})-q(x,d(x))\right)+\frac{1}{2}\left\langle\nabla h(x+d^k),\hat{d}-d(x)\right\rangle.\end{aligned}$$

Multiplying by $(1-\varepsilon_k)$ and using $(1-\varepsilon_k)(\hat{d}-d(x))=\hat{d}-d^k$, we have

$$(1-\varepsilon_k)\left[\left\langle\hat{\xi},\hat{d}-d(x)\right\rangle-\left(q(x,\hat{d})-q(x,d(x))\right)\right]\leq\frac{1}{2}\left\langle\nabla h(x+d^k),\hat{d}-d^k\right\rangle.$$

Set $z=x+\hat{d}$ and $y^k=x+d^k$. As $k\to\infty$, $y^k\to x+d(x)\in\mathrm{bd}S$. Thus

$$(1-\varepsilon_k)\left[\left\langle\hat{\xi},\hat{d}-d(x)\right\rangle-\left(q(x,\hat{d})-q(x,d(x))\right)\right]\leq\frac{1}{2}\left\langle\nabla h(y^k),z-y^k\right\rangle.$$

Since h is boundary coercive, as $k\to\infty$, we have $\langle\nabla h(y^k),z-y^k\rangle\to-\infty$ while the left hand side of the inequality becomes $[\langle\hat{\xi},\hat{d}-d(x)\rangle-(q(x,\hat{d})-q(x,d(x)))]$ as $k\to\infty$, which is finite. This is of course not possible. Hence $\partial\left(q(x,\cdot)+\frac{1}{2}h(x+\cdot)\right)(d(x))=\emptyset$ whenever $x+d(x)\in\mathrm{bd}S$. Since $\partial\left(q(x,\cdot)+\frac{1}{2}h(x+\cdot)\right)(d(x))\neq\emptyset$, we must have $x+d(x)\in S$, and this completes the first assertion of the lemma. The proof that $d(x)\in\mathrm{int}D(x)$ is very simple and thus omitted. This completes the proof. □

What we are proposing here is an interior point approach to solve a strongly convex multiobjective problem constrained by a closed convex set with a particular property, by augmenting a Newton type method with a Bregman function to generate the direction of descent. In some sense this is a new approach since we are able to force every iterate to remain in the interior of the feasible region by augmenting with a Bregman function. However, one may argue that our problem format is very special and our approach may not be able to tackle some important multiobjective problems which arise in practice.

Consider the important problem of finding the weak Pareto minimizer of a strongly convex vector objective function under affine constraints. This is to basically find the weak minimizer over a polyhedral set. We mark this problem as (MOP-2).

$$\begin{aligned}&\min F(x)=(f_1(x),\ldots,f_m(x))\\ &\text{subject to } H(x)=0,\quad x\geq 0,\end{aligned}$$

where each f_i is strongly convex and $H:\mathbb{R}^n\to\mathbb{R}^k$ and $x\geq 0$ means $x_i\geq 0$ for all $i=1,\ldots,n$. Now let us consider a sequence of "penalized" version of the above problem. Each problem is marked as (P_k) and is given as $\min_{x\geq 0}F(x)+\rho_k\|H(x)\|^2e$, where $\rho_k>0$ and $\rho_k\to\infty$ and $e\in\mathbb{R}^m$ such that each of its component is one. Let us denote, for each $k\in\mathbb{N}$, the function φ_k by $\varphi_k(x)=F(x)+\rho_k\|H(x)\|^2e$. It is clear that all the components of $\varphi_k(x)$ are strongly convex in x. Thus, for each $k\in\mathbb{N}$, problems (P_k) are in the problem format (S-MOP) with $\overline{S}=\mathbb{R}^n_+$. Thus $S=\mathbb{R}^n_{++}$. This particular S is the zone of the celebrated Kullback–Leibler divergence, which is an important class of Bregman functions. Let us assume that the problem (MOP-2) has a solution, i.e., it has a weak Pareto minimizer. Since each f_i is strongly convex, we can show that the set of weak Pareto minimizers of (MOP-2) is non-empty. Let $\bar{x}$ be weak Pareto minimizer of (MOP-2). Thus, for any x feasible to (MOP-2), $F(x)-F(\bar{x})\in W$, where we recall that W is a closed cone,

not convex. Moreover it can be demonstrated that, for each k, problem (P_k) has a solution. Let x^k be a weak Pareto minimizer of (P_k) for each k. Assume $\bar{x}$ to be a limit point of $\{x^k\}$. Without loss of generality, we assume that $x^k \to \bar{x}$. Using standard separation arguments, we can show that, for each k, there exists $w^k \in S^m$, where $S^m = \{w \in \mathbb{R}^m_+ : \sum_{j=1}^m w_j = 1\}$, such that x^k solves the following scalar optimization problem which we mark as (SP_k), $\min_{x\geq 0}\langle w^k, \varphi_k(x)\rangle$. Let x_0 be a feasible point of (MOP-2). Then x_0 is also feasible for (P_k) and (SP_k). Hence, for each k, $\langle w^k, \varphi_k(x^k)\rangle \leq \langle w^k, \varphi_k(x_0)\rangle$, which shows

$$\langle w^k, F(x^k)\rangle + \rho_k \|H(x^k)\|^2 \leq \langle w^k, F(x_0)\rangle + \rho_k \|H(x_0)\|^2.$$

Since x_0 is feasible to (MOP-2), we have $H(x_0) = 0$. Thus the above inequality reduces to

$$\langle w^k, F(x^k)\rangle + \rho_k \|H(x^k)\|^2 \leq \langle w^k, F(x_0)\rangle. \tag{3.1}$$

Since S^m is compact, without loss of generality, we can assume that $w^k \to \bar{w}$. Thus $\langle w^k, F(x^k)\rangle \to \langle \bar{w}, F(\bar{x})\rangle$ and $\langle w^k, F(x_0)\rangle \to \langle \bar{w}, F(x_0)\rangle$. For inequality (3.1) to hold as $\rho_k \to \infty$, we must have $\|H(x^k)\| \to 0$ as $k \to \infty$. Let us now demonstrate why this happens. On the contrary, we assume that $\|H(x^k)\|$ is bounded away from zero, i.e., there exists $\delta > 0$ such that $\|H(x^k)\| > \delta$ for all $k \in \mathbb{N}$. As $\{\langle w^k, F(x^k)\rangle\}$ and $\{\langle w^k, F(x_0)\rangle\}$ are bounded sequences, we can conclude that there exist $M_1 > 0$ and $M_2 > 0$ such that $|\langle w^k, F(x^k)\rangle| \leq M_1$, and $|\langle w^k, F(x_0)\rangle| \leq M_2$, for all $k \in \mathbb{N}$. From (3.1), we have, for any $k \in \mathbb{N}$, $-M_1 + \rho_k \|H(x^k)\|^2 \leq M_2$. This shows that, for all $k \in \mathbb{N}$, $\rho_k \|H(x^k)\|^2 \leq M_2 + M_1$. Since $\{H(x^k)\}$ is bounded away from zero by δ and $\rho_k \to \infty$, we have $\rho_k \delta^2 \leq M_1 + M_2$. Since $\rho_k \to \infty$ as $k \to \infty$, it is clear that the above inequality is easily violated for a large k. This shows that $\|H(x^k)\|^2 \to 0$ as $k \to \infty$. Since $x^k \to \bar{x}$, we have $H(\bar{x}) = 0$ proving that $\bar{x}$ is a feasible solution for (MOP-2). We now demonstrate that $\bar{x}$ is a weak minimizer for (MOP-2). Let us consider an arbitrary feasible point x_0 for (MOP-2). Then it is also a feasible solution for each (P_k). Hence, for each k, $(F(x_0) + \rho_k \|H(x_0)\|^2 e) - (F(x^k) + \rho_k \|H(x^k)\|^2 e) \in W$. Since x_0 is feasible, we have $F(x_0) - F(x^k) - \rho_k \|H(x^k)\|^2 e \in W$. Observe that $\rho_k \|H(x^k)\|^2 e \in \mathbb{R}^m_+$ and note that $W + \mathbb{R}^m_+ \subset W$. It follows that $F(x_0) - F(x^k) \in W$. Since W is closed, as $k \to \infty$, we have $F(x_0) - F(\bar{x}) \in W$. Since x_0 was an arbitrary feasible point of (MOP-2), we conclude that $\bar{x}$ is a weak minimizer of (MOP-2). The discussion above can now be summarized in the form of a following theorem.

Theorem 3.2. *For each $k \in \mathbb{N}$, let x^k be weak Pareto minimizer of (P_k). Then any limit point of $\{x^k\}$ is a weak Pareto minimizer of (MOP-2).*

If $\{x^k\}$ turns out to be a bounded sequence, then there exists a limit point, the weak Pareto minimizer. In this case, it is the Pareto minimizer too. Thus we have demonstrated that an important class of multiobjective problems can be solved by successively solving a penalized problem of the form (P_k). Let us now go back to problem $(P(x))$. This problem is equivalent to the problem $(\hat{P}(x))$, given by

$$\begin{aligned} &\min_{(t,d)} \; t + \frac{1}{2} D_h(x+d, x) \\ &\text{subject to} \quad q(x,d) \leq t. \end{aligned}$$

Note that $(P(x))$ is an unconstrained optimization problem with a strongly convex objective. In fact, if $(t, \hat{d})$ is a solution to $(\hat{P}(x))$, then $q(x, \hat{d}) \leq t$ and $\hat{d}$ is the unique minimizer of $(P(x))$.

The advantage of representing problem (P(x)) in the form ($\hat{P}(x)$) is that by writing the Karush-Kuhn-Tucker conditions for the latter problem we can provide an interesting link between the solution of ($\hat{P}(x)$) and the problem (S-MOP) thus linking the solution of (P(x)) and (S-MOP). This justifies the approach used here based on the Bregman function. Observe that problem ($\hat{P}(x)$) can also be stated as

$$\min_{(t,d)} \; t + \frac{1}{2} D_h(x+d,x)$$
$$\text{subject to } d \in D(x), \langle \nabla f_j(x), d\rangle + \frac{1}{2}\langle d, \nabla^2 f_j(x) d\rangle \leq t, j \in \Lambda_m.$$

The Slater condition holds for problem ($\hat{P}$ (x)) since, for any $t > 0$, $(t,0)$ is also feasible. Also, we assume that h is boundary coercive, which guarantees that $x + d(x) \in S$. If $(t(x), d(x))$ is a solution to problem ($\hat{P}$ (x)), then there exists $\lambda(x) \in \mathbb{R}^m_+$ such that the KKT conditions holds, i.e.,

i): $\sum_{j\in\Lambda_m} \lambda_j(x) = 1$
ii): $\frac{1}{2}(\nabla h(x+d(x)) - \nabla h(x)) + \sum_{j\in\Lambda_m} \lambda_j(x)(\nabla f_j(x) + \nabla^2 f_j(x) d(x)) = 0$
iii): $\lambda_j(x)(\langle \nabla f_j(x), d(x)\rangle + \frac{1}{2}\langle d(x), \nabla^2 f_j(x) d(x)\rangle - t(x)) = 0 \quad \forall j \in \Lambda_m$.

The vector function $\lambda : x \mapsto \lambda(x)$ is called the Lagrange multiplier function associated with the problem ($\hat{P}$ (x)). As $(0,0)$ is a feasible solution to the problem ($\hat{P}$ (x)), we have $t(x) + \frac{1}{2} D_h(x+d(x),x) \leq 0$. An interesting question that one may ask at this point is the following. For what values of $t(x)$ and $d(x)$, x will be a weak Pareto minimizer of (S-MOP). The answer will surprisingly turn out to be $t(x) = 0$ and $d(x) = 0$. Let us try to demonstrate this. Before we proceed, let us set, for any $x \in S$, $\theta_x(t,d) = t + \frac{1}{2} D_h(x+d,x)$. Further, if $(t(x), d(x))$ denote the solution of ($\hat{P}$ (x)), then we set $\Theta(x) = \theta_x(t(x), d(x))$. The importance of ($\hat{P}$ (x)) is clear as the next proposition shows us that, under mild conditions, a solution of problem ($\hat{P}$ (x)) is a solution of (S-MOP).

Proposition 3.1. *Let the Bregman distance be induced by a function h which is boundary coercive and twice continuously differentiable on S. Let $x^* \in S$ be such that $(t(x^*), d(x^*))$ is a solution of ($\hat{P}(x^*)$) and $d(x^*) = 0$. Further, assume that Slater condition holds for the problem ($\hat{P}(x^*)$). Then x^* is a weak minimizer of (S-MOP).*

Proof. We see that the Slater condition holds for problem ($\hat{P}(x^*)$). Hence there exists $\lambda(x^*) \in \mathbb{R}^m_+$ such that KKT conditions i), ii) and iii) as described above holds. From i), it is clear that there exists at least one $j \in \Lambda_m$ such that $\lambda_j(x^*) > 0$. In view of iii) and $d(x^*) = 0$, we conclude that $t(x^*) = 0$. Thus $\Theta(x^*) = \theta_{x^*}(t(x^*), d(x^*)) = 0$. Our aim is to show that if $d(x^*) = 0$, then x^* is a critical point of (S-MOP) and hence a weak minimizer of that problem. On the contrary, let us assume that it is not a critical point. Then there exists $d \in \mathbb{R}^n$ such that $\langle \nabla f_j(x^*), d\rangle < 0$ for all $j \in \Lambda_m$. Let $\xi^* = \max_{j\in\Lambda_m} \langle \nabla f_j(x^*), d\rangle$. It is clear that $\xi^* < 0$. Then for each $j \in \Lambda_m$ we can find $\alpha_j > 0$ such that for all $\alpha \in (0, \alpha_j)$

$$\alpha \langle \nabla f_j(x^*), d\rangle + \frac{\alpha^2}{2}\langle d, \nabla^2 f_j(x^*) d\rangle \leq \frac{\alpha \xi^*}{2} < 0.$$

Since $x^* \in S$ and S is open, we may also choose the α_j's so small that $x^* + \alpha d \in S$ for all $\alpha \in (0, \alpha_j)$. Let $\bar{\alpha} = \min_{j\in\Lambda_m} \alpha_j$. This shows that $\left(\frac{\alpha\xi^*}{2}, \alpha d\right)$ is feasible for ($\hat{P}(x^*)$) for every

$\alpha \in (0, \bar{\alpha})$. Observe that

$$\theta_{x^*}\left(\frac{\alpha\xi^*}{2}, \alpha d\right) = \frac{\alpha\xi^*}{2} + \frac{1}{2}D_h(x^* + \alpha d, x^*).$$

As $x^* \in S$ and $x^* + \alpha d \in S$ for $\alpha \in (0, \bar{\alpha})$, by Taylor's theorem there exists $\eta_\alpha \in (0,1)$ such that

$$\begin{aligned}\theta_{x^*}\left(\frac{\alpha\xi^*}{2}, \alpha d\right) &= \frac{\alpha\xi^*}{2} + \frac{1}{2}(h(x^* + \alpha d) - h(x^*) - \langle \nabla h(x^*), \alpha d\rangle) \\ &= \frac{\alpha\xi^*}{2} + \frac{\alpha^2}{4}\langle d, \nabla^2 h(x^* + \eta_\alpha \alpha d)d\rangle.\end{aligned}$$

Since h is twice continuously differentiable on S, the term $\langle d, \nabla^2 h(x^* + \eta_\alpha \alpha d)d\rangle$ remains bounded for all small $\alpha > 0$. Hence we can choose $\alpha \in (0, \bar{\alpha})$ sufficiently small such that

$$\theta_{x^*}\left(\frac{\alpha\xi^*}{2}, \alpha d\right) < 0 = \theta_{x^*}(0,0) = \Theta(x^*).$$

This contradicts the fact that $(t(x^*), d(x^*)) = (0,0)$ is a solution of $(\hat{P}(x^*))$. Hence x^* is a critical point of (S-MOP) and thus a weak Pareto minimizer. This completes the proof. □

3.1. **An algorithm for the strongly convex case.** We are now in a position to develop a descent algorithm for problem (S-MOP). Our approach is an interior point approach, and we generate a sequence $x^k \in S$, $k \in \mathbb{N}$, and each such that $(t(x^k), d(x^k))$ is a solution of $(P(x^k))$. Starting with an initial solution $x^0 \in S$, we generate a sequence $x^k \in S$, which is updated as $x^{k+1} = x^k + \alpha_k d^k$, where $d^k = d(x^k)$, and $\alpha_k > 0$, a suitable step length which can be obtained by an Armijo back-tracking procedure given as

$$f_j(x^k + \alpha_k d^k) \leq f_j(x^k) + \beta\alpha_k t^k \tag{3.2}$$

for some $\beta \in (0,1)$. We now state our algorithm.

Algorithm 1: (Newton method for (S-MOP) with Bregman distance)

Step 1: Initialization Step: Choose a strictly convex and boundary coercive Bregman function h. Choose initial approximation x^0, scalars $r, \beta \in (0,1)$ and $\varepsilon > 0$. Set $k := 0$.

Step 2: Solve the sub-problem $\hat{P}(x^k)$ to find (t^k, d^k).

Step 3: If $||d^k|| < \varepsilon$, then stop. Else go to Step 4.

Step 4: Choose a suitable step length α_k as the first element of geometric sequence $\{1, r, r^2, \dots\}$ for which the Armijo criteria as given above is fulfilled.

Step 5: Update $x^{k+1} = x^k + \alpha_k d^k$; set $k := k+1$ and go to step 2.

Basic Assumptions:

- **A1):** ∇f_j is Lipschitz continuous over $\mathbb{R}^n$ with Lipschitz constant $L_j > 0$, for every $j \in \Lambda_m$, and we set $L = \max_{j\in\Lambda_m} L_j$;
- **A2):** The level set $M = \{x \in \bar{S} : F(x) \leq F(x^0)\}$ is bounded for every $x^0 \in S$, where the inequality is understood componentwise;
- **A3):** For any bounded sequence $\{x^k\}$ in S, with (t^k, d^k), the solution of (P(x^k)) for each $k \in \mathbb{N}$, $\lim_{r\to\infty}(\nabla h(x_r^k + d_r^k) - \nabla h(x_r^k)) \in N_{\bar{S}}(x^*)$, for every subsequence $\{x_r^k\}$ of $\{x^k\}$ with $x_r^k \to x^*$ as $r \to \infty$.

Our first result is to demonstrate that the Armijo criterion, which we discussed above, actually works.

Theorem 3.3. *Let (t^k, d^k) be a solution of $\hat{P}(x^k)$, for some $x^k \in S$. Then $t^k \leq -\frac{\rho}{2}||d^k||^2$, where $\rho = \min\{\rho_1, \rho_2, ..., \rho_m\}$. Further, if $d^k \neq 0$, then the Armijo criterion holds for every $\alpha > 0$, sufficiently small.*

Proof. If (t^k, d^k) is a solution to $\hat{P}(x^k)$, then there exists a KKT-multiplier $\lambda^k = \lambda(x^k) \in \mathbb{R}^m_+$ such that KKT optimality condition holds. Using the condition (iii) of the KKT condition, we have $\lambda^k_j \left(\langle \nabla f_j(x^k), d^k\rangle + \frac{1}{2}\langle d^k, \nabla^2 f_j(x^k) d^k\rangle - t^k\right) = 0$ for all $j \in \Lambda_m$. Summing over all $j \in \Lambda_m$ and using i), we have

$$\sum_{j\in\Lambda_m} \lambda^k_j \langle \nabla f_j(x^k), d^k\rangle + \sum_{j\in\Lambda_m} \frac{1}{2}\lambda^k_j \langle d^k, \nabla^2 f_j(x^k) d^k\rangle = t^k. \tag{3.3}$$

From ii), we have $\frac{1}{2}(\nabla h(x^k + d^k) - \nabla h(x^k)) + \sum_{j\in\Lambda_m} \lambda^k_j (\nabla f_j(x^k) + \nabla^2 f_j(x^k) d^k) = 0$. Multiplying both sides of the above equation by d^k and using the mean-value theorem, we have

$$\frac{1}{2}\langle d^k, \nabla^2 h(x^k + \eta_k d^k) d^k\rangle + \sum_{j\in\Lambda_m} \lambda^k_j \langle \nabla f_j(x^k), d^k\rangle + \sum_{j\in\Lambda_m} \lambda^k_j \langle d^k, \nabla^2 f_j(x^k) d^k\rangle = 0.$$

where $\eta_k \in (0,1)$. Using (3.3), we have

$$t^k + \frac{1}{2}\langle d^k, \nabla^2 h(x^k + \eta_k d^k) d^k\rangle + \frac{1}{2}\sum_{j\in\Lambda_m} \lambda^k_j \langle d^k, \nabla^2 f_j(x^k) d^k\rangle = 0. \tag{3.4}$$

Since h is strictly convex, we see that $\langle d^k, \nabla^2 h(x^k + \eta_k d^k) d^k\rangle \geq 0$. It follows from (3.4) that

$$t^k + \frac{1}{2}\sum_{j\in\Lambda_m} \lambda^k_j \langle d^k, \nabla^2 f_j(x^k) d^k\rangle \leq 0. \tag{3.5}$$

Since each f_j is strongly convex, we have $\langle d^k, \nabla^2 f_j(x^k) d^k\rangle \geq \rho_j ||d^k||^2$. If $\rho = \min\{\rho_1, \rho_2, ..., \rho_m\}$, we have $\langle d^k, \nabla^2 f_j(x^k) d^k\rangle \geq \rho ||d^k||^2$ for all $j \in \Lambda_m$. Thus, $\frac{1}{2}\sum \lambda^k_j \langle d^k, \nabla^2 f_j(x^k) d^k\rangle \geq \frac{\rho}{2}||d^k||^2$. The last step was achieved using i) in the KKT conditions. From (3.5) we have $t^k \leq -\frac{\rho}{2}||d^k||^2$. Hence, if $||d^k|| \neq 0$, we have $t^k < 0$. We now show that if $||d^k|| \neq 0$, then, for any $j \in \Lambda_m$,

$$f_j(x^k + \alpha d^k) = f_j(x^k) + \alpha \langle \nabla f_j(x^k), d^k\rangle + O(\alpha^2).$$

Since (t^k, d^k) is feasible to $(\hat{P}(x^k))$, we have $\langle \nabla f_j(x^k), d^k\rangle + \frac{1}{2}\langle d^k, \nabla^2 f_j(x^k) d^k\rangle \leq t^k$. Since $\frac{1}{2}\langle d^k, \nabla^2 f_j(x^k) d^k\rangle \geq \frac{1}{2}\rho_j ||d^k||^2 > 0$, we see that $\langle \nabla f_j(x^k), d^k\rangle < t^k$. As $\alpha > 0$, we have, for all $j \in \Lambda_m$, $f_j(x^k + \alpha d^k) < f_j(x^k) + \alpha t^k + O(\alpha^2)$. Thus, for $\beta \in (0,1)$,

$$\frac{f_j(x^k + \alpha d^k) - f_j(x^k) - \alpha\beta t^k}{\alpha} < t^k(1-\beta) + \frac{O(\alpha^2)}{\alpha}.$$

Since $\beta \in (0,1)$ and $t^k < 0$, we have, for α sufficiently small, $t^k(1-\beta) + \frac{O(\alpha^2)}{\alpha} < 0$. Thus,

$$f_j(x^k + \alpha d^k) < f_j(x^k) + \alpha\beta t^k.$$

Hence we can find an α which satisfies the Armijo criterion. This completes the proof. □

Theorem 3.4. *Let $\{x^k\}$ be a sequence of iterates generated by Algorithm 1. Let assumptions A_1 and A_2 hold. Then $\lim_{k\to\infty} d^k = 0$.*

Proof. The first step is to show that there exists $\bar{\alpha} > 0$ such that $\alpha_k \geq \bar{\alpha}$ for all $k \in \mathbb{N}$, where α_k is the step-length at the k-th iteration which guarantees that the Armijo criterion holds. From Step 4 of Algorithm 1, we have either $\alpha_k = 1$ or there exists a $k_1 \in \mathbb{N}$ such that $\alpha_k = r^{k_1}$. In the latter case, there exists at least one $j \in \Lambda_m$ such that

$$f_j(x^k + r^{k_1-1}d^k) > f_j(x^k) + r^{k_1-1}\beta t^k. \tag{3.6}$$

Using the Lipschitz continuity of ∇f_j, we have

$$f_j(x^k + r^{k_1-1}d^k) \leq f_j(x^k) + r^{k_1-1}\langle \nabla f_j(x^k), d^k\rangle + \frac{1}{2}(r^{k_1-1})^2 L \parallel d^k \parallel^2 .$$

Note that $\langle \nabla f_j(x^k), d^k\rangle < t^k$. It follows from (3.6) that

$$f_j(x^k) + r^{k_1-1}\beta t^k < f_j(x^k) + r^{k_1-1}t^k + \frac{1}{2}(r^{k_1-1})^2 L \parallel d^k \parallel^2 .$$

Thus $(1-\beta)(-t^k) < \frac{1}{2}Lr^{k_1-1} \parallel d^k \parallel^2$. Since $-t^k \geq \frac{\rho}{2} \parallel d^k \parallel^2$ by Theorem 3.3, we obtain $\frac{\rho}{2}(1-\beta) \parallel d^k \parallel^2 < \frac{1}{2}Lr^{k_1-1} \parallel d^k \parallel^2$, that is, $r^{k_1} > \frac{1}{L}(1-\beta)\rho r$. Hence, setting $\bar{\alpha} = \min\{1, \frac{1}{L}(1-\beta)\rho r\}$, we have $\alpha_k \geq \bar{\alpha}$ for all $k \in \mathbb{N}$. If $d^k \neq 0$, we have $t^k < 0$. In that case, the Armijo rule implies that $f_j(x^{k+1}) < f_j(x^k)$, for each $k \in \mathbb{N}$, and this shows that the sequence $\{x^k\}$ generated by Algorithm 1 is in the level set M and thus is bounded. Hence there exists a subsequence $\{x^{k_r}\}$ of $\{x^k\}$, which converges to x^* as $r \to \infty$. For any $N \in \mathbb{N}$, we see that, for any $j \in \Lambda_m$,

$$f_j(x^N) - f_j(x^0) \leq \beta \sum_{k=0}^{N} \alpha_k t^k \leq -\frac{1}{2}\rho\beta \sum_{k=0}^{N} \alpha_k \parallel d^k \parallel^2 \leq -\frac{1}{2}\bar{\alpha}\rho\beta \sum_{k=0}^{N} \parallel d^k \parallel^2 .$$

Thus, for any x^{k_r} in the subsequence, $f_j(x^{k_r}) - f_j(x^0) \leq -\frac{1}{2}\bar{\alpha}\rho\beta \sum_{k=0}^{k_r} \parallel d^k \parallel^2$ and $\frac{1}{2}\bar{\alpha}\rho\beta \sum_{k=0}^{k_r} \parallel d^k \parallel^2 \leq f_j(x^0) - f_j(x^{k_r})$. As $r \to \infty$, we have $\frac{1}{2}\bar{\alpha}\rho\beta \sum_{k=0}^{\infty} \parallel d^k \parallel^2 \leq f_j(x^0) - f_j(x^*) < +\infty$. Thus $\sum_{k=0}^{\infty} \parallel d^k \parallel^2 < +\infty$. This shows that $\parallel d^k \parallel \to 0$ as $k \to \infty$ and this establishes the result. □

Theorem 3.5. *Let assumptions A1,A2 and A3 hold and $\{x^k\}$ be a convergent sequence generated by Algorithm 1. Then $\{x^k\}$ is bounded and any limit point x^* of $\{x^k\}$ is a weak Pareto minimizer of (S-MOP). In fact, such a limit point is actually a Pareto minimizer.*

Proof. Note that $\{x^k\}$ is bounded (see the proof of Theorem 3.4). Further, for each $k \in \mathbb{N}$, assume that (t^k, d^k) solves $\hat{P}(x^k)$ with multipliers $\lambda^k \geq 0$ that satisfies the KKT conditions i), ii), and iii). Now, we consider the condition i) of the KKT conditions. For each $k \in \mathbb{N}$, $\sum_{j\in\Lambda_m} \lambda_j^k = 1$. Thus, for each $j \in \Lambda_m$, $\{\lambda_j^k\}$ is a bounded sequence. Without loss of generality, we assume that $\lim_{k\to\infty} \lambda_j^k = \lambda_j^*$, which shows that $\sum_{j\in\Lambda_m} \lambda_j^* = 1$. We also see from (ii) of the KKT conditions that

$$\frac{1}{2}(\nabla h(x^k + d^k) - \nabla h(x^k)) + \sum_{j\in\Lambda_m} \lambda_j^k (\nabla f_j(x^k) + \nabla^2 f_j(x^k)d^k) = 0.$$

As x^k is bounded, we see that there exists a convergent subsequence which converges to x^*. In fact, to avoid relabelling, let us assume, without loss of generality, that $x^k \to x^*$. By Assumption A3, let $\lim_{k\to\infty}(\nabla h(x^k + d^k) - \nabla h(x^k)) = w \in N_{\bar{S}}(x^*)$. As $k \to \infty$, $d^k \to 0$ by Theorem 3.4. We obtain from ii) that $\frac{1}{2}w + \sum_{j\in\Lambda_m} \lambda_j^* \nabla f_j(x^*) = 0$. Thus $-\sum_{j\in\Lambda_m} \lambda_j^* \nabla f_j(x^*) \in N_{\bar{S}}(x^*)$, which is necessary and sufficient for x^* to be a weak Pareto minimizer of (S-MOP). Thus we conclude that x^* is indeed the weak Pareto minimizer of (S-MOP). Further, for (S-MOP), we see that

every weak Pareto minimizer of (S-MOP) is also a Pareto minimizer. This completes the result. □

It is important to see the role played by the assumption A3. This is the crucial condition which allows us to factor in the fact that x^* can be in the boundary of S, and we cannot say anything about the continuity of ∇h on the boundary of S. However, if $x^* \in S$, then the above condition shows that $\lim_{r\to\infty}(\nabla h(x^{k_r}+d^{k_r})-\nabla h(x^{k_r}))=0$.

Under the additional assumption that $x^* \in S$, we have the following corollary.

Corollary 3.1. *Let the sequence $\{x^k\}$ be generated by the Algorithm 1. Assume that the assumptions A1 and A2 hold. If $\{x^k\}$ has limit points and among them at least one limit point is in S, then the limit point is a weak Pareto minimizer of (S-MOP).*

4. The Convex Case

In the previous section, we have shown that if we have a multiobjective optimization problem, with strongly convex objective and constrained by a convex set which is the closure of the zone of a Bregman function, then an effective algorithm can be developed by regularizing the Newton's method for multiobjective problems with a Bregman distance.

Now one question naturally arises: Can one modify the algorithm developed for the strongly convex case to the convex case? In fact, some basic changes can even allow us to use the results of the previous section in this case. We now consider the problem (MOP)

$$\min\left(f_1(x),\ldots,f_m(x)\right),$$
$$x\in\bar{S},$$

where each f_j, $j\in\Lambda_m$ is a convex function and $\bar{S}$ is the closure of the open convex set S, which is the zone of a Bregman function. Following the approach of the previous section, we begin with the quadratic approximation, $\bar{q}_j(x,d)$, for each $j\in\Lambda_m$ which is given as

$$\bar{q}_j(x,d)=\left\langle\nabla f_j(x),d\right\rangle+\frac{1}{2}\left\langle d,\nabla^2 f_j(x)d\right\rangle.$$

It is clear that $\bar{q}_j(x,d)$ is not necessarily a strongly convex function in d for a given x, though $\bar{q}_j(x,d)$ is a convex function. Our aim is now to regularize $\bar{q}_j(x,d)$ in such a way that the regularized function $\tilde{q}_j(x,d)$ becomes a strongly convex function. Construct the following regularized version of $\bar{q}_j(x,\cdot)$

$$\tilde{q}_j(x,d)=\bar{q}_j(x,d)+\frac{\mu_j}{2}\|d\|^2,\quad \mu_j\geq 0.$$

It is easy to see that, for $\mu_j>0$, $\tilde{q}_j(x,d)$ is a strongly convex function for each $j\in\Lambda_m$, with μ_j as the modulus of strong convexity. However, it might be interesting to choose μ_j's in such a way that all $\tilde{q}_j$'s have the same modulus of strong convexity. The following proposition proposes a way to do it.

Proposition 4.1. *Let $\tilde{q}_j(x,d)$ be as defined above, and let $\sigma>0$ be a given number. Let μ_j, $j\in\Lambda_m$ satisfy either of the following two conditions:*

(1) $\min\{\mu_1,\mu_2,\ldots,\mu_m\}\geq\sigma$

(2) $\mu_j=\begin{cases}0, & \text{if } \lambda_{\min}\left(\nabla^2 f_j(x)\right)\geq\sigma\\ \sigma-\lambda_{\min}\left(\nabla^2 f_j(x)\right), & \text{if } \lambda_{\min}\left(\nabla^2 f_j(x)\right)<\sigma,\end{cases}$

where $\lambda_{\min}(\nabla^2 f_j(x))$ is the minimum eigenvalue of the Hessian of f_j evaluated at x. Then, for any $j \in \Lambda_m$, $\tilde{q}_j(x,d)$ is a strongly convex function with modulus $\sigma > 0$.

Proof. Observe that $\nabla_d^2 \tilde{q}_j(x,d) = \nabla^2 f_j(x) + \mu_j I$, where I is the $n \times n$ identity matrix and ∇_d^2 symbolizes that the Hessian is computed with respect to d. Note that the Hessian of $\tilde{q}_j(x,d)$ is independent of d. For the given x, $\nabla_d^2 \tilde{q}_j(x,d)$ has the same value for any d. Further, for any $w \in \mathbb{R}^n$, $\langle w, \nabla_d^2 \tilde{q}_j(x,d)w\rangle = \langle w, \nabla^2 f_j(x)w\rangle + \mu_j\|w\|^2$ for any $d \in \mathbb{R}^n$. Thus

$$\langle w, \nabla_d^2 \tilde{q}_j(x,d)w\rangle \geq \langle w, \nabla^2 f_j(x)w\rangle + \sigma\|w\|^2 \tag{B}$$

if condition (1) holds. Observe that the Hessian matrix of f_j is positive semidefinite. It follows that

$$\langle w, \nabla_d^2 \tilde{q}_j(x,d)w\rangle \geq \sigma\|w\|^2. \tag{A}$$

As $\tilde{q}_j(x,\cdot)$ is convex for any $j \in \Lambda_m$, we see that

$$\tilde{q}_j(x,d) = \tilde{q}_j(x,d_0) + \langle \nabla_d \tilde{q}_j(x,d_0), d-d_0\rangle + \frac{1}{2}\langle d-d_0, \nabla_d^2 \tilde{q}_j(x,\bar{d})(d-d_0)\rangle,$$

where $\bar{d} \in (d,d_0)$. From (A), we have $\tilde{q}_j(x,d) \geq \tilde{q}_j(x,d_0) + \langle \nabla_d \tilde{q}_j(x,d_0), d-d_0\rangle + \frac{1}{2}\sigma\|d-d_0\|^2$. Since d and d_0 are arbitrarily chosen, we conclude that $\tilde{q}_j(x,\cdot)$ is strongly convex with modulus $\sigma > 0$. Let us now assume that condition (2) holds. Recall that $\langle w, \nabla_d^2 \tilde{q}_j(x,d)w\rangle = \langle w, \nabla^2 f_j(x)w\rangle + \mu_j\|w\|^2$. Let $\lambda_{\min}(\nabla^2 f_j(x)) \geq \sigma > 0$. In this case $\mu_j = 0$. Hence $\nabla^2 f_j(x)$ is a positive definite matrix. Thus

$$\langle w, \nabla^2 f_j(x)w\rangle \geq \lambda_{\min}(\nabla^2 f_j(x))\|w\|^2 \geq \sigma\|w\|^2.$$

Hence, $\langle w, \nabla_d^2 \tilde{q}_j(x,d)w\rangle \geq \sigma\|w\|^2$. This implies that $\tilde{q}_j(x,\cdot)$ is a strongly convex function in d, with modulus $\sigma > 0$. If $\lambda_{\min}(\nabla^2 f_j(x)) < \sigma$, then $\mu_j = \sigma - \lambda_{\min}(\nabla^2 f_j(x))$. Thus

$$\begin{aligned}\langle w, \nabla_d^2 \tilde{q}_j(x,d)w\rangle &= \langle w, \nabla^2 f_j(x)w\rangle + \mu_j\|w\|^2 \\ &= \langle w, \nabla^2 f_j(x)w\rangle + \left(\sigma - \lambda_{\min}(\nabla^2 f_j(x))\right)\|w\|^2.\end{aligned}$$

Note that $\lambda_{\min}(\nabla^2 f_j(x)) = \min_{\|w\|\neq 0} \frac{\langle w, \nabla^2 f_j(x)w\rangle}{\|w\|^2}$. Thus, for $w \neq 0$,

$$\langle w, \nabla^2 f_j(x)w\rangle \geq \lambda_{\min}(\nabla^2 f_j(x))\|w\|^2.$$

It follows that

$$\langle w, \nabla_d^2 \tilde{q}_j(x,d)w\rangle \geq \lambda_{\min}(\nabla^2 f_j(x))\|w\|^2 + \sigma\|w\|^2 - \lambda_{\min}(\nabla^2 f_j(x))\|w\|^2 = \sigma\|w\|^2.$$

Thus we conclude that $\tilde{q}_j(x,\cdot)$ is strongly convex with modulus $\sigma > 0$. This completes the proof. □

In view of $\tilde{q}_j(x,d)$, we now construct $\tilde{q}_v(x,d) = \max_{j\in\Lambda_m} \tilde{q}_j(x,d)$. We now further regularize $\tilde{q}_v(x,d)$ via a Bregman function. Consider the problem: $(\tilde{P}(x))$

$$\min_{d\in D(x)} \tilde{q}_v(x,d) + \frac{1}{2}D_h(x+d,x).$$

Since $\tilde{q}_v(x,d)$ is strongly convex, and $D(x)$ is a closed convex set, then problem $(\tilde{P}(x))$ has a unique solution which we mark again as $d(x)$, assuming that no confusion will arise.

Proposition 4.2. *Let $d(x)$ be a unique solution of $(\tilde{P}(x))$. Then $x+d(x) \in S$ and $d(x) \in$ int $D(x)$.*

This result also plays an important role in the convex case. In fact, we can consider an equivalent version of $(\tilde{P}(x))$ by considering $(\tilde{P}_0(x))$:

$$\min_{(t,d)\in\mathbb{R}\times\mathbb{R}^n} \quad t+\frac{1}{2}D_h(x+d,x)$$

$$\text{subject to } d\in D(x)$$

$$\left\langle \nabla f_j(x),d\right\rangle+\frac{1}{2}\left\langle d,(\nabla^2 f_j(x)+\mu_j I)d\right\rangle\leq t,\quad \forall j.$$

If $(t(x),d(x))$ is a solution to the above problem, then the KKT condition holds. Hence, there exist multipliers $\lambda_j(x)\geq 0$, $j\in\Lambda_m$, such that

a): $\sum_{j\in\Lambda_m}\lambda_j(x)=1$;
b): $\sum_{j\in\Lambda_m}\lambda_j(x)\left[\nabla f_j(x)+(\nabla^2 f_j(x)+\mu_j I)d(x)\right]+\frac{1}{2}(\nabla h(x+d(x))-\nabla h(x))=0$;
c): $\lambda_j(x)\left(\left\langle \nabla f_j(x),d(x)\right\rangle+\frac{1}{2}\left\langle d(x),(\nabla^2 f_j(x)+\mu_j I)d(x)\right\rangle-t(x)\right)=0$.

Since $d\in\operatorname{int}D(x)$, we have $N_{D(x)}(d(x))=\{0\}$. Consider $h(x)=\frac{1}{2}\left\langle x,Qx\right\rangle$, where Q is a positive definite matrix. Thus $\nabla h(x+d)=Q(x+d)$. Using (b), we can show that

$$d(x)=-\left(\frac{1}{2}Q+\sum_{j\in\Lambda_m}\lambda_j(x)(\nabla^2 f_j(x)+\mu_j I)\right)^{-1}\sum_{j\in\Lambda_m}\lambda_j(x)\nabla f_j(x).$$

If $\mu_j=0$ for $j\in\Lambda_m$, then we obtain $d(x)$ for the case where each f_j is strongly convex. Observe that, in computing $d(x)$ even in this simple case, we rely on the Lagrange multipliers of $(\tilde{P}_0(x))$. In fact Lagrangian multipliers here play the role of scalarizing weights. This is a key difference between the scalar case and the vector case.

The algorithm for (MOP) remains the same as that of (S-MOP), with the only change that we solve $(\tilde{P}_0(x))$ instead of $(\hat{P}(x))$. In fact, all the convergence results for (S-MOP) remains valid even in this case under similar conditions. So, we do not repeat them here.

5. Numerical Experiments

We now present some numerical experiments to show the efficacy of our algorithm. There are two types of problems. The first problem is a bi-objective convex problem for which the Newton method is not a very robust method. However we show that our approach works efficiently. The second problem has both the objectives to be strongly convex. Even in this setting, we show that for certain starting points as the number of iterations increase Newton's method may fail to generate descent direction and thus we will not reach the Pareto minimizing point. However, from the same starting point, our algorithm reaches the solution. In both cases, we compare our algorithm with the popular method for solving multiobjective problems, the weighted sum method. Given a feasible set $\bar{S}$ which is represented by box constraints, we construct a strictly convex function h which satisfies all the properties of the Bregman function and has $S=\operatorname{int}\bar{S}$ as its zone. h is constructed by using the data of the feasible set and is boundary coercive. For $\bar{S}=\{x\in\mathbb{R}^n | lb\leq x\leq ub\}$, $lb<ub$, we consider the following Bregman function from [7]

$$h(x):=\sum_{i=1}^{n}(x_i-lb_i)\log(x_i-lb_i)+\sum_{i=1}^{n}(ub_i-x_i)\log(ub_i-x_i). \tag{5.1}$$

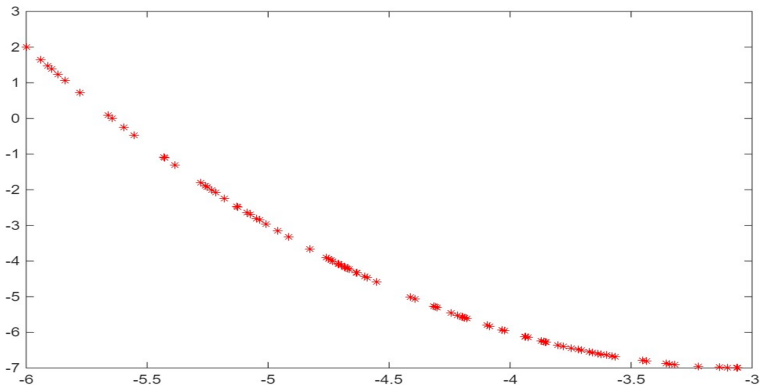


FIGURE 1. Approximate Pareto front of $(VU2)$

In this case,

$$D_h(x,y) := \sum_{i=1}^{n}(x_i - lb_i)\log\left(\frac{x_i - lb_i}{y_i - lb_i}\right) + \sum_{i=1}^{n}(ub_i - x_i)\log\left(\frac{ub_i - x_i}{ub_i - y_i}\right).$$

The Convex Case. Consider the following box-constrained test problem from [17]

$$(VU2): \min_{x\in\mathbb{R}^2} \left(x_1 + x_2, x_1^2 + 2x_2 - 1\right)$$
$$\text{s.t. } -3 \le x_1, x_2 \le 3.$$

Clearly, both objective functions are convex (the first one is actually affine), but not strictly convex. Hence, Newton method in [12] would not work here. We have solved this problem using our method with the above Bregman function. MATLAB (2025b) code is developed for Algorithm 1 adapted to the convex case with Bregman function h.

Consider $x^0 = (1.8883, 2.4348)^T$. Then $f(x^0) = (4.3231, 7.4353)^T$. The sub-problem $P(x^k)$ is solved by using *CVX* solver *SeDuMi*. Solution of $P(x^0)$ is obtained as $t^0 = -4.6919$, $d^0 = (-2.6498, -2.0421)^T$. Further, $\alpha_0 = 1$ satisfies the Armijo type condition given by (3.2). Hence $x^1 = x^0 + \alpha_0 d^0 = (-0.7614, 0.3927)^T$ is the next iterative point. Clearly,

$$f(x^1) = (-0.3687, 0.3651)^T < f(x^0).$$

The final solution is obtained as $x^* = (-1.7324, -3.0000)^T$ after the 7th iteration.

Next, we show that $x^* = (-1.7324, -3.0000)^T$ is a critical point. Clearly, $g(x) = -3 - x_2 \le 0$ is an active box constraint at x^* (other box constraints are ignored as they are inactive). Hence, from [19, Theorem 3.1.1], there exists $\lambda_1^*, \lambda_2^*, \mu^* \ge 0$, $(\lambda_1, \lambda_2) \ne (0,0)$ such that $\lambda_1^* \nabla f_1(x^*) + \lambda_2^* \nabla f_2(x^*) + \mu^* \nabla g(x^*) = 0$, which implies $\lambda_1^*(1,1)^T + \lambda_2^*(-3.4648, 2)^T + \mu^*(0,-1)^T = (0,0)^T$. Clearly $\lambda_1^* = 3.4648$, $\lambda_2^* = 1$, and $\mu^* = 5.4648$ satisfy these equations. Hence x^* is a critical point of $(VU2)$. Solution of a multi-objective optimization problem does not consist of isolated optimum points but a set of efficient solutions (X^*). We use the multi-start technique to generate a well-distributed approximation of X^*. In this process, a set of 100 uniformly distributed initial approximations is considered. Algorithm 1 is executed individually. The set of non-dominated points of the approximation solutions collection is considered as an approximate efficient solution. We generated approximate Pareto fronts of this problem and provided them in Figure 1.

Strongly convex function. Consider the two objective optimization problem

$$\min_{x\in\mathbb{R}^2} (f_1(x), f_2(x)),$$

where

$$\begin{aligned} f_1(x) &= \frac{1}{2}\left(x_1^2+x_2^2\right)+\frac{1}{4}\left(x_1^4+x_2^4\right), \\ f_2(x) &= \frac{1}{2}\left((x_1-2)^2+(x_2-2)^2\right)+\frac{1}{4}\left((x_1-2)^4+(x_2-2)^4\right). \end{aligned}$$

Clearly f_1 is strictly convex since

$$\nabla^2 f_1(x) = \begin{bmatrix} 1+3x_1^2 & 0 \\ 0 & 1+3x_2^2 \end{bmatrix}$$

has eigenvalues $1+3x_1^2, 1+3x_2^2 \geq 1$. Since these eigenvalues are at least 1, f_1 is also strongly convex. Similarly f_2 is strongly convex since

$$\nabla^2 f_2(x) = \begin{bmatrix} 1+3(x_1-2)^2 & 0 \\ 0 & 1+3(x_2-2)^2 \end{bmatrix}$$

has eigenvalues $1+3(x_1-2)^2, 1+3(x_2-2)^2 \geq 1$. For $x^0 = (150,150)^T$, both $\nabla^2 f_1(x^0)^{-1}$ and $\nabla^2 f_2(x^0)^{-1}$ become nearly singular. Hence Newton method fails to generate a descent direction. However, in our method, if we use Bregman function (5.1) with $lb = (-10,-10)^T$ and $ub = (160,160)^T$ ($ub > x^0$), then $x^* = (1.999, 1.9999)^T \approx (2,2)$. One can verify that $x^* = (2,2)^T$ is an efficient solution of this problem by using the weighted sum method with weight $(0,1)$.

REFERENCES

[1] M.A.T. Ansary, G. Panda, A sequential quadratically constrained quadratic programming technique for a multi-objective optimization problem, Eng. Optim. 51 (2019), 22–41.

[2] M.A.T. Ansary, G. Panda, A globally convergent SQCQP method for multi-objective optimization problems, SIAM J. Optim. 31 (2021), 91—113.

[3] M.A.T. Ansary, G. Panda, A modified quasi-Newton method for vector optimization problem, Optimization 64 (11), 2289–2306.

[4] M.A.T. Ansary, G. Panda, A sequential quadratic programming method for constrained multi-objective optimization problems, J. Appl. Math. Comput. 64 (2020), 379-–397.

[5] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, Nonlinear Programming: Theory and Algorithms, Wiley, 2006.

[6] A. Beck, First-order Methods in Optimization, SIAM, 2017.

[7] R.S. Burachik, A.N. Iusem, A generalized proximal point algorithm for the variational inequality problem in a Hilbert space, SIAM J. Optim. 8 (1998), 197–216.

[8] G. Chen, M. Teboulle, Convergence analysis of a proximal-like minimization algorithm using bregman functions, SIAM J. Optim. 3 (1993), 538–543.

[9] K. Deb, Multi-Objective Optimization Using Evolutionary Algorithms, Wiely, New Delhi, 2003.

[10] K. Deb, S. Agrawal, A., Pratap, T. Meyarivan, A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II. In: M. Schoenauer, et al. (ed.) Parallel Problem Solving from Nature PPSN VI. PPSN 2000. Lecture Notes in Computer Science, vol 1917, Springer, Berlin, Heidelberg, 2000.

[11] N. Doikov, Y. Nesterov, Gradient regularization of Newton method with Bregman distances, Math. Program. A, 204 (2024), 1-25.

[12] J. Fliege, L.M.G. Drummond, B.F. Svaiter, Newton's method for multiobjective optimization, SIAM J. Optim. 20 (2009), 602–626.

[13] J. Fliege, B.F. Svaiter, Steepest descent methods for multicriteria optimization, Math. Methods Oper. Res. 51 (2000), 479–494.
[14] J. Fliege, A.I.F. Vaz, A method for constrained multiobjective optimization based on SQP techniques, SIAM J. Optim. 26 (2016), 2091–2119.
[15] D. Ghosh, N. Kishor, X. Zhao, A Newton method for uncertain multiobjective optimization problems with finite uncertainty sets, J. Nonlinear Var. Anal. 9 (2025), 81-110.
[16] N. Hayek, A note on multiobjective optimal control in presence of asymptotical constraints, Commun. Optim. Theory 2025 (2025) 18.
[17] S. Huband, P. Hingston, L. Barone, L. While, A review of multiobjective test problems and a scalable test problem toolkit, IEEE Trans. Evol. Comput. 10 (2006), 477–506.
[18] D.T. Luc, Theory of Vector Optimization, Springer, 1989.
[19] K. Miettinen, Nonlinear Multiobjective Optimization, Kluwer, Boston, 1999.
[20] Ž. Povalej, Quasi-Newton's method for multiobjective optimization, J. Comput. Appl. Math. 255 (2014), 765–777.
[21] S. Qu, M. Goh, F.T.S. Chan, Quasi-Newton methods for solving multiobjective optimization, Oper. Res. Lett. 39 (2011), 397–399.
[22] N.M. Tung, N.M. Cuong, Second-order optimality conditions for nonsmooth multiobjective optimization subjected to mixed constraints, Appl. Set-Valued Anal. Optim. 7 (2025), 39–55.