Noname manuscript No.
(will be inserted by the editor)

A Proximal Gradient Method for Multi-objective Optimization Problems Using Bregman Functions

Md Abu Talhamainuddin Ansary ·
Joydeep Dutta

Received: date / Accepted: date

Abstract In this paper, a globally convergent proximal gradient method is developed for convex multi-objective optimization problems using Bregman distance. The proposed method is free from any kind of a priori chosen parameters or ordering information of objective functions. At every iteration of the proposed method, a subproblem is solved to find a descent direction. This subproblem uses a Bregman distance induced by a strongly convex function. An Armijo type line search is conducted to find a suitable step length. A sequence is generated using the descent direction and step length. It is justified that this sequence converges to a weak efficient solution under some mild assumptions. The method is verified and compared with one existing method using a set of test problems.

Keywords convex optimization · multi-objective optimization · proximal gradient method · critical point · Bregman functions

Mathematics Subject Classification (2010) 90C25 · 90C29 · 49M37 · 97N40

1 Introduction

In a multi-objective optimization problem, several objective functions are minimized simultaneously. If any feasible solution minimizes all objective functions,

Md Abu Talhamainuddin Ansary · Joydeep Dutta
Department of Economic Science,
Indian Institute of Technology Kanpur,
India-208016

Md Abu Talhamainuddin Ansary (✉)
E-mail: md.abutalha2009@gmail.com

Joydeep Dutta
E-mail: jdutta@iitk.ac.in

then this is called an ideal solution. But quite often decrease of one objective function causes an increase in other objective functions. So the concept of optimality is replaced by efficiency. Classical methods of solving multi-objective optimization problems are scalarization methods (see [14,24]), which reduce the original problem to a single objectiveoptimization problem using a set of priori chosen parameters. These methods are user dependent and often fail to generate Pareto front. Heuristic methods like evolutionary algorithms (see [15,16]), are often used to find approximate Pareto front but can not guarantee any convergence property. Recently many researchers have developed new techniques for nonlinear multi-objective optimization problems (see [18, 17,2,1,3,30,9,27]) which does not use any priori chosen parameter or ordering information of objective functions. The techniques developed in [18,17,2, 1,3] are possible extension of gradient based techniques for single objective optimization to multi-objective case. These techniques are restricted to continuously differentiable multi-objective problems. Apart from gradient based techniques, recently several methods for solving non-differentiable single objective optimization problems are extended to multi-objective optimization problems by several researchers. This extensions include proximal point methods ([8,9]), sub-gradient methods ([7,25,27]), proximal gradient methods ([30]) for non-differentiable multi-objective optimization problems.

Proximal gradient methods are widely used to solve single objective optimization problems where objective function can be represented as sum of a convex and continuously differentiable function and a proper convex and lower semicontinuous but not necessarily a differentiable function (see [4,6,5,26]). The idea of proximal method in single objective optimization problems is extended in different directions. In state of using Euclidian distances, Bregman distance induced by a strongly convex function([11]) is used in [13,12,20].

Recently Tanabe et al. ([30]) have developed proximal gradient method for multi-objective optimizations. This method combines the ideas of steepest descent method and proximal point method for multi-objective optimization problems developed in [18] and [9] respectively. At every iteration of the proximal gradient method in [30], a subproblem is solved to find a descent direction which involves Euclidian distance. In this paper, we have used the ideas of [13, 20,21,9] and modified the ideas in [30]. Our key idea here is to replace the Euclidian distance induced by the Euclidean norm by a non-Euclidean distance called *Bregman distance* induced by a strictly convex function.

The paper is organized as follows. Some prerequisites are discussed in Section 2. A proximal gradient method using Bregman distance is developed in Section 3. An algorithm is proposed and convergence of this algorithm is established in Section 4. In Section 5, the proposed method is compared with existing method using a set of test problems.

The notations used in this paper is fairly standard. For any given vectors x and y we shall interchangeably denote by $\langle x, y\rangle$ to mean the inner product. For a given set S we shall denote by $\overline{S}$ the closure of S and by $intS$ the interior of S.

2 Preliminaries

Consider the multi-objective optimization problem:

$$(MOP): \quad \min_{x\in\mathbb{R}^n} \ F(x) = (F_1(x), F_2(x), ..., F_m(x)),$$

where S is a nonempty open convex set. Suppose $F_j : \mathbb{R}^n \to \bar{\mathbb{R}}$ is defined by $F_j(x) = f_j(x) + g_j(x)$ where f_j is convex and continuously differentiable and g_j is proper convex and lower semi-continuous (l.s.c.) but not necessary differentiable function, for $j = 1, 2, ..., m$. We will assume that for each g_j the interior of the its domain denoted as $int(dom\ g_j)$ is non-empty. Denote $\Lambda_m = \{1, 2, ..., n\}$ denote the index set of the objective functions. Inequality in $\mathbb{R}^m$ is understood component wise. If there exists $x \in \mathbb{R}^n$ such that x minimizes all objective functions simultaneously then it is an ideal solution. But in practice, decrease of one objective function may cause increase of another objective function. So the theory of multi-objective optimization *optimality* is replaced by *efficiency*. In order to define the notion of efficiency we need the notion of a partial order on the space $\mathbb{R}^m$. We shall use for our discussions the natural ordering between elements of $\mathbb{R}^m$, which is done in the following way. For u and v in $\mathbb{R}^m$, we say $u \geq v$ if $u - v \in \mathbb{R}^m_+$ or $u_i \geq v_i$ for all $i = 1, \ldots, m$. Further we mark $u > v$ if $u - v \in \mathbb{R}^m_{++}$, or in other words $u_i > v_i$ for all $i = 1, \ldots, m$. A feasible point $x^* \in X$ is said to be an efficient solution of the (MOP) if there does not exist $x \in \mathbb{R}^n$ such that $F(x) \leq F(x^*)$ and $F(x) \neq F(x^*)$. A feasible point $x^* \in \mathbb{R}^n$ is said to be a weak efficient solution of the (MOP) if there does not exist $x \in \mathbb{R}^n$ such that $F(x) < F(x^*)$. It is clear that every efficient solution of (MOP) is a weak efficient solution, but the converse is not true. If each F_j, $j \in \Lambda_m$ are strictly convex then every weak efficient solution is an efficient solution. For $x, y \in \mathbb{R}^n$, we say y dominates x, if and only if $F(y) \leq F(x)$, $F(y) \neq F(x)$. If X^* is the set of all efficient solutions of the (MOP), then $F(X^*)$ is said to be the Pareto front of the (MOP) and it lies on the boundary of $F(\mathbb{R}^n)$.
Assume that $x^* \in \bar{S} \subseteq \bigcap\limits_{j\in\Lambda_m} int(dom\ g_j)$ and x^* is an weak efficient solution of the (MOP). Then,

$$\left(F_1^{'}(x^*; d), F_2^{'}(x^*; d), ..., F_m^{'}(x^*; d)\right) \notin -\mathbb{R}^m_{++},$$

for all $d \in \mathbb{R}^n$. This shows that

$$\max_{j\in\Lambda_m} F_j^{'}(x^*; d) \geq 0 \qquad \text{for all } d \in \mathbb{R}^n \tag{1}$$

The inequality in (1) is sometimes refereed to in the literature as the criticality condition or the first order necessary for (MOP) and x^* satisfying (1) is often called a critical point for the (MOP). Further convexity of each F_j, $j \in \Lambda_m$ ensures that every critical point of (MOP) is weak efficient solution. Further note that if each F_j, $j \in \Lambda_m$ is a strictly or strongly convex function then the critical point of (MOP) is an efficient solution. Note that if either f_j or g_j is

strictly or strongly convex then is so F_j.
We would like to note that in the problem (MOP) through we have considered that each g_j , $i \in \Lambda_m$, to be a proper lower-semicontinuous convex function but for most applications one usually has g_j to be a finite-valued convex function. Thus when we develop the algorithm and carry out a convergence analysis we shall consider the case where each g_j to be finite-valued. Let us consider how such a problem (MOP) can naturally arise where each g_j is finite-valued. In the their recent book on applied linear algebra, Boyd and Vendenberghe [10] considers the following multi-objective least square problems , i,e, where we seek to find the Pareto minimizer of the following multi-objective optimization problem

$$\min\left(\|A_1x - b_1\|^2, \ldots, \|A_mx - b_m\|^2\right),$$

where each A_j is a $k_j \times n$ matrix, $x \in \mathbb{R}^n$ and each $b_j \in \mathbb{R}^{k_j}$. Such a problem has been referred to as the multi-objective least square problem in [10]. In fact it was shown in [10] why such problems naturally arise. The standard single objective least square naturally arises in several applications and also in the important problem of image de-blurring. However in the problem of image de-blurring one seeks a sparse solution and in such cases a regularized version on the least square problem is solved by adding the a multiple of the l_1 norm of x to the main objective. For more details see the seminal paper of Beck and Teboulle [5]. This motivates us to consider a regularized version of the above multi-objective least square problem in order generate Pareto solutions which are sparse in nature. The regularized version can be given as

$$\min\left(\|A_1x - b_1\|^2 + \mu_1\|x\|_1, \ldots, \|A_mx - b_m\|^2 + \mu_m\|x\|_1\right),$$

where $\mu_j > 0$, $i = 1, \ldots m$ are the regularization parameters. Thus the regularized version of the multi-objective least square problem is in the form of (MOP) with $f_j(x) = \|A_jx - b_j\|^2$ and $g_j(x) = \mu_j\|x\|_1$. Note that here each g_i is finite-valued though not differentiable everywhere on $\mathbb{R}^n$.
Suppose S be a non-empty open convex subset of $\mathbb{R}^n$. Let h be a strictly convex function which is differentiable on S. Define the function $D_h : \bar{S} \times S \to \mathbb{R}$ as

$$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y\rangle. \tag{2}$$

The function h is called a Bregman function if it satisfies the properties given in the definition below

Definition 1 *(Definition 2.1, [13])*
Let $S \in \mathbb{R}^n$ be an open convex subset. Then $h : \bar{S} \to \mathbb{R}$ is called a **Bregman function** *with zone S if the following hold:*

(i) h is continuously differentiable on S.
(ii) h is strictly convex and continuous on $\bar{S}$.
(iii) For every $\alpha \in \mathbb{R}$ the partial level sets $L_1(y, \alpha) = \{x \in \bar{S} : D_h(x, y) \leq \alpha\}$ and $L_2(x, \alpha) = \{y \in S : D_h(x, y) \leq \alpha\}$ are bounded for every $y \in S$ and $x \in \bar{S}$.

(iv) *If* $\{y^k\} \in S$ *converges to* y^*, *then* $D_h(y^*, y^k) \to 0$.
(v) *If* $\{x^k\}$ *and* $\{y^k\}$ *are sequences in* S *such that* $y^k \to y^* \in \bar{S}$, $\{x^k\}$ *is bounded, and if* $D_h(x^k, y^k) \to 0$, *then* $x^k \to y^*$.

Clearly $D_h(x, y) \geq 0$ for every $(x, y) \in \bar{S} \times S$ and $D_h(x, y) = 0$ if and only if $x = y$. But $D_h(x, y)$ is not symmetric and triangular inequality does hold for D_h. In the literature the function D_h is often called the Bregman distance induced by the function h. For any $x \in \bar{S}$ and $y, z \in S$ following result (known as 'three points identity') is derived in [13].

Lemma 1 *(Lemma 3.1, [13]) Let* h *is a Bregman function. Then for any* $y, z \in S$ *and* $x \in \bar{S}$ *the following identity holds:*

$$D_h(x, y) - D_h(y, z) - D_h(x, z) = \langle \nabla h(z) - \nabla h(y), x - y \rangle. \tag{3}$$

Definition 2 *A Bregman function* h *is said to be boundary coercive if for any sequence* $\{y^k\}$ *in* S, $\lim_{k\to\infty} y^k = y \in bdS$ *holds then* $\lim_{k\to\infty} \langle \nabla h(y^k), x - y^k \rangle = -\infty$, *for all* $x \in S$.

If h is twice continuously differentiable on S, then by using Taylor expansion of second order for h,

$$D_h(x, y) = \frac{1}{2} \langle x - y, \nabla^2 h\left(y + \xi(x - y)\right)(x - y) \rangle$$

for some $\xi \in (0, 1)$. If h is a σ-strongly convex function on S, then
$D_h(x, y) \geq \frac{\sigma}{2} \|x - y\|^2$.

Lemma 2 *If* h *is* σ*-strongly convex on* $\bar{S}$ *and continuous on* $\bar{S}$, *and twice continuously differentiable on* S, *then for any* $x \in S$

$$\langle d, \nabla^2 h(x) d \rangle \geq \sigma \|d\|^2$$

for any finite d.

Definition 3 *([4]) Suppose* $F_0 : \mathbb{R}^n \to (-\infty, \infty]$ *be a proper convex function and* $x \in dom(F_0)$. *Then sub-differential of* F_0 *at* x *is denoted by* $\partial F_0(x)$ *and defined as*

$$\partial F_0(x) := \{\xi \in \mathbb{R}^n : F_0(y) \geq F_0(x) + \langle \xi, y - x \rangle \quad \forall\, y \in \mathbb{R}^n\}.$$

If $x \notin dom\ F_0$ *then we define* $\partial F_0(x) = \emptyset$.

It is simple to observe that $x_0 = arg\ \min_{x \in \mathbb{R}^n} F_0(x)$ if and only if $0 \in \partial F_0(x)$. Associated with the notion of the subdifferential of a convex function is the notion of the directional derivative. For the given convex function F_0 the directional derivative at $x \in \mathbb{R}^n$ in the direction $v \in \mathbb{R}^n$ is given as

$$F_0'(x, v) = \lim_{t \downarrow 0} \frac{F(x + td) - F(x)}{t}.$$

The directional exists and is finite when x is the interior of $domF_0$. Further the subdifferential and the directional derivative is linked through the following relation,

$$F_0'(x,d) = \max_{\xi \in \partial_0(x)} \langle \xi, d \rangle.$$

More details on the subdifferential and the directional derivative of a convex function see for example Rockafellar [29].

3 A proximal gradient method for (MOP) using Bregman function

In this section we shall develop the theoretical basis which will allow us to build a descent algorithm for (MOP) motivated by based on proximal gradient method for scalar convex optimization problems. Though our algorithm will be aimed for the case where the functions g_i are finite, our theoretical development will continue to consider the function g_i to be proper and lower-semicontinuous and we shall specifically mention that g_i is finite-valued when we shall need it in a particular result. One might feel that why not do the complete analysis for the case where g_i is assumed to be finite. However there are several crucial results regarding the nature of critical points for (MOP) which seem to hold even in the more general setting where each g_i is proper and lower-semicontinuous.
We shall begin by stating our main assumption which we will assume throughout this article.

$$\bar{S} \subset int \left(\bigcap_{j \in \Lambda_m} dom\ F_j \right) = \bigcap_{j \in \Lambda_m} (int\ dom\ F_j)\,.$$

For each $j \in \Lambda$ let us define the following a function $\Psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ as follows,

$$\boxed{\Psi_j(x,d) := \langle \nabla f_j(x), d \rangle + g_j(x+d) - g_j(x).} \tag{4}$$

Let us now introduce for any $x \in \mathbb{R}^n$, the function $\Psi_x : \mathbb{R}^n \to \mathbb{R}$ given as,

$$\boxed{\Psi_x(d) := \max_{j \in \Lambda_m} \Psi_j(x,d).} \tag{5}$$

Then Ψ_j is jointly continuous in x and d if $x, x+d \in int(dom(F_j))$. Hence Ψ_x is jointly continuous in x and d if $x, x+d \in int \left(\bigcap_{j \in \Lambda_m} dom(F_j) \right)$. Given $x, d \in \mathbb{R}^n$, denote by $J(x,d)$ the active index set associated with the function Ψ_x as follows,

$$J(x,d) := \{ j \in \Lambda_m | \Psi_x(d) = \Psi_j(x,d) \}.$$

Then from convexity of F_j, we have

$$F_j(x+d) - F_j(x) \geq \Psi_x(d)$$

holds for every $j \in J(x,d)$.

Lemma 3 *If $\Psi_x(d) \geq 0,\ \forall\ d \in \mathbb{R}^n$, then x is a weak efficient solution of (MOP).*

Proof: On the contrary if x is not a weak minimizer of (MOP) then there exists x', such that

$$F_j(x') < F_j(x) \quad \forall j \in \Lambda_m \tag{6}$$

By convexity of f_j, we have for all $j \in \Lambda_m$

$$\langle \nabla f_j(x), x' - x \rangle \leq f_j(x') - f_j(x)$$
$$\langle \nabla f_j(x), x' - x \rangle + g_j(x') - g_j(x) \leq f_j(x') - f_j(x) + g_j(x') - g_j(x).$$

Hence from (6),

$$\Psi_j(x, x' - x) \leq F_j(x') - F_j(x) < 0 \quad \forall j \in \Lambda_m.$$

This implies $\Psi_x(x'-x) < 0$. Thus we have shown the existence of a $d = x' - x$ which contradicts the hypothesis of the lemma, Hence the result. $\square$.

For any $x \in S$, define set $D(x) := \{d \in \mathbb{R}^n : x + d \in \bar{S}\}$. The set $D(x)$ is non-empty since $0 \in D(x)$. Note further that $D(x)$ is a closed set. Now for $\lambda > 0$ and $d \in D(x)$ set

$$\boxed{P_{\lambda,x}(d) := \Psi_x(d) + \frac{1}{\lambda} D_h(x+d, d).} \tag{7}$$

Lemma 4 *For any $x \in S$ and $\lambda > 0$, $d \mapsto P_{\lambda,x}(d)$ is lower-semicontinuous on $D(x)$. Further if $h : \bar{S} \to \mathbb{R}$ is strongly convex, then for each $x \in S$, the function*
$d \mapsto D_h(x+d, x)$ is strongly convex over $D(x)$. Thus

$$\underset{d \in D(x)}{argmin} \ P_{\lambda,x}$$

is a singleton set.

Proof: Suppose $\{d^k\}$ be any sequence in $D(x)$ such that $d^k \to \bar{d}$ as $k \to \infty$. Then for any $j \in \Lambda_m$,

$$\begin{aligned}
\liminf_{k\to\infty} \Psi_{j,x}(d^k) &= \liminf_{k\to\infty} \left\{ \langle \nabla f_j(x), d^k \rangle + g_j(x + d^k) - g_j(x) \right\} \\
&\geq \liminf_{k\to\infty} \{ \langle \nabla f_j(x), d^k \rangle - g_j(x) \} + \liminf_{k\to\infty} \ g_j(x + d^k) \\
&\geq \langle \nabla f_j(x), \bar{d} \rangle + g_j(x + \bar{d}) - g_j(x) \\
&= \Psi_{j,x}(\bar{d}).
\end{aligned}$$

Second inequality follows from lower semicontinuty of g. Hence $\Psi_{j,x}$ is l.s.c. over $D(x)$. Then Ψ_x is l.s.c. over $D(x)$. So
$\lim_{k\to\infty} \Psi_x(d^k) \geq \Psi_x(\bar{d})$.
From Definition 1(ii), D_h is continuous in $d \in D(x)$ for any fixed x. So $\liminf_{k\to\infty} D_h(x+d^k,x) = D_h(x+\bar{d},x)$. Now

$$\begin{aligned}\liminf_{k\to\infty} P_{\lambda,x}(d^k) &= \liminf_{k\to\infty}\{\Psi_x(d^k) + \frac{1}{\lambda}D_h(x+d^k,x)\}\\ &\geq \liminf_{k\to\infty} \Psi_x(d^k) + \frac{1}{\lambda} \liminf_{k\to\infty} D_h(x+d^k,x)\\ &\geq \Psi_x(\bar{d}) + +\frac{1}{\lambda}D_h(x+\bar{d},x)\\ &= P_{\lambda,x}(\bar{d}).\end{aligned}$$

Hence $P_{\lambda,x}$ is l.s.c. over $D(x)$.
Let h be strongly convex with $\sigma > 0$ as the modulus of strong convexity. Let d_1 and d_2 be two elements of $D(x)$ for a given $x \in S$. Now observe the following. For any $\lambda \in [0,1]$, we have,

$$\begin{aligned}D_h\left(x + (\lambda d_1 + (1-\lambda)d_2), x\right) &= D_h\left(\lambda(x+d_1) + (1-\lambda)(x+d_2), x\right)\\ &= h\left(\lambda(x+d_1) + (1-\lambda)(x+d_2)\right) - h(x)\\ &\quad -\langle \nabla h(x), \lambda d_1 + (1-\lambda)d_2\rangle.\end{aligned}$$

Now as h is strongly convex we have

$$\begin{aligned}h\left(\lambda(x+d_1) + (1-\lambda)(x+d_2)\right) &\leq \lambda h(x+d_1) + (1-\lambda)h(x+d_2)\\ &\quad -\frac{\sigma}{2}\lambda(1-\lambda)\|d_1-d_2\|^2.\end{aligned}$$

Hence

$$\begin{aligned}D_h\left(x + (\lambda d_1 + (1-\lambda)d_2), x\right) &\leq \lambda\left[h(x+d_1) - h(x) - \langle \nabla h(x), d_1\rangle\right]\\ &\quad +(1-\lambda)\left[h(x+d_2) - h(x) - \langle \nabla h(x), d_2\rangle\right]\\ &\quad -\frac{\sigma}{2}\lambda(1-\lambda)\|d_1-d_2\|^2.\end{aligned}$$

This implies

$$\begin{aligned}D_h\left(x + (\lambda d_1 + (1-\lambda)d_2), x\right) + \frac{\sigma}{2}\lambda(1-\lambda)\|d_1-d_2\|^2 &\leq \lambda D_h(x+d_1)\\ &\quad +(1-\lambda)D_h(x+d_2).\end{aligned}$$

This proves that $d \mapsto D_h(x+d,x)$ is strongly convex over $D(x)$. Further as Ψ_x is convex by definition we have $P_{\lambda,x}(d)$ to be strongly convex over $D(x)$. Now as $D(x)$ is a closed convex set and $P_{\lambda,x}$ is l.s.c. and strongly convex in $d \in D(x)$, there exists unique minimizer of $P_{\lambda,x}$ over $D(x)$. Hence the set,

$$\operatorname*{argmin}_{d\in D(x)} \; P_{\lambda,x}$$

is a singleton set. This completes the proof. □

Let us now define

$$\boxed{d_\lambda(x) := \underset{d \in D(x)}{\text{argmin}} \;\; P_{\lambda,x}(d).} \tag{8}$$

Then we denote

$$\boxed{\theta_\lambda(x) := P_{\lambda,x}(d_\lambda(x)).} \tag{9}$$

In the rest of the article we shall denote by $(P_{\lambda(x)})$ the problem

$$\boxed{\min_{d \in D(x)} P_{\lambda,x}(d).} \tag{10}$$

Using Lemma 4, we note that d_λ is well defined over S. Observe that $\theta_\lambda(x) \leq 0$ for all x, since

$$\theta_\lambda(x) \leq P_\lambda(0) = 0, \tag{11}$$

as $0 \in D(x)$.

Lemma 5 *Let us consider that Bregman distance D_h is induced by a strongly convex and twice continuously differentiable function h on S. Let $x \in S$ and $x \in int \left(\bigcap_{j \in \Lambda_m} dom\ g_j \right)$. Then $\theta_\lambda(x) < 0$ if and only if x is a non-critical point. In other words x is a critical point if and only if $\theta_\lambda(x) = 0$.*

Proof: Let $x \in S$ be a non-critical point of (MOP); it is clear x is not a weak efficient solution and by Lemma 3, there exists $d \in \mathbb{R}^n$, such that $\Psi_x(d) < 0$.

Observe that Ψ_x is a convex function in $d \in \mathbb{R}^n$. Further since $\Psi_x(0) = 0$, we have for $\alpha \in (0,1)$,

$$\begin{aligned} \Psi_x(\alpha d) &= \Psi_x(\alpha d + (1-\alpha)0) \\ &\leq \alpha \Psi_x(d) + (1-\alpha)\Psi_x(0) \\ &= \alpha \Psi_x(d). \end{aligned}$$

Now since the Bregman distance D_h is induced by a twice continuously differentiable function, we now observe the following.

$$D_h(x,y) = h(x) - h(y) - \langle \nabla h(x), x - y \rangle.$$

Now using Taylor's expansion we have

$$D_h(x,y) = \frac{1}{2!} \langle x - y, \nabla^2 h(y + \xi(x-y)) x - y \rangle,$$

where $\xi \in (0,1)$. Note that $y + \xi(x-y) \in S$.

Since S is open convex set, $x \in S$, and $x+d \in \bar{S}$ then for $\alpha \in (0,1)$ we have $x+\alpha d \in S$. Hence $x+\alpha d \in \bar{S}$. Hence $\alpha d \in D(x)$. Thus

$$\begin{aligned}
\theta_\lambda(x) &\leq P_{\lambda,x}(\alpha d)\\
\theta_\lambda(x) &\leq \Psi_x(\alpha d) + \frac{1}{\lambda} D_h(x+\alpha d, x)\\
&\leq \alpha \Psi_x(d) + \frac{1}{\lambda}\frac{\alpha^2}{2}\langle d, \nabla^2 h(x+\xi(\alpha d))d\rangle \text{ where } \xi \in (0,1).
\end{aligned}$$

As $\alpha > 0$ we have

$$\frac{\theta_\lambda(x)}{\alpha} \leq \Psi_x(d) + \frac{\alpha}{2\lambda}\langle d, \nabla^2 h(x+\xi(\alpha d))d\rangle.$$

Note that strong convexity of h shows that $\langle d, \nabla^2 h(x+\xi(\alpha d))d\rangle > 0$. Since $\Psi_x(d) < 0$, we can always choose a $\alpha_0 \in (0,1)$ which is sufficiently small such that $\forall\ \alpha \in (0, \alpha_0)$ we have

$$\Psi_x(d) + \frac{\alpha}{2\lambda}\langle d, \nabla^2 h(x+\xi(\alpha d))d\rangle < 0.$$

Thus for α sufficiently small $\frac{\theta_\lambda(x)}{\alpha} < 0$. As $\alpha \in (0,1)$ we have $\theta_\lambda(x) < 0$.
Conversely suppose $\theta_\lambda(x) < 0$. Then from (9), $\Psi_x(d_\lambda(x)) < \theta_\lambda(x) < 0$. Hence from Lemma 3, x is a non-critical point.

Lemma 6 *If x is a critical point if and only $d_\lambda(x) = 0$ for any $\lambda > 0$.*

Proof: By Lemma 4, we know that $P_{\lambda,x}(d)$ has a unique minimizer over $D(x)$. Let that be $d_\lambda(x)$, that is

$$\{d_\lambda(x)\} = \underset{d \in D(x)}{\text{argmin}}\ P_{\lambda,x}(d).$$

Hence if $d_\lambda(x) \neq 0$ then,

$$P_{\lambda,x}(d_\lambda(x)) < P_{\lambda,x}(0)$$

This shows that $\theta_\lambda(x) < 0$. This implies x is not a critical point by Lemma 5. Thus if x is a critical point we have $d_\lambda(x) = 0$.
Conversely we have to show that if $d_\lambda(x) = 0$, where $\lambda > 0$, then x is a critical point. So $d_\lambda(x) = 0$ implies $\{0\} = \underset{d \in D(x)}{\text{argmin}}\ P_{\lambda,x}(d)$. Hence

$$\theta_\lambda(x) = P_{\lambda,x}(0) = 0.$$

This implies x is a critical point by Lemma 5. □

Lemma 7 *Suppose the Bregman distance D_h is induced by a boundary coercive Bregman function h, then for every $x \in S$, $x + d_\lambda(x) \in S$ and $d_\lambda(x) \in int(D(x))$.*

Proof: We know that $d_\lambda(x) = \underset{d\in D(x)}{\operatorname{argmin}}\ P_{\lambda,x}(d)$. Since $P_{\lambda,x}(d)$ is strongly convex, from the standard optimality conditions of convex optimization we have,

$$0 \in \partial P_{\lambda,x}\left(d_\lambda(x)\right) + N_{D(x)}\left(d_\lambda(x)\right)$$

This shows that

$$0 \in \partial\left(\Psi_x + \frac{1}{\lambda}h\left(x+.\right)\right)\left(d_\lambda(x)\right) - \frac{1}{\lambda}\nabla h(x) + N_{D(x)}\left(d_\lambda(x)\right).$$

Let $\omega \in N_{D(x)}\left(d_\lambda(x)\right)$ such that

$$\frac{1}{\lambda}\nabla h(x) - \omega \in \partial\left(\Psi_x + \frac{1}{\lambda}h\left(x+.\right)\right)\left(d_\lambda(x)\right).$$

Thus

$$\partial\left(\Psi_x + \frac{1}{\lambda}h\left(x+.\right)\right)\left(d_\lambda(x)\right) \neq \emptyset. \tag{12}$$

We shall show that if h is boundary coercive then

$$\partial\left(\Psi_x + \frac{1}{\lambda}h\left(x+.\right)\right)(d) = \emptyset \tag{13}$$

whenever $x+d \in bdS$. On the contrary let $d \in \mathbb{R}^n$ be such that $x+d \in bdS$ and

$$\partial\left(\Psi_x + \frac{1}{\lambda}h\left(x+.\right)\right)(d) \neq \emptyset. \tag{14}$$

Then by definition $d \in D(x)$. Now let $\xi \in \partial\left(\Psi_{\lambda,x} + \frac{1}{\lambda}h\left(x+.\right)\right)(d)$. Let $\bar{d} \in \mathbb{R}^n$, be such that $x+\bar{d} \in S$. Since $x \in S$ and S is open we can choose $\bar{d}$ with sufficiently small such that $x+\bar{d} \in S$. Now let us consider a decreasing sequence $\{\epsilon_p\}$, such that $0 < \epsilon_p < 1$ and $\lim\limits_{p\to\infty} \epsilon_p = 0$. For each $p \in \mathbb{N}$, let us set

$$d^p = (1-\epsilon_p)d + \epsilon_p\bar{d}.$$

Since $D(x)$ is convex $d^p \in D(x)$ and we will show that $x+d^p \in S$. Observe that

$$\begin{aligned} x+d^p &= x + (1-\epsilon_p)d + \epsilon_p\bar{d} \\ &= (1-\epsilon_P)(x+d) + \epsilon_p(x+\bar{d}). \end{aligned}$$

Since $x+d \in bdS$ and $x+\bar{d} \in S$, we have $x+d^p \in S$. Also observe that

$$\lim_{p\to\infty} d^p = d$$

By the convexity of h we have

$$\begin{aligned} & h(x+d) - h(x+d^p) \geq \langle \nabla h(x+d^p), d-d^p\rangle \\ i.e.\quad & h(x+d^p) - h(x+d) \leq \langle \nabla h(x+d^p), d^p-d\rangle \end{aligned} \tag{15}$$

Now

$$\begin{aligned}\epsilon_p\langle \xi, \bar{d}-d\rangle &= \langle \xi, \epsilon_p(\bar{d}-d)\rangle \\ &= \langle \xi, d^p-d\rangle \\ &\le \Psi_x(d^p)-\Psi_x(d)+\left[\frac{1}{\lambda}h(x+d^p)-\frac{1}{\lambda}h(x+d)\right] \\ &\le \Psi_x(d^p)-\Psi_x(d)+\frac{1}{\lambda}\langle \nabla h(x+d^p), d^p-d\rangle \qquad (16)\end{aligned}$$

Now using the convexity of Ψ_x we have

$$\begin{aligned}&\Psi_x(d^p)-\Psi_x(d) \\ &= \Psi_x\left((1-\epsilon_p)d+\epsilon_p\bar{d}\right)-\Psi_x(d) \\ &\le (1-\epsilon_p)\Psi_x(d)+\epsilon_p\Psi_x(\bar{d})-\Psi_x(d) \\ &= \epsilon_p\left(\Psi_x(\bar{d})-\Psi_x(d)\right) \qquad (17)\end{aligned}$$

Now

$$\begin{aligned}\bar{d}-d^p &= \bar{d}-(1-\epsilon_p)d-\epsilon_p\bar{d} \\ &= (1-\epsilon_p)(\bar{d}-d).\end{aligned}$$

Therefore

$$\frac{\epsilon_p}{1-\epsilon_p}(\bar{d}-d^p) = d^p-d \qquad (18)$$

Now using (16) and (17) in (18) we have

$$\langle \xi, \epsilon_p(\bar{d}-d)\rangle \le \epsilon_p\left(\Psi_x(\bar{d})-\Psi_x(d)\right)+\frac{\epsilon_p}{1-\epsilon_p}\langle \nabla h(x+d^p), \bar{d}-d^p\rangle.$$

This shows that

$$(1-\epsilon_p)\left[\langle \xi, \bar{d}-d\rangle-\left(\Psi_x(\bar{d})-\Psi_x(d)\right)\right] \le \langle \nabla h(x+d^p), \bar{d}-d^p\rangle$$

If we set $z=x+d$, $y^p=x+d^p$, then $y^p\to x+d\in bdS$ as $p\to\infty$ and pass the above inequality we have

$$(1-\epsilon_p)\left[\langle \xi, \bar{d}-d\rangle-\left(\Psi_x(\bar{d})-\Psi_x(d)\right)\right] \le \langle \nabla h(y^p), z-y^p\rangle$$

Now using the fact that h is boundary coercive we have

$$\langle \nabla h(y^p), z-y^p\rangle \to -\infty$$

as $p\to\infty$, while the left-side has a finite limit. Hence our assumption that

$$\partial\left(\Psi_x+h(x+.)\right)\neq\emptyset$$

where $x+d\in bdS$, is not true. Hence for any d such that $x+d\in bdS$, we must have

$$\partial\left(\Psi_x+h(x+.)\right)=\emptyset.$$

Thus if $x + d_\lambda(x) \in bdS$, then we will have

$$\partial \left(\Psi_x + h(x + .)\right)(d_\lambda(x)) = \emptyset, .$$

which contradicts (12). Hence $x + d_\lambda(x) \in S$.
Since S is an open set, there exists $\epsilon_1 > 0$ such that $x + d_\lambda(x) + d \in S$ for every $d \in \mathbb{R}^n$ such that $\|d\| \leq \epsilon_1$. This implied $d_\lambda(x) + d \in D(x)$ for every $d \in \mathbb{R}^n$ such that $\|d\| \leq \epsilon_1$. That is $d_\lambda(x) \in int(D(x))$. □

Lemma 8 *Consider the problem (MOP) where for each $j \in \Lambda_m$ the functions g_j are finite-valued convex functions on $\mathbb{R}^m$ and $x^k \in S$ be such that $x^k \to x^*$. Given any $d \in \mathbb{R}^n$ there exists a subsequence $\{x^{k_p}\}$ such that*

$$\lim_{p\to\infty} \Psi_{x^{k_p}}(d) = \Psi_{x^*}(d).$$

Proof: For any k we have

$$\Psi_{x^k}(d) = \max_{j\in\Lambda_m} \Psi_j(x^k, d)$$

where

$$\Psi_j(x^k, d) = \langle \nabla f_j(x^k), d\rangle + g_j(x^k + d) - g_j(x^k).$$

Using the counting principle we can find $r \in \Lambda_m$, and subsequence $\{x^{k_p}\}$ of $\{x^k\}$ such that

$$\begin{aligned}\Psi_{x^k}(d) &= \Psi_r(x^{k_p}, d) \quad \forall p = 1, 2, ...\\ &= \langle \nabla f_r(x^{k_p}), d\rangle + g_r(x^{k_p} + d) - g_r(x^{k_p}).\end{aligned}$$

Since g_r is a finite-valued convex function it is continuous. Therefore

$$\begin{aligned}\lim_{p\to\infty} \Psi_{x^{k_p}}(d) &= \langle \nabla f_r(x^*), d\rangle + g_r(x^* + d) - g_r(x^*)\\ &= \Psi_r(x^*, d) \leq \Psi_{x^*}(d). \end{aligned} \tag{19}$$

Further

$$\Psi_{x^{k_p}}(d) \geq \Psi_j(x^{k_p}, d), \quad \forall j \in \Lambda_m.$$

Therefore

$$\begin{aligned}&\lim_{p\to\infty} \Psi_{x^{k_p}}(d) \geq \langle \nabla f_j(x^*), d\rangle + g_j(x^* + d) - g_j(x^*)\\ &\lim_{p\to\infty} \Psi_{x^{k_p}}(d) \geq \Psi_j(x^*, d), \quad \forall j \in \Lambda_m\end{aligned}$$

Hence

$$\lim_{p\to\infty} \Psi_{x^{k_p}}(d) \geq \Psi_{x^*}(d) \tag{20}$$

Thus from (19) and (20) we have that

$$\lim_{p\to\infty} \Psi_{x^{k_p}}(d) = \Psi_{x^*}(d).$$

Hence the result. □.

Lemma 9 *Assume that g_j's are continuous. Let $x^k \in S \ \forall \ k \in \mathbb{N}$ and $d_\lambda(x^k)$ be such that*

$$\lim_{k\to\infty} d_\lambda(x^k) = 0.$$

Let us assume that $x^k \to x^*$. Then there exists a subsequence of $\{x^{k_p}\}$ of $\{x^k\}$ such that

$$\lim_{p\to\infty} \Psi_{x^{k_p}}(d_\lambda(x^{k_p})) = \Psi_{x^*}(0) = 0.$$

Proof: From the definition of $\Psi_{x^k}(d_\lambda(x^k))$, using the counting principle we conclude that there exists a sub subsequence $\{x^{k_p}\}$ of $\{x^k\}$ and $r \in \Lambda_m$ such that $\forall \ j \in \mathbb{N}$

$$\Psi_{x^{k_p}}(d_\lambda(x^{k_p})) = \Psi_r(x^{k_p}, d_\lambda(x^{k_p})).$$

Therefore

$$\Psi_{x^{k_p}}(d_\lambda(x^{k_p})) = \langle \nabla f_r(x^{k_p}), d_\lambda(x^{k_p})\rangle + g_r(x^{k_p} + d_\lambda(x^{k_p})) - g_r(x^{k_p}).$$

This implies

$$\lim_{p\to\infty} \Psi_{x^{k_p}}(d_\lambda(x^{k_p})) = 0 = \Psi_{x^*}(0).$$

This proves the result. □

4 The algorithm and convergence analysis

In this section we will present an algorithm for solving (MOP). As mentioned earlier our algorithm can be viewed as a generalization of the proximal point algorithm for the scalar case to the vector case. However instead of using the Euclidean distance we use the Bregmann distance.
At any iterate x^k, we know that if x^k is not critical point then $\theta_\lambda(x^k) < 0$. We also know that for a given $\lambda > 0$,

$$\theta_\lambda(x^k) = P_{\lambda,x^k}(d_\lambda(x^k)).$$

Hence the vector $d_\lambda(x^k)$ can be viewed as a direction which mimics the idea of a descent direction at given iterate in scalar optimization. Thus it motivates us to define an iteration step as

$$x^{k+1} = x^k + \alpha_k d_\lambda(x^k)$$

where $\alpha_k > 0$, is a suitably chosen step length. Let us now try to explain why such an iteration makes sense. If x^k is a non-critical point, from the proof of Lemma 5, we see that

$$\Psi_{x^k}(d_\lambda(x^k)) < \theta_\lambda(x^k) < 0.$$

Now by the definition of $\Psi_{x^k}(d_\lambda(x^k))$ we have $\forall\ j \in \Lambda_m$,

$$\langle \nabla f_j(x^k), d_\lambda(x^k)\rangle + g_j(x^k + d_\lambda(x^k)) - g_j(x^k) < 0.$$

Hence by the convexity of g_j for each $j \in \Lambda_m$ we conclude that

$$g'(x^k, d_\lambda(x^k)) + \langle f_j(x^k), d_\lambda(x^k)\rangle < 0.$$

Hence for each $j \in \Lambda_m$ we have

$$F_j'(x^k, d_\lambda(x^k)) < 0$$

which shows that $d_\lambda(x^k)$ acts as a descent direction, for each F_j. This encourages us to consider an Armijo type back tracking criteria at each iteration. At the k-th iteration the Armijo type backtracking criteria seeks to find $\alpha_k > 0$ such that for each $j \in \Lambda_m$, find

$$F_j(x^k + \alpha_k d_\lambda(x^k)) \leq F_j(x^k) + \beta\alpha_k \Psi_{x^k}(d_\lambda(x^k)) \quad \forall j \in \Lambda_m, \quad \beta \in (0,1). \quad (21)$$

The use of the Armijo type sufficient decrease condition can be justified using the following lemma.

Lemma 10 *Suppose x^k is a noncritical point and $d_\lambda(x^k)$ is the solution of $P_\lambda(x^k)$. Then the line search condition (21) holds for every $\alpha_k > 0$ sufficiently small.*

Proof: The result follows from Lemma 3.3 of [30]. □

Though it is clear but we would like to mention that using Lemma 7 we can easily show that $x^{k+1} \in S$. Thus our descent scheme is well-defined. The iterates generated by the scheme lies in the lies in S.For the economy of the writing we will use henceforth use the following short cuts, we set

$$\theta_\lambda(x^k) := \theta_\lambda^k \quad \text{and} \quad d_\lambda(x^k) := d_\lambda^k.$$

We will state our **basic assumption** for the problem (MOP) for which the algorithm stated below is designed.

i) For each $j \in \Lambda_m$ the function f_j is a finite-valued differentiable convex function on $\mathbb{R}^n$
ii) For each $j \in \Lambda_m$ the function g_j is finite-valued on convex function on $\mathbb{R}^n$.

The above two conditions will be assumed on the rest of the articles. Further it is important to note that every differentiable convex function on $\mathbb{R}^n$ is continuously differentiable and every finite-valued convex function is continuous.

Algorithm 1 *(Proximal gradient method using Bregman function)*

Step 1 *Choose σ-strongly convex and boundary coercive Bergman function h, initial approximation x^0, scalars $\gamma, \beta \in (0,1)$, $\lambda > 0$ and $\epsilon > 0$.Set $k := 0$*

Step 2 *Solve the subproblem* $P_\lambda(x^k)$ *to find* d_λ^k *and* θ_λ^k.
Step 3 *If* $\|d_\lambda^k\| < \epsilon$, *then stop. Else go to Step 4.*
Step 4 *Choose a suitable step length* α_k *as first element of the sequence* $\{1, \gamma, \gamma^2, ...\}$ *satisfying (21).*
Step 5 *Update* $x^{k+1} = x^k + \alpha_k d_\lambda^k$, *set* $k := k+1$, *and go to Step 2.*

We are in a position to state our first main result towards the convergence analysis of the above mentioned algorithm referred to as Algorithm 1 . Before stating our first result namely Theorem 1 we shall state the following lemma which will be used in it's proof. The proof of the lemma is straight forward and hence avoided.

Lemma 11 *Consider a sequence* $\{\alpha_k w^k\}$ *where* $\alpha_k \in \mathbb{R}$ *and* $w^k \in \mathbb{R}^n$ *for all* $k \in \mathbb{N}$. *If* $\alpha_k \to \alpha^*$ *and* $w^k \to w^*$ *as* $k \to \infty$, *then* $\alpha_k w^k \to \alpha^* w^*$ *as* $k \to \infty$.

Theorem 1 *Suppose* $\{x^k\}$ *is a sequence generated by Algorithm 1, Bregman distance is induced by a* σ*-strongly convex function* h, ∇f_j *is Lipschitz continuous for every* $j \in \Lambda_m$ *with Lipschtz constant* $L > 0$, *and the level set* $M := \{x \in \mathbb{R}^n : F(x) \leq F(x^0)\}$, *where* x^0 *is the starting point of Algorithm 1, is bounded. Then* $\lim\limits_{k\to\infty} d_\lambda^{k_j} = 0$ *holds for every convergent subsequence* $\{x^{k_j}\}$ *of* $\{x^k\}$

Proof: First we show that there exists $\bar{\alpha} > 0$ such that $\alpha_k \geq \bar{\alpha}$ holds for every k. Since ∇f_j is Lipschitz continuous, for any α we have

$$f_j(x^k + \alpha d_\lambda^k) \leq f_j(x^k) + \alpha\langle \nabla f_j(x^k), d_\lambda^k\rangle + \frac{L}{2}\alpha^2 \|d_\lambda^k\|^2 \;\;\forall\;\; j\Lambda_m. \tag{22}$$

From convexity of g_j for any $\alpha \in [0,1]$, we have

$$\begin{aligned} g_j(x^k + \alpha d_\lambda^k) - g_j(x^k) &\leq \alpha(g_j(x^k + d_\lambda^k)) + (1-\alpha) g_j(x^k) - g_j(x^k) \\ &= \alpha(g_j(x^k + d_\lambda^k) - g_j(x^k)) \end{aligned} \tag{23}$$

Hence from (22) and (23) for any $\alpha \in [0,1]$,

$$\begin{aligned} F_j(x^k + \alpha d_\lambda^k) - F_j(x^k) &\leq \alpha \Psi_j(x^k, d_\lambda^k) + \frac{L}{2}\alpha^2 \|d_\lambda^k\|^2 \\ &\leq \alpha \Psi_{x^k}(d_\lambda^k) + \frac{L}{2}\alpha^2 \|d_\lambda^k\|^2 \;\;\forall\;\; j \in \Lambda_m. \end{aligned} \tag{24}$$

From Step 4 of Algorithm 1, either $\alpha_k = 1$ or there exists $k_1 \in \mathbb{N}$ such that $\alpha_k = \gamma^{k_1}$ holds. If $\alpha_k = \gamma^{k_1}$ then there exists at least one $\hat{j} \in \Lambda_m$ such that

$$F_{\hat{j}}(x^k + \gamma^{k_1 - 1} d_\lambda^k) - F_{\hat{j}}(x^k) > \gamma^{k_1-1}\beta \Psi_{x^k}(d_\lambda^k). \tag{25}$$

Then from (24),

$$\begin{aligned} \gamma^{k_1-1}\beta \Psi_{x^k}(d_\lambda^k) &< \gamma^{k_1-1}\Psi_{x^k}(d_\lambda^k) + \frac{L}{2}\gamma^{2(k_1-1)}\|d_\lambda^k\|^2 \\ -(1-\beta)\Psi_{x^k}(d_\lambda^k) &< \frac{L}{2}\gamma^{(k_1-1)}\|d_\lambda^k\|^2. \end{aligned} \tag{26}$$

From (11), $\theta_\lambda(x^k) = \Psi_{x^k}(d_\lambda^k) + \frac{1}{\lambda} D_h(x^k + d_\lambda^k, x^k) \le 0$. That is, $\Psi_{x^k}(d_\lambda^k) \le -\frac{1}{\lambda} D_h(x^k + d_\lambda^k, x^k)$. Since the Bregman distance is induced by a σ-strongly convex function h,

$$\begin{aligned}\Psi_{x^k}(d_\lambda^k) &\le -\frac{1}{\lambda} D_h(x^k + d_\lambda^k, x^k) \\ &\le -\frac{\sigma}{2\lambda}\|d^k\|^2\end{aligned}$$

This implies

$$\frac{\sigma}{2\lambda}\|d^k\|^2 \le -\Psi_{x^k}(d_\lambda^k) \tag{27}$$

Using (27) in (26),

$$\begin{aligned}\frac{\sigma(1-\beta)}{2\lambda}\|d^k\|^2 &< \frac{L}{2}\gamma^{k_1-1}\|d^k\|^2 \\ \gamma^{k_1} &> \frac{\sigma(1-\beta)\gamma}{\lambda L}.\end{aligned}$$

Choose $\bar{\alpha} = \min\{1, \frac{\sigma(1-\beta)\gamma}{\lambda L}\}$. Then we have $\alpha_k \ge \bar{\alpha}$ for every k.
Now from Step 4 of Algorithm 1 for any $N \in \mathbb{N}$ and $j \in \Lambda_m$ by repeated application of the Armijo decrease criteria we have

$$\begin{aligned}F_j(x^{N+1}) - F_j(x^0) &\le \beta \sum_{k=0}^{N} \alpha_k \Psi_{x^k}(d_\lambda^k) \\ &\le -\frac{\beta}{\lambda}\sum_{k=0}^{N} \alpha_k D_h(x^k + d_\lambda^k, x^k) \quad \text{(from (11))} \\ &\le -\frac{\beta\bar{\alpha}}{\lambda}\sum_{k=0}^{N} D_h(x^k + d_\lambda^k, x^k) \quad (\because \alpha_k \ge \bar{\alpha})\end{aligned} \tag{28}$$

This shows that $x^k \in M = \{x \in \mathbb{R}^n | F(x) \le F(x^0)\}$ for every k. That is $\{x^k\}$ is bounded. Since for each $j \in \Lambda_m$ the function F_j is continuous and so is F. This shows that M is closed and as we have already assumed that M is bounded and M is compact. Hence the sequence $\{F_j(x^k)\}$ is bounded. Then from (28),

$$\frac{\beta\bar{\alpha}}{\lambda}\sum_{k=0}^{\infty} D_h(x^k + d_\lambda^k, x^k) < \infty.$$

This implies that

$$\lim_{k\to\infty} D_h(x^k + d_\lambda^k, x^k) = 0. \tag{29}$$

Next we show that the sequence $\{x^k + d_\lambda^k\}$ is bounded. On contrary suppose $\|x^k + d_\lambda^k\| \to \infty$ as $k \to \infty$. From the descent scheme we have

$$\begin{aligned} x^{k+1} &= x^k + \alpha_k d_\lambda^k \\ &= (1-\alpha_k)x^k + \alpha_k(x^k + d_\lambda^k). \end{aligned}$$

This shows that

$$\frac{x^{k+1}}{\|x^k + d_\lambda^k\|} - \frac{1}{\|x^k + d_\lambda^k\|} + \frac{\alpha_k x^k}{\|x^k + d_\lambda^k\|} = \alpha_k \frac{x^k + d_\lambda^k}{\|x^k + d_\lambda^k\|}. \tag{30}$$

Let us set

$$w^k = \frac{x^k + d_\lambda^k}{\|x^k + d_\lambda^k\|}.$$

Now as $\|w^k\| = 1$ for all $k \in \mathbb{N}$, we see that $\{w^k\}$ is a bounded sequence and without loss of generality we assume that $w^k \to w^*$ where $\|w^*\| = 1$.
Also α_k is bounded since form Step 4 of Algorithm 1 and Lemma 10 we have $\alpha_k \in (0, 1]$ for every k. We conclude using the boundedness of $\{x^k\}$ that the left side of (30) converges to zero. This shows that

$$\lim_{k\to\infty} \alpha_k w^k = 0. \tag{31}$$

Since α^k is bounded we have that there is a subsequence $\{\alpha_{k_j}\}$ such $\alpha_{k_j} \to \alpha^*$. Then using (31) we can write that

$$\lim_{j\to\infty} \alpha_{k_j} w^{k_j} = 0$$

Now by using Lemma 11 we conclude that $\alpha^* w^* = 0$. Hence $\alpha^*\|w^*\| = 0$. Since $\|w^*\| = 1$ we have that $\alpha^* = 0$. This contradicts the fact $\alpha_k \geq \overline{\alpha} > 0$.
Since the sequence $\{x^k\}$ is bounded there exists a convergent subsequence. Consider any convergent subsequence of $\{x^{k_j}\}$ of $\{x_k\}$. Assume that x^{k_j} converges to say x^* as $j \to \infty$. Now by using (29) we conclude that

$$\lim_{j\to\infty} D_h(x^{k_j} + d_\lambda^{k_j}, x^{k_j}) = 0.$$

Hence by using property (v) in Definition 1 we conclude that

$$\lim_{j\to\infty} x^{k_j} + d_\lambda^{k_j} = x^*$$

From the the above equality it is straight-forward to show that $d_\lambda^{k_j} \to 0$ as $j \to \infty$. Hence the result. □
We shall now present below our main convergence result for the problem (MOP). In the following result we show that every limit point of the sequence of iterates generated by Algorithm 1 is actually a weak minimizer of (MOP).

Theorem 2 *Let $\{x^k\}$ be a sequence generated by Algorithm 1. Assume further that ∇f_j is Lipschitz on $\mathbb{R}^n$ for every $j \in \Lambda_m$ with Lipschitz constant $L > 0$. Further assume that the level set $M = \{x \in \mathbb{R}^n : F(x) \leq F(x_0)\}$, where x^0 is the starting point of Algorithm 1 is bounded. Then $\{x^k\}$ is bounded and any limit point of the sequence is a weak minimizer of (MOP).*

Proof: From Lemma 8, for any $d \in \mathbb{R}^n$, there exists a subsequence $\{x^{k_p}\}$ of $\{x^k\}$ such that

$$\lim_{p\to\infty} \Psi_{x^{k_p}}(d) = \Psi_{x^*}(d) \tag{32}$$

Now from $\{x^{k_p}\}$ we can extract a subsequence $\{x^{k_{p_l}}\}$ such that

$$\lim_{l\to\infty} \Psi_{x^{k_{p_l}}}(d_\lambda^{k_{p_l}}) = \Psi_{x^*}(0), \tag{33}$$

from Lemma 9. Let $\xi^{k_{p_l}} \in \partial\Psi_{x^{k_{p_l}}}(d_\lambda^{k_{p_l}})$. Then

$$\Psi_{x^{k_{p_l}}}(d) - \Psi_{x^{k_{p_l}}}(d_\lambda^{k_{p_l}}) \geq \langle \xi^{k_{p_l}}, d - d_\lambda^{k_{p_l}} \rangle$$

As $l \to \infty$, we have

$$\Psi_{x^*}(d) - \Psi_{x^*}(0) \geq \langle \lim_{l\to\infty} \xi^{k_{p_l}}, d \rangle \tag{34}$$

We will show that $\lim_{l\to\infty} \xi^{k_{p_l}} = 0$. Now from the definition of $d_\lambda^k = d_\lambda^k(x^k)$, we have

$$0 \in \partial\Psi_{x^k}(d_\lambda^k) + \nabla h(x^k + d_\lambda^k) - \nabla h(x^k) + N_{D(x^k)}(d_\lambda^k)$$

As $d_\lambda^k \in$ *int* $D(x^k)$, $N_{D(x^k)}(d_\lambda^k) = \{0\}$. This implies

$$0 \in \partial\Psi_{x^k}(d_\lambda^k) + \nabla h(x^k + d_\lambda^k) - \nabla h(x^k).$$

Thus there exists $\xi^k \in \partial\Psi_{x^k}(d_\lambda^k)$ such that

$$0 = \xi^k + \nabla h(x^k + d_\lambda^k) - \nabla h(x^k)$$

As $k \to \infty$, we have

$$0 = \lim_{k\to\infty} \xi^k$$

Thus $\lim_{l\to\infty} \xi^{k_{p_l}} = 0$. Thus from (34), we have

$$\Psi_{x^*}(d) - \Psi_{x^*}(0) \geq \langle 0, d \rangle$$

This implies that for any $d \in \mathbb{R}^n$ we have $\Psi_{x^*}(d) \geq \Psi_{x^*}(0) = 0$. Thus we have for any $d \in \mathbb{R}^n$

$$\Psi_{x^*}(d) \geq 0.$$

Now by applying Lemma 3 we see that x^* is a weak minimizer of (MOP). □

Remark 1 *It is apparent by the very definition of a Bregman distance that the solution obtained by choosing a particular type of Bregman distance will only lie in its zone of definition. If there is a weak Pareto minimizer which is not lying in the zone of that specific Bregman function inducing the distance then it cannot be captured by using that particular Bregman function. However, as different Bregman functions have different zones, for various different starting points we can use the different Bregman functions and obtain as much part of the approximate Pareto front. This is the built in flexibility of using the Bregman distance. Example 1 shows that the approach as described above is actually feasible.*

5 Numerical examples

In this section Algorithm 1 is verified and compared with proximal gradient method for (MOP) developed in [30] using a set of problems. MATLAB (2019a) code is developed for both methods. Equivalent form of $P_\lambda(x^k)$ is

$$\big(P_0(x^k)\big) \min_{t,d} \;\; t + \frac{1}{\lambda} D_h(x^k + d, x^k)$$
$$\text{s. t. } \langle \nabla f_j(x^k), d\rangle + g_j(x^k + d) - g_j(x^k) \le t.$$

Suppose (t^k, d^k) is solution of $\big(P_0(x^k)\big)$, then $\Psi_{x^k}(d^k_\lambda) = t^k$ and $\theta_\lambda(x^k) = t^k + \frac{1}{\lambda} D_h(x^k + d^k_\lambda, x^k)$. We have used the solver *CVX* to solve $\big(P_0(x^k)\big)$. Throughout the numerical computation, $\lambda = 2$ is considered and $|\theta_\lambda(x^k)| < 10^{-5}$ or maximum 200 iterations is considered as stopping criteria.
Following different Bregman functions are considered.

1. **Squared elliptic norm:** $h_1 : \mathbb{R}^n \to \mathbb{R}$ defined by $h_1(x) = \langle x, Qx\rangle$, where Q is a symmetric positive definite matrix. In this case

   $$D_{h_1}(x, y) = \langle (x - y), Q(x - y)\rangle.$$

   Suppose eig_{ij} be the i^{th} eigenvalue of $\nabla^2 f_j(x^0)$, $i \in \Lambda_n$, $j \in \Lambda_m$. Define $\nu = \min_{j\in\Lambda_m} \min_{i\in\Lambda_n} eig_{ij}$. In our numerical computations we have considered $Q = \mu I_n$, where $\mu = \begin{cases} \nu & \text{if } \nu \ge 1, \\ 1.01 & \text{otherwise} \end{cases}$.
   Then $D_{h_1}(x, y)$ satisfies all properties of Definition 1. Next we justify that D_{h_1} and h_1 satisfy the assumptions used in theoretical developments.

   Suppose $\{y^k\}$ be a sequence converging to ∞. Then for any $x \in \mathbb{R}^n$,

   $$\langle \nabla h_1(y^k), x - y^k\rangle = \langle 2Qy^k, x - y^k\rangle = 2\left(\langle Qy^k, x\rangle - \langle Qy^k, y^k\rangle\right).$$

   This $\to -\infty$ as $k \to \infty$ since second term of this expression use quadratic expression of y^k. This implies h_1 is boundary coercive.
   Now for any $x \in \mathbb{R}^n$, $\langle \nabla^2 h_1(x)d, d\rangle = \langle d, Qd\rangle = \mu\|d\|^2$ holds for every $d \in \mathbb{R}^n$. Hence h_1 is a μ-strongly convex function.

2. **Kullback-Leibler divergence:** $h_2 : \mathbb{R}^n_+ \to \mathbb{R}$ defined by $h_2(x) = \sum_{i=1}^n x_i \log(x_i)$ along with the conversion $0 \log(0) = 0$. In this case

$$D_{h_2}(x, y) = \sum_{i=1}^n \left(x_i \log(\frac{x_i}{y_i}) - x_i + y_i \right).$$

Then $D_{h_2}(x, y)$ satisfies all properties of Definition 1. Similar to previous function, we justify that D_{h_1} and h_1 satisfy the assumptions used in theoretical developments.

For any $x \in \mathbb{R}^n_{++}$, $\nabla h_2(x) = (1 + log(x_1), 1 + log(x_2), \ldots, log(x_n))$ and $\nabla^2 h_2(x) = diag(\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n})$. Suppose $\{y^k\}$ be a sequence converging to some $y \in bd(\mathbb{R}^n_+)$. Then $y_i = 0$ holds for at least one i. Now for any x,

$$\langle \nabla h_2(y^k), (x - y^k) \rangle = \sum_{i=1}^n x_i - \sum_{i=1}^n y_i^k + \sum_{i=1}^n x_i \log(y_i^k) - \sum_{i=1}^n y_i^k \log(y_i^k).$$

This $\to -\infty$ as $k \to \infty$ since there exists at least one i such that $x_i \log(y_i^k) \to x_i \log(0)$ as $k \to \infty$.
Next suppose $\sigma = \min\{\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n}\}$ then for any $x \in \mathbb{R}^n_{++}$ $\langle d, \nabla^2 h_2(x) d \rangle \geq \sigma \|d\|^2$ holds for every $d \in \mathbb{R}^n$. This implies h_2 is a σ-strongly convex function on any bounded subset of $\mathbb{R}^n_{++}$.

Note 1 *Suppose $h : \mathbb{R}^n \to \mathbb{R}$ is defined by $h(x) = \|x\|^2$. Then $d_h(x, y) = \|x - y\|^2$. In this case Algorithm 1 coincides with Algorithm 3.1 of [30].*

Solution of a multi-objective optimization problem is not isolated optimum points but a set of efficient solutions. To generate an approximate set of efficient solutions we have considered multi-start technique. Suppose lb and ub are lower and upper bound of x respectively. A set of uniformly distributed random initial points between lb and ub are considered and Algorithm 1 is executed individually. If $\mathcal{WX}^*$ is the collection of approximate critical points then the non dominated set $\mathcal{WX}^*$ is considered as an approximate set of efficient solutions. In $prox_2$, lb is considered in $\mathbb{R}^n_+$.

In the following example we have compared the approximate Pareto fronts using $Prox_1$ and $Prox_2$.

Example 1 *Consider the problem*

$$\min_{x \in \mathbb{R}^2} \quad (f_1(x) + g_1(x), f_2(x) + g_2(x))$$

where $f_1(x) = \left((x_1 - 5)^2 + (x_2 - 5)^2\right)$, $f_2(x) = \left((x_1 + 5)^2 + (x_2 + 5)^2\right)$ *and* $g_j(x) = \max\{\langle u_j, x \rangle, \langle \bar{u}_j, x \rangle\}$, $j = 1, 2$, *where* u_j *is chosen uniformly distributed random vector satisfying* $-.1 \leq u_{i,j} \leq .1$ *and* $\bar{u}j$ *is a uniformly distributed random vector such that* $-.1 \leq (A\bar{u})_{i,j} \leq .1$, $i, j = 1, 2$ *and* A *is an* 2×2 *random matrix.*

Since elliptic norm is defined for any $x, y \in \mathbb{R}^n$, we can obtain approximation of entire Pareto front using suitable set of initial approximations. As KL divergence is defined for $x, y \in \mathbb{R}^n_{++}$, using a set of positive initial approximations we can obtain Pareto front partially (only the image of positive efficient solutions). In Example 1, for elliptic norm we have considered a set of 100 initial approximations uniformly distributed in $(-10, -10)^T$ and $(10, 10)^T$. For KL divergence, we have considered a set of 100 initial approximations uniformly distributed in $(0, 0)^T$ and $(10, 10)^T$. One can observe from Figure 1, approximate Pareto front obtained by KL divergence contained in a part of the approximate Pareto front obtained by elliptic norm.

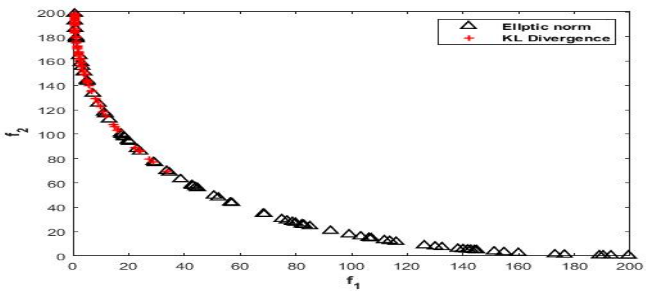

Fig. 1: Comparison of approximate Pareto fronts using $Prox_1$ and $Prox_2$

Similar to Example 1, we have considered a set of problems $F_j = f_j + g_j$ $j = 1, 2, \ldots, m$, where f_j is nonlinear test problem collected from different sources and g_j is of the form $g_j(x) = \max\{\langle u_j, x\rangle, \langle \bar{u}_j, x\rangle\}$, $j = 1, \ldots, m$, where u_j is uniformly distributed random vector such that $-.1 \le U_{i,j} \le .1$ $i = 1, 2, \ldots, n$ $j = 1, 2, \ldots, m$ and $\bar{u}_j$ is a uniformly distributed random vector such that $-.1 \le (A\bar{u})_{i,j} \le .1$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$ and A is an $n \times n$ random matrix.

Computational details for a set of problems are summarized in Table 1. Source of nonlinear test problems are provided in '*Source*'. Here $Prox_i$, $i = 1, 2, 3, 4$ denote the proximal gradient methods with Bregman function h_i and $Prox_0$ denote the proximal gradient method developed in [30]. For $Prox_0$, and $Prox_1$ lb and ub provided in Table 1. For $Prox_2$, we have considered $\bar{lb}_l = 0$, for every $lb_l < 0$ $l = 1, 2, \ldots, n$, where n is the number of variables. Here m, n, $\#\ It$, $\#F$, $\#T(s)$ denote number of objective functions, number of variables, number of iterations, number of function evaluations, and CPU time in seconds restively. One can observe from Table 1 that, $Prox_1$ requires less number of iterations, function evaluations, CPU time than $Prox_0$ in most cases. Lowest number of iterations, function evaluations, and CPU time between $Prox_1$ and

Problem	*Source*	(m,n)	lb^T	ub^T	$Prox_0$			$Prox_1$			$Prox_2$		
					$\#\ It$	$\#F$	$\#T(s)$	$\#\ It$	$\#F$	$\#T(s)$	$\#\ It$	$\#F$	$\#T(s)$
Example 1	-	(2, 2)	$(-10,-10)$	$(10,10)$	9615	9713	5472.48	**200**	**200**	**75.95**	1685	6268	813.32
BK1	[19]	(2,2)	$(-5,-5)$	$(10,10)$	6693	6793	3516.11	**200**	**200**	**74.00**	2158	7798	1205.85
FDSa	[17]	(3,3)	$(-2,-2,-2)$	$(2,2,2,)$	1154	2377	416.42	**700**	**700**	**281.98**	1455	3803	6420.51
FDSb	[17]	(3,5)	$(-2,\dots,-2)$	$(2,\dots,2)$	2915	7628	3882.34	**1972**	**4283**	**869.93**	3270	10210	1786.75
Jin1a	[22]	(2,2)	(-2,-2)	(4,4)	**200**	**200**	**73.81**	244	244	88.50	1944	3557	888.42
Jin1b	[22]	(2,5)	$(-2,\dots,-2)$	$(4,\dots,4)$	**1008**	**1008**	2004.55	1027	1027	**391.94**	578	578	329.24
Jin1c	[22]	(2,10)	$(-2,\dots,-2)$	$(4,\dots,4)$	**2059**	**2059**	3981.75	2084	2084	**828.52**	816	816	493.99
lovison1	[23]	(2,2)	(0,0)	(3,3)	320	456	112.49	**267**	**267**	**93.15**	955	2425	412.90
lovison4	[23]	(2,2)	(0,-1)	(6,1)	419	643	147.56	**322**	**392**	**113.11**	673	770	294.30
LRS1	[19]	(2,2)	(-50,-50)	(50,50)	2993	3122	1197.44	**263**	**263**	**104.16**	218	218	97.47
MHHM1	[19]	(3,1)	0	1	188	274	72.92	184	**184**	70.79	240	375	104.93
MHHM2	[19]	(3,2)	(0,0)	(1,1)	400	480	170.88	**200**	**200**	**83.41**	323	498	178.05
MLF2	[19]	(2,2)	(-2,-2)	(2,2)	1512	1659	746.54	**1502**	**1636**	bf 554.26	1182	1784	752.77
MOP1	[19]	(2,1)	-500	500	4109	4183	2058.53	**324**	**324**	**123.61**	200	200	85.10
MOP7	[19]	(3,2)	(-20,-20)	(20,20)	**5451**	**5451**	**2300.19**	5914	5914	2465.06	1652	1778	530.41
PNRa	[28]	(2,2)	(-2,-2)	(2,2)	1438	3826	528.91	**923**	**3161**	**332.41**	877	3474	392.38
PNRb	[28]	(2,2)	(-2,-2)	(2,2)	719	2711	259.46	**666**	**2236**	**239.34**	1014	5129	456.41
PNRc	[28]	(2,2)	(-2,-2)	(2,2)	1232	3409	481.9	**885**	**2520**	**313.60**	887	3314	381.91
PNRd	[28]	(2,2)	(-2,-2)	(2,2)	**536**	**1895**	**196.57**	641	2138	231.65	1136	5604	500.99
PNRe	[28]	(2,2)	(-2,-2)	(2,2)	868	3229	343.46	**651**	**2040**	**238.82**	1007	4938	454.28
SP1	[19]	(2,2)	(-1,-1)	(5,5)	655	1579	226.79	**588**	**1418**	**204.41**	608	2944	296.01
SSFYY1	[19]	(2,2)	(-100,-100)	(100,100)	808	939	307.80	**303**	**303**	**114.71**	202	203	89.32
VFM1	[19]	(3,2)	(-2,-2)	(2,2)	245	292	**92.52**	**214**	**240**	105.72	243	301	93.36
ZLT1a	[19]	(3,3)	$(-50,-50,-50)$	$(50,50,50)$	6994	7121	13894.99	**291**	**291**	**113.93**	207	209	94.17
ZLT1b	[19]	(3,5)	$(-50,\dots,-50)$	$(50,\dots,50)$	3872	4003	1951.28	**330**	**330**	**131.07**	441	672	276.78
ZLT1c	[19]	(3,10)	$(-50,\dots,-50)$	$(50,\dots,50)$	3292	3435	5497.86	**410**	**410**	**162.57**	1386	2777	975.42

Table 1: Computation details

$Prox_0$ for each problem is identified with bold number. Since feasible sets in $Prox_2$ is different from $Prox_1$ and $Prox_0$, we have not considered $Prox_2$ for comparison.

6 Conclusion

In this paper, we have developed a proximal gradient method nonlinear convex multi-objective optimization problems. In state of euclidian distance, we have used Bregman distance to find a suitable descent direction. This method is free from any kind of priori chosen parameters or ordering information of objective function. The ideas of this paper can be extended to non convex multi-objective optimization problems, which can be considered as a future scope of this paper. We have used multi-start technique to generate approximate Pareto fronts, which fails to generate a well distributed approximate Pareto front in some cases. In future we want to develop some initial point selection technique to generate a well distributed approximate Pareto front. Few generations of genetic algorithms can be considered in this process. To improve convergence rate we want to develop some higher order proximal gradient methods for nonlinear multi-objective optimization problems.

References

1. Ansary, M.A.T., Panda, G.: A sequential quadratically constrained quadratic programming technique for a multi-objective optimization problem. Eng. Optim. **51**(1), 22–41 (2019)
2. Ansary, M.A.T., Panda, G.: A sequential quadratic programming method for constrained multi-objective optimization problems. J. Appl. Math. Comput. **64**, 379–397 (2020)
3. Ansary, M.A.T., Panda, G.: A globally convergent SQCQP method for multi-objective optimization problems. SIAM J. Optim. **31**(1) (2021)
4. Beck, A.: First-order methods in optimization, vol. 25. SIAM (2017)
5. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
6. Beck, A., Teboulle, M.: A fast dual proximal gradient algorithm for convex minimization and applications. Oper. Res. Lett. **42**(1), 1–6 (2014)
7. Bello Cruz, J.: A subgradient method for vector optimization problems. SIAM J. Optim. **23**(4), 2169–2182 (2013)
8. Bento, G.C., Cruz Neto, J.X., Lopez, G., Soubeyran, A., Souza, J.C.O.: The proximal point method for locally lipschitz functions in multiobjective optimization with application to the compromise problem. SIAM J. Optim. **28**(2), 1104–1120 (2018)
9. Bonnel, H., Iusem, A.N., Svaiter, B.F.: Proximal methods in vector optimization. SIAM J. Optim. **15**(4), 953–970 (2005)
10. Boyd, S., Vandenberghe, L.: Introduction to applied linear algebra: vectors, matrices, and least squares. Cambridge university press (2018)
11. Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. U.S.S.R. Comput. Math. and Math. Phys. **7**(3), 200–217 (1967)
12. Censor, Y., Zenios, S.A.: Proximal minimization algorithm with D-functions. J. Optim. Theory Appl. **73**(3), 451–464 (1992)
13. Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using bregman functions. SIAM J. Optim. **3**(3), 538–543 (1993)

14. Deb, K.: Multi-Objective Optimization Using Evolutionary Algorithms. Wiely India Pvt. Ltd., New Delhi, India (2003)
15. Deb, K., Agrawal, S., Pratap, A., Meyarivan, T.: A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. In: International conference on parallel problem solving from nature, pp. 849–858. Springer (2000)
16. Deb, K., Jain, H.: An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. IEEE T. Evolut. Comput. **18**(4), 577–601 (2013)
17. Fliege, J., Drummond, L.M.G., Svaiter, B.F.: Newton's method for multiobjective optimization. SIAM J. Optim **20**(2), 602–626 (2009)
18. Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res. **51**(3), 479–494 (2000)
19. Huband, S., Hingston, P., Barone, L., While, L.: A review of multiobjective test problems and a scalable test problem toolkit. IEEE Trans. Evol. Comput. **10**(5), 477–506 (2006)
20. Iusem, A.N.: Some properties of generalized proximal point methods for quadratic and linear programming. J. Optim. Theory Appl. **85**(3), 593–612 (1995)
21. Iusem, N.A., Solodov, V.M.: Newton-type methods with generalized distances for constrained optimization. Optimization **41**(3), 257–278 (1997)
22. Jin, Y., Olhofer, M., Sendhoff, B.: Dynamic weighted aggregation for evolutionary multi-objective optimization: Why does it work and how? In: Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation, pp. 1042–1049 (2001)
23. Lovison, A.: Singular continuation: Generating piecewise linear approximations to pareto sets via global analysis. SIAM J. Optim. **21**(2), 463–490 (2011)
24. Miettinen, K.M.: Nonlinear Multiobjective Optimization. Kluwer, Boston (1999)
25. Montonen, O., Karmitsa, N., Mäkelä, M.: Multiple subgradient descent bundle method for convex nonsmooth multiobjective optimization. Optimization **67**(1), 139–158 (2018)
26. Nesterov, Y.: Gradient methods for minimizing composite functions. Math. Program. **140**(1), 125–161 (2013)
27. Neto, J.D.C., Da Silva, G.J.P., Ferreira, O.P., Lopes, J.O.: A subgradient method for multiobjective optimization. Comput. Optim. Appl. **54**(3), 461–472 (2013)
28. Preuss, M., Naujoks, B., Rudolph, G.: Pareto set and emoa behavior for simple multimodal multiobjective functions. In: Parallel Problem Solving from Nature-PPSN IX, pp. 513–522. Springer (2006)
29. Rockafellar, R.T.: Convex analysis, vol. 36. Princeton university press (1970)
30. Tanabe, H., Fukuda, E.H., Yamashita, N.: Proximal gradient methods for multiobjective optimization and their applications. Comput. Optim. Appl. **72**(2), 339–361 (2019)

Ackowledgement The first author Md Abu Talhamainuddin Ansary is grateful to the support provided by the Post-doctoral research fellowship of the Indian Institute of Technology, Kanpur. The second author Joydeep Dutta is gratefully acknowledges the support received through the SERB grant for the MATRICS project no MTR/2019/00694 for the project titled :**A Study of First Order Methods in Scalar and Vector Optimization**. This is funded by he Science and Energy Research Board of the the Government of India.