J. Math. Math. Sci. (2026) 75–96

A Bregman projection for split feasibility problem with multiple output sets: A self-adaptive inertial extragradient-type algorithm

Le Xuan Ly (Hanoi, Vietnam)

Dedicated to Professor Do Van Luu on the occasion of his 80th birthday

(Received Mar. 16, 2026; accepted May 8, 2026)

Abstract. The purpose of this paper is to investigate a Bregman projection algorithm for solving split feasibility problem with multiple output sets. The proposed algorithm is motivated by the ideas of the Halpern method, the CQ method and the Tseng method. Our proposed algorithm employs the inertial technique and a self-adaptive step size to guarantee a high rate of convergence. The strong convergence theorem is established without prior knowledge of the operator norm and the Lipschitz continuous assumption on the operators involved. Numerical experiments with graphical illustrations are presented to demonstrate the effectiveness and the performance of our proposed algorithm in comparison with some existing ones.

1. Introduction

Let $\mathcal{H}_j$, $j = 0, 1, \dots, N$, be real Hilbert spaces with the inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\| \cdot \|$. In this paper, we study the split feasibility problem

---

Key words and phrases: Hilbert spaces; split variational inequality problem; inertial term; self-adaptive; Bregman projection.
2020 Mathematics Subject Classification: 49J40; 90C33; 47H05; 47H09
The Project is supported by the the Ministry-level research project of the Ministry of Education and Training (MOET) under grant number **B2026-BKA-02**.

with multiple output sets (SFP-MOS) [17,22,25,28–30],

$$\text{find } u^* \in \mathcal{H}_0 \text{ such that } \mathcal{F}_j u^* \in \mathcal{C}_j,\ \forall j = 0, 1, \ldots, N, \tag{1.1}$$

where $\mathcal{C}_j \subseteq \mathcal{H}_j$, $j = 0, 1, \ldots, N$, are nonempty closed convex subsets, $\mathcal{F}_0 : \mathcal{H}_0 \to \mathcal{H}_0$ is the identity mapping I, and $\mathcal{F}_j : \mathcal{H}_0 \to \mathcal{H}_j$, $j = 1, \ldots, N$, are bounded linear mappings. The solution set of the SFP-MOS is denoted by Ω, that is

$$\Omega := \big\{u^* \in \mathcal{H}_0 \mid \mathcal{F}_j u^* \in \mathcal{C}_j,\ \forall j = 0, 1, \ldots, N\big\}.$$

A special case of the SFP-MOS, when $N = 1$, is the split feasibility problem (SFP):

$$\text{find } u^* \in \mathcal{H}_0 \text{ such that } u^* \in \mathcal{C}_0 \text{ and } \mathcal{F}_1 u^* \in \mathcal{C}_1.$$

Originally, the SFP was introduced in Euclidean spaces [7] and afterward extended to infinite dimensional spaces. The SFP was applied successfully in the field of intensity-modulated radiation therapy treatment planning [8,9]. It plays an important role in medical image reconstruction and in signal processing [3,4].

For each $j = 1, 2, \ldots, N$, let us define the mappings $\mathcal{T}_j : \mathcal{H}_0 \to \mathcal{H}_0$ by

$$\mathcal{T}_j := P_{\mathcal{C}_0}\big[I - \gamma_j \mathcal{F}_j^*(I - P_{\mathcal{C}_j})\mathcal{F}_j\big], \tag{1.2}$$

where $\gamma_j \in \big(0, 2/\|\mathcal{F}_j\|^2\big)$. Then $\mathcal{T}_j$ is a nonexpansive mapping and its set of fixed points $\mathrm{Fix}(\mathcal{T}_j) = \{x \in \mathcal{H}_0 \mid x = \mathcal{T}_j x\}$ coincides with the solution set of the SFP finding $u^* \in \mathcal{C}_0$ such that $\mathcal{F}_j u^* \in \mathcal{C}_j$. Thus, the set of common fixed points of a finite family of nonexpansive operators is coincident with the solution set of the SFP-MOS, i.e.,

$$\bigcap_{j=1}^{N} \mathrm{Fix}(\mathcal{T}_j) \equiv \Omega.$$

A well-known method for solving the SFP is Byrne's CQ algorithm (see [3]). As has already been mentioned by Byrne, a special case of this method was introduced by Landweber [14]. It is known that $u^* \in \mathcal{C}_0$ solves the SFP if and only if u^* solves the fixed point problem:

$$u^* = P_{\mathcal{C}_0}\big(u^* - \gamma \mathcal{F}_1^*(I - P_{\mathcal{C}_1})\mathcal{F}_1 u^*\big),$$

where $P_{\mathcal{C}_0}$ and $P_{\mathcal{C}_1}$ are the metric projections of $\mathcal{H}$ and $\mathcal{H}_1$ onto $\mathcal{C}_0$ and $\mathcal{C}_1$, respectively, I is the identity operator on $\mathcal{H}_0$ or $\mathcal{H}_1$, $\mathcal{F}_1^*$ is the adjoint operator of $\mathcal{F}_1$, and $\gamma \in (0, 2/L)$ with L is the spectral radius of $\mathcal{F}_1^*\mathcal{F}_1$. The weak convergence theorem requires to calculate the spectral norm of the operator

$\mathcal{F}_1$. The implementation of this method depends on the knowledge of the norm of the bounded linear operator. It also should be noted that strong convergence results are much more desirable than weak convergence results in infinite dimensional Hilbert spaces. To guarantee the strong convergence result, some methods, such as the Halpern method, the viscosity approximation method, and the hybrid projection method, can be employed. Some new results on the CQ method to solve the SFP and some related problems can be seen in [24–27,29–32] and the references therein.

In order to improve the convergence rate of the algorithms, inertial acceleration is widely applied [16,17,25,30–32]. It is firstly proposed by Polyak in 1964 [21] for solving the smooth convex minimization problems.

Most of the aforementioned methods employ the norm distances and the induced metric projections. The Bregman distance is an elegant and effective distance function introduced by Bregman in 1976 [5]. It generalizes a wide range of measures, such as the squared Euclidean distance, the Itakura–Saito divergence, and the Kullback–Leibler divergence. The Bregman distance, which is capable of exploring the nonlinear correlation of data features, has found applications in various areas, including machine learning, computational geometry, operations research, and information theory [11,18]. Several methods for solving problem SFP with Bregman projections can be found in [13,15,23].

Motivated by the results above and the ongoing research interest in this direction, we introduce an algorithm based on Bregman projections for solving the SFP-MOS in Hilbert spaces. Our proposed method has key features as follows.

1. The proposed method utilizes the Bregman distance and Bregman projection instead of the standard Euclidean metric. This allows the algorithm to adapt to the specific geometry of the underlying Hilbert spaces, providing a more flexible framework for solving the SFP-MOS under relaxed regularity conditions.

2. The proposed method is combined with one-step inertial techniques to improve its convergence.

3. Implementation of the algorithm does not require the knowledge of operator norms nor the Lipschitz continuous assumption on the operators.

The remaining part of this paper is organized as follows: the next section displays some lemmas that will be used for the validity and convergence of the algorithm. The third section is devoted to describing our proposed algorithms and the strong convergence results. In the next section, an example is provided to illustrate the performance of the algorithm using certain Bregman projections.

2. Preliminaries

In this section, we introduce some mathematical symbols, definitions, and lemmas which can be used in the proof of our main result. Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. In what follows, we write $x^k \rightharpoonup x$ to indicate that the sequence $\{x^k\}$ converges weakly to x while $x^k \to x$ indicates that the sequence $\{x^k\}$ converges strongly to x.

Let $g : \mathcal{H} \to (-\infty, +\infty]$. The function g is said to be

1. strictly convex if $\forall x, y \in \text{dom } g,\ x \neq y$,

$$g[tx + (1-t)y] < tg(x) + (1-t)g(y),\ \forall t \in (0,1);$$

2. uniformly convex with modulus ϕ if $\forall t \in (0,1),\ \forall x, y \in \text{dom } g$,

$$g[tx + (1-t)y] < tg(x) + (1-t)g(y) - t(1-t)\phi(\|x-y\|),$$

where ϕ is an increasing function vanishing only at 0;

3. strongly convex with a constant $\sigma > 0$ if $\forall t \in (0,1),\ \forall x, y \in \text{dom } g$,

$$g[tx + (1-t)y] < tg(x) + (1-t)g(y) - \frac{\sigma}{2}t(1-t)\|x-y\|^2.$$

Note that strong convexity implies strict convexity.

Let $g : \mathcal{H} \to (-\infty, +\infty]$ be a strictly convex and differentiable (on the interior of its domain) function with gradient ∇g and domain

$$\text{dom } g = \{x \in \mathcal{H} \mid g(x) < +\infty\} \neq \emptyset.$$

The Fenchel conjugate function of g is the convex function $g^* : \mathcal{H} \to (-\infty, +\infty]$ defined by

$$g^*(x^*) := \sup_{x \in \mathcal{H}} \{\langle x^*, x \rangle - g(x)\}.$$

Remark 2.1. Let $g^* : \mathcal{H} \to (-\infty, +\infty]$ be Fenchel conjugate function of g. Then $\nabla g^*(\nabla g(x)) = x$ for all $x \in \text{dom } \nabla g$ and $\nabla g(\nabla g^*(x^*)) = x^*$ for all $x^* \in \text{dom } g^*$.

A function $g : \mathcal{H} \to (-\infty, +\infty]$ is called Legendre if it satisfies

1. $\text{int}(\text{dom } g) \neq \emptyset$ and the subdifferential ∂g is single-valued on its domain;

2. $\text{int}(\text{dom } g^*) \neq \emptyset$ and ∂g^* is single-valued on its domain.

The Bregman bifunction (distance) $D_g : \mathrm{dom}\ g \times \mathrm{int}(\mathrm{dom}\ g) \to [0, +\infty)$ corresponding to the strictly convex function g is defined by the formula [5]

$$D_g(x, y) := g(x) - g(y) - \langle \nabla g(y), x - y \rangle, \ \forall x \in \mathrm{dom}\ g, \ \forall y \in \mathrm{int}(\mathrm{dom}\ g).$$

The Bregman divergence is one kind of measurement of the difference between two points (or distributions in statistics) on a differentiable convex function of Legendre type. Note that the Bregman divergence is not a usual metric because it is asymmetric and does not satisfy the triangle inequality. The Bregman divergence with respect to various types of g can be seen as follows (Bauschke et al. [2]; Hieu and Cholamjiak [12]): let $x = (x_1, x_2, \ldots, x_m)^\top$ and $y = (y_1, y_2, \ldots, y_m)^\top$ be two points in $\mathbb{R}^m$.

1. The squared Euclidean divergence

$$D_g^{\mathrm{SE}}(x, y) = \frac{1}{2}\|x - y\|^2$$

generated by the function $g^{\mathrm{SE}}(x) = \frac{1}{2}\|x\|^2$ with its domain $\mathrm{dom}\ g^{\mathrm{SE}} = \mathbb{R}^m$ and its gradient $\nabla g^{\mathrm{SE}}(x) = x$.

2. The Kullback–Leibler divergence

$$D_g^{\mathrm{KL}}(x, y) := \sum_{i=1}^{m} x_i \Big(\log\left(\frac{x_i}{y_i}\right) - 1 \Big) + \sum_{i=1}^{m} y_i.$$

generated by the Shannon's function $g^{\mathrm{KL}}(x) := \sum_{i=1}^{m} x_i \log x_i$ with its domain $\mathrm{dom}\ g^{\mathrm{KL}} = \{x = (x_1, x_2, \ldots, x_m)^\top \in \mathbb{R}^m, \ x_i > 0, \ i = 1, \ldots, m\}$ and its gradient

$$\nabla g^{\mathrm{KL}}(x) = (1 + \log x_1, 1 + \log x_2, \ldots, 1 + \log x_m)^\top.$$

In statistics, the Kullback–Leibler divergence is used to measure the difference between two probability distributions.

3. The squared Mahalanobis divergence

$$D_g^{\mathrm{SM}}(x, y) := \frac{1}{2}(x - y)^\top Q(x - y)$$

generated by the function $g^{\mathrm{SM}}(x) := \frac{1}{2}x^\top Qx$ with its domain $\mathrm{dom}\ g^{\mathrm{SM}} = \mathbb{R}^m$ and its gradient $\nabla g^{\mathrm{SM}}(x) = Qx$, where Q is a positive definite symmetric matrix. The squared Mahalanobis divergence is used to measure the difference between standard deviation and mean in a normal distribution.

The relationship between D_g and norm $\|\cdot\|$ is guaranteed when g is strongly convex with strong convexity constant $\sigma > 0$ (see [33]):

$$D_g(x,y) \geq \frac{\sigma}{2}\|x-y\|^2,\ \forall x \in \text{dom } g,\ y \in \text{int}(\text{dom } g). \tag{2.1}$$

We also have the following three-point identity: $\forall x \in \text{dom } g,\ y,z \in \text{int}(\text{dom } g)$,

$$D_g(x,y) + D_g(y,z) - D_g(x,z) = \langle \nabla g(z) - \nabla g(y), x-y \rangle.$$

Let $\mathcal{C} \subseteq \mathcal{H}$ be a nonempty closed convex subset. The metric projection $P_{\mathcal{C}} : \mathcal{H} \to \mathcal{C}$ is defined by

$$P_{\mathcal{C}}x := \arg\min\{\|x-y\| \mid y \in \mathcal{C}\},\ x \in \mathcal{H}.$$

A fundamental characteristic property (see [1]) is that

$$z = P_{\mathcal{C}}(x) \text{ if and only if } \langle x-z, y-z \rangle \leq 0,\ \forall y \in \mathcal{C}. \tag{2.2}$$

Moreover, we have

$$\|P_{\mathcal{C}}x - y\|^2 \leq \|x-y\|^2 - \|x - P_{\mathcal{C}}x\|^2. \tag{2.3}$$

Throughout this paper, we consider a more general projection, namely the Bregman projection, which is defined as follows: The Bregman projection, with respect to the function g, of a point $z \in \text{int}(\text{dom } g)$ is the unique point in $\mathcal{C}$ defined by

$$P^g_{\mathcal{C}}x := \arg\min\{D_g(y,x) \mid y \in \mathcal{C}\},\ x \in \mathcal{H}.$$

In the special case where $g(x) = \frac{1}{2}\|x\|^2$, we have $D_g(x,y) = \frac{1}{2}\|x-y\|^2$ and thus $P^g_{\mathcal{C}} = P_{\mathcal{C}}$. In general, the Bregman projection $P^g_{\mathcal{C}}$ depends not only on the set $\mathcal{C}$, but also on the function g. Like the metric projection, the Bregman projection also has the following properties, for each $x \in \mathcal{H}$ (see [6]):

$$z = P^g_{\mathcal{C}}(x) \text{ if and only if } \langle \nabla g(x) - \nabla g(z), y-z \rangle \leq 0,\ \forall y \in \mathcal{C} \tag{2.4}$$

and

$$D_g(y, P^g_{\mathcal{C}}x) + D_g(P^g_{\mathcal{C}}x, x) \leq D_g(y,x),\ \forall y \in \mathcal{C}. \tag{2.5}$$

We also know that if $g : \mathcal{H} \to \mathbb{R}$ is uniformly Fréchet differentiable and bounded on bounded subsets of $\mathcal{H}$, then ∇g is uniformly continuous on bounded subsets of $\mathcal{H}$. Moreover, if g is assumed to be σ-strongly convex, Fréchet differentiable and bounded on bounded subsets of $\mathcal{H}$, then for any two sequences $\{x^k\}$ and $\{y^k\}$ in $\mathcal{H}$,

$$\lim_{k\to\infty} D_g(x^k, y^k) = 0 \Longrightarrow \lim_{k\to\infty} \|x^k - y^k\| = 0. \tag{2.6}$$

Let V_g : dom $g \times$ dom $g^* \to [0, +\infty)$ associated with a Legendre function g be defined by

$$V_g(x, x^*) := g(x) - \langle x^*, x \rangle + g^*(x^*), \ \forall x \in \text{dom } g, \ x^* \in \text{dom } g^*.$$

Some properties of the function V_g can be summarized as follows (see [20]):

1. V_g is nonnegative and convex in the second variable;
2. $V_g(x, x^*) = D_g(x, \nabla g^*(x^*)) \ \forall x \in \text{dom } g, \ x^* \in \text{int(dom) } g^*$;
3. $V_g(x, x^*) \le V(x^*+y^*, x) - \langle y^*, \nabla g^*(x^*) - x \rangle \ \forall x \in \text{dom } g, \ x^*, y^* \in \text{dom } g^*$.

Since V_g is convex in the second variable, then, for $N \in \mathbb{N}$

$$D_g\Big(x, \nabla g^*\Big(\sum_{j=1}^{N} \lambda_j \nabla g(y_j)\Big)\Big) \le \sum_{j=1}^{N} \lambda_j D_g(x, y_j), \ \forall x \in \text{dom } g, \tag{2.7}$$

where $y_j \in \mathcal{H}$, $\lambda_j \in [0,1]$ for $j = 1, \dots, N$, and $\sum_{j=1}^{N} \lambda_j = 1$.

The following lemmas are needed to prove a result in the next section.

Lemma 2.1 (see [10]). *Let $\{s_k\}$, $\{a_k\}$, $\{b_k\}$, and $\{c_k\}$ be sequences of nonnegative real numbers such that $s_{k+1} \le (1 - a_k - b_k)s_k + b_k s_{k-1} + a_k c_k$, $\forall k \ge 1$, where $\sum_{k=1}^{\infty} a_k = \infty$, $\{b_k\} \subset [0, 1/2]$, and $\limsup_{k\to\infty} c_k \le 0$. Then $\lim_{k\to\infty} s_k = 0$.*

Lemma 2.2 (see [19]). *Let $\{s_k\}$ be a nonnegative real sequence such that there exists a subsequence $\{k_l\}$ of $\{k\}$ such that $s_{k_l} < s_{k_l+1}$ for all $l \in \mathbb{N}$. Then there exists an increasing sequence $\{\phi(l)\} \subset \mathbb{N}$ such that $\lim_{l\to\infty} \phi(l) = \infty$ and the following properties are satisfied by all (sufficiently large) numbers $m \in \mathbb{N}$: $s_{\phi(m)} < s_{\phi(m)+1}$ and $s_m < s_{\phi(m)+1}$. In fact, $\phi(m) = \max\{j \le m \ | s_j < s_{j+1} \ \}$.*

3. The algorithm and convergence analysis

In this section, we introduce a new algorithm based on the Bregman distance and the Bregman projection for solving the SFP-MOS in Hilbert spaces. Our algorithm is constructed around the following method: Byrne's CQ method, Polyak's gradient method, Halpern method, and hybrid projection method. In order to establish the strong convergence of the algorithm, we make the following assumptions

(A1) $\mathcal{F}_0 : \mathcal{H}_0 \to \mathcal{H}_0$ is the identity operator, $\mathcal{F}_j : \mathcal{H}_j \to \mathcal{H}_j$ is bounded linear operator, $j = 1, \dots, N$.

(A2) The functions $g_j : \mathcal{H}_j \to \mathbb{R}$, $j = 0, 1, \dots, N$ (where $g_0 := g$) are σ_j-strongly convex, Legendre, which is bounded and uniformly Fréchet differentiable on bounded subsets of $\mathcal{H}_j$, $j = 1, 2$.

(A3) $\Omega = \{x^* \in \mathcal{H}_0 \mid \mathcal{F}_j x^* \in C_j \; j = 0, 1, \dots, N\}$, the solution set of SFP-MOS, is a nonempty set.

We require that the control parameters satisfy the following conditions for $k \geq 1$.

(C1) $\{\alpha_k\} \subset (0, 1)$, $\lim\limits_{k\to\infty} \alpha_k = 0$, $\sum_{k=1}^{\infty} \alpha_k = \infty$.

(C2) $\{\eta_k\} \subset (0, \infty)$ such that $\lim\limits_{k\to\infty} \eta_k/\alpha_k = 0$.

Algorithm 3.1. Initialization: Given $\theta \in (0, \infty)$, $\mu \in (0, \infty)$, $\tau \in (0, 1)$, $\nu \in (0, 1)$. Let x^0, $x^1 \in \mathcal{H}_1$ be arbitrary.

Iterative Steps: Given the iterates x^k, x^{k-1} for $k \geq 1$, calculate x^{k+1} as follows.

Step 1. Set $w^k = \nabla g^* \left[\nabla g(x^k) + \theta_k \left(\nabla g(x^{k-1}) - \nabla g(x^k)\right)\right]$, where

$$\theta_k = \begin{cases} \min\left\{\frac{\eta_k}{\|\nabla g(x^{k-1}) - \nabla g(x^k)\|}, \theta\right\}, & \text{if } \nabla g(x^k) \neq \nabla g(x^{k-1}), \\ \theta, & \text{otherwise.} \end{cases} \tag{3.1}$$

Step 2. Compute $y_j^k = \nabla g_j^* \left[\mathcal{F}_j^* \left(\nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P_{C_j}^{g_j} \mathcal{F}_j w^k)\right)\right]$, $j = 0, 1, \dots, N$.

Step 3. Compute $z^k = \nabla g^* \left(\nabla g(w^k) - \gamma_k \sum_{j=0}^{N} \nabla g_j(y_j^k)\right)$, where $\gamma_k = \mu\nu^{\kappa_k}$ with κ_k being the smallest non-negative integer κ satisfying

$$\mu\nu^{\kappa} \sum_{j=0}^{N} D_{g_j}(\mathcal{F}_j z^k, \mathcal{F}_j w^k) \leq \tau D_g(z^k, w^k). \tag{3.2}$$

If $w^k = z^k$, then stop; w^k is a solution of SFP-MOS. Otherwise, go to Step 4.

Step 4. Compute $x^{k+1} = \nabla g^* \left[\alpha_k \nabla g(x^0) + (1 - \alpha_k) \nabla g(z^k)\right]$.

Set $k := k + 1$ and go to Step 1.

The combination of the inertial term θ_k and the self-adaptive step size γ_k in Algorithm 3.1 significantly enhances the performance and stability of the iterative process. Intuitively, the inertial term θ_k acts as a momentum factor that utilizes information from previous iterations to accelerate the convergence rate toward the solution set Ω. Simultaneously, the adaptive step-size γ_k adjusts the movement based on the local geometry of the operators, which eliminates the requirement for prior knowledge of global operator norms $\|F_j\|$. These components work together to balance rapid progress with numerical stability.

Lemma 3.1. The Armijo-line search rules (3.2) is well-defined. Moreover, $\gamma_k \in (0, \mu]$ for all $k \geq 0$.

Proof. Let consider two possible cases.

Case 1. Suppose that $\sum_{j=0}^{N} D_{g_j}(\mathcal{F}_j z^k, \mathcal{F}_j w^k) = 0$. In this case, we obtain that $\kappa_k = 0$ holds.

Case 2. Suppose that $\sum_{j=0}^{N} D_{g_j}(\mathcal{F}_j z^k, \mathcal{F}_j w^k) \neq 0$. In this case, we additionally suppose that $D_g(z^k, w^k) = 0$. Then, $z^k = w^k$ and so $\mathcal{F}_j z^k = \mathcal{F}_j w^k$ for all $j = 1, \ldots, N$. Therefore, $\sum_{j=0}^{N} D_{g_j}(\mathcal{F}_j z^k, \mathcal{F}_j w^k) = 0$. Which is contradiction. Hence, $D_g(z^k, w^k) \neq 0$. We assume that

$$\mu\nu^{\kappa} \sum_{j=1}^{N} D_{g_j}(\mathcal{F}_j z^k, \mathcal{F}_j w^k) > \tau D_g(z^k, w^k) \text{ for any integer } \kappa. \tag{3.3}$$

Since $\mu \in (0, \infty)$, $\nu \in (0, 1)$, it follows from (3.3) that

$$0 = \lim_{\kappa \to \infty} \mu\nu^{\kappa} \sum_{j=1}^{N} D_{g_j}(\mathcal{F}_j z^k, \mathcal{F}_j w^k) > \tau D_g(z^k, w^k) > 0.$$

This is a contradiction. Therefore, there exists a finite nonnegative integer $\tilde{\kappa} \in \mathbb{N}$ such that $\mu\nu^{\tilde{\kappa}} \sum_{j=1}^{N} D_{g_j}(\mathcal{F}_j z^k, \mathcal{F}_j w^k) \leq \tau D_g(z^k, w^k)$. Then (3.2) holds. This implies that $\{\gamma^k\} \in (0, \mu]$.

■

Lemma 3.2. *Let $\{w^k\}$ and $\{z^k\}$ be sequences generated by Algorithm 3.1. If $w^k = z^k$ holds for some integer k, then $w^k \in \Omega$.*

Proof. We assume that $w^k = z^k$. By using Steps 1-3 and Remark 2.1, the assumption can be rewritten as

$$\begin{aligned} \nabla g(w^k) = \nabla g(z^k) &= \nabla g(w^k) - \gamma_k \sum_{j=0}^{N} \nabla g_j(y_j^k) \\ &= \nabla g(w^k) - \gamma_k \sum_{j=0}^{N} \left[\mathcal{F}_j^* \left(\nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k)\right)\right]. \end{aligned} \tag{3.4}$$

It follows from $\gamma_k \in (0, \mu]$ and (3.4) that

$$0 = \sum_{j=0}^{N} \left[\mathcal{F}_j^* \left(\nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k) \right) \right]. \tag{3.5}$$

Let $u \in \Omega$. It follows from (3.5) that

$$0 = \sum_{j=0}^{N} \left\langle \nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k), \mathcal{F}_j w^k - \mathcal{F}_j u \right\rangle. \tag{3.6}$$

From the three-point identity of the Bregman distance and the property (2.4) of the Bregman projection, we have

$$\begin{aligned}
&\langle \nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k), \mathcal{F}_j w^k - \mathcal{F}_j u \rangle \\
&\qquad = D_{g_j}(\mathcal{F}_j u, \mathcal{F}_j w^k) + D_{g_j}(\mathcal{F}_j w^k, P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k) - D_{g_j}(\mathcal{F}_j u, P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k) \\
&\qquad \geq D_{g_j}(\mathcal{F}_j w^k, P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k) \geq 0 \ \forall j = 0, 1, \ldots, N. \qquad (3.7)
\end{aligned}$$

It follows from (3.6) and (3.7) that

$$\left\langle \nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k), \mathcal{F}_j w^k - \mathcal{F}_j u \right\rangle = 0, \ \forall j = 0, 1, \ldots, N. \tag{3.8}$$

From (3.7) and (3.8), we obtain

$$D_{g_j}(\mathcal{F}_j w^k, P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k) = 0 \ \forall j = 0, 1, \ldots, N. \tag{3.9}$$

Hence, (3.9) asserts that $\|\mathcal{F}_j w^k - P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k\| = 0$ for all $j = 0, 1, \ldots, N$. This implies that $\mathcal{F}_j w^k \in \mathcal{C}_j$ for all $j = 0, 1, \ldots, N$. Thus, $w^k \in \Omega$. ■

Remark 3.1. Lemma 3.2 implies that if the iterative sequence generated by Algorithm 3.1 terminates within finite steps, then the current iterative point must be a solution of the SFP-MOS. Without loss of generality, we assume that Algorithm 3.1 generates an infinite iterative sequence for the following convergence analysis.

Lemma 3.3. *Suppose that all conditions* (A1)-(A3), (C1)-(C2) *are satisfied. Let* $\{w^k\}$ *and* $\{z^k\}$ *be two sequences generated by Algorithm 3.1. Then for* $u \in \Omega$*, we have*

$$\begin{aligned}
D_g(u, z^k) &\leq D_g(u, w^k) - D_g(z^k, w^k) \\
&\quad - \gamma_k \sum_{j=0}^{\infty} \left[D_{g_j}(\mathcal{F}_j z^k, P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k) + D_{g_j}(P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k, \mathcal{F}_j w^k) \right] + \tau D_g(z^k, w^k).
\end{aligned}$$

Proof. From the three-point identity of the Bregman distance, we have

$$D_g(u,z^k) = D_g(u,w^k) - D_g(z^k,w^k) + \langle \nabla g(w^k) - \nabla g(z^k), u - z^k\rangle \tag{3.10}$$

and for all $j = 0, 1, \dots, N$,

$$\begin{aligned} &\big\langle \nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k)\,,\, P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k - \mathcal{F}_j z^k\big\rangle \\ &\quad = -D_{g_j}(\mathcal{F}_j z^k, P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k) + D_{g_j}(\mathcal{F}_j z^k, \mathcal{F}_j w^k) - D_{g_j}(P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k, \mathcal{F}_j w^k). \end{aligned} \tag{3.11}$$

From the property (2.4) of the Bregman projection, we have

$$\big\langle \nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k)\,,\, \mathcal{F}_j u - P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k\big\rangle \le 0,\ \forall j = 0, 1, \dots, N. \tag{3.12}$$

Be the definition of y_j^k, it follows that, for all $j = 0, 1, \dots, N$,

$$\begin{aligned} \langle \nabla g_j(y_j^k), u - z^k\rangle &= \big\langle \nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k)\,,\, \mathcal{F}_j u - \mathcal{F}_j z^k\big\rangle \\ &= \big\langle \nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k)\,,\, \mathcal{F}_j u - P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k\big\rangle \\ &\quad + \big\langle \nabla g_j(\mathcal{F}_j w^k) - \nabla g_j(P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k)\,,\, P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k - \mathcal{F}_j z^k\big\rangle. \end{aligned} \tag{3.13}$$

By using the definition of z^k, one has

$$\langle \nabla g(w^k) - \nabla g(z^k), u - z^k\rangle = \gamma_k \Big\langle \sum_{j=0}^{N} \nabla g_j(y_j^k), u - z^k\Big\rangle. \tag{3.14}$$

Substituting the results from (3.11) to (3.14) into (3.10), we obtain that

$$\begin{aligned} D_g(u,z^k) &\le D_g(u,w^k) - D_g(z^k,w^k) \\ &\quad -\gamma_k \sum_{j=0}^{N} \Big(D_{g_j}(\mathcal{F}_j z^k, P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k) + D_{g_j}(P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k, \mathcal{F}_j w^k)\Big) + \tau D_g(z^k,w^k). \end{aligned}$$

This complete the proof. ∎

We are now in a position to prove the strong convergence result of Algorithm 3.1.

Theorem 3.1. *Suppose that all conditions* (A1)-(A3), (C1)-(C2) *are satisfied. Then the sequence* $\{x^k\}$ *generated by Algorithm 3.1 converges strongly to* $u^* \in \Omega$ *with* $u^* = P^g_\Omega(x^0)$.

Proof. Let $u \in \Omega$. By combining the definition of w^k and (2.7), one deduces that

$$D_g(u, w^k) \leq (1-\theta_k) D_g(u, x^k) + \theta_k D_g(u, x^{k-1}). \tag{3.15}$$

This, combined with Lemma 3.3, implies

$$D_g(u, z^k) \leq (1-\theta_k) D_g(u, x^k) + \theta_k D_g(u, x^{k-1}). \tag{3.16}$$

This implies that

$$D_g(u, z^k) \leq \max\big\{D_g(u, x^k), D_g(u, x^{k-1})\big\}. \tag{3.17}$$

By virtue of the definition of x^{k+1}, (2.7), and (3.17), we obtain

$$\begin{aligned} D_g(u, x^{k+1}) &\leq \alpha_k D_g(u, x^0) + (1-\alpha_k) D_g(u, z^k) \\ &\leq \max\big\{D_g(u, x^0), D_g(u, x^k), D_g(u, x^{k-1})\big\} \\ &\ \ \vdots \\ &\leq \max\big\{D_g(u, x^0), D_g(u, x^1)\big\}. \end{aligned} \tag{3.18}$$

This implies that the sequence $\{D_g(u, x^k)\}$ is bounded, which implies that $\{x^k\}$ is bounded too. By using (3.15) and (3.16), one finds that $\{w^k\}$ and $\{z^k\}$ are also bounded. By using Lemma 3.3 and (3.18), we obtain

$$\begin{aligned} &(1-\alpha_k)\Big[(1-\tau) D_g(z^k, w^k) + \gamma_k \sum_{j=0}^{N} \big(D_{g_j}(\mathcal{F}_j z^k, P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k) \\ &\qquad + D_{g_j}(P_{\mathcal{C}_j}^{g_j} \mathcal{F}_j w^k, \mathcal{F}_j w^k)\big)\Big] \\ &\leq\ \alpha_k D_g(u, x^0) + D_g(u, x^k) - D_g(u, x^{k+1}) - \theta_k\big[D_g(u, x^k) - D_g(u, x^{k-1})\big]. \end{aligned} \tag{3.19}$$

Let $u^* = P^g_\Omega(x^0)$. It follows from definition of V_g and x^{k+1} that

(3.20)
$$\begin{aligned}
D_g(u^*, x^{k+1}) &= V_g\big(u^*,\ \alpha_k \nabla_g(x^0) + (1-\alpha_k)\nabla_g(z^k)\big)\\
&\le V_g\big(\alpha_k \nabla_g(x^0) + (1-\alpha_k)\nabla_g(z^k) - \alpha_k\big[\nabla_g(x^0) - \nabla_g(u^*)\big], u^*\big)\\
&\quad + \big\langle \alpha_k\big(\nabla_g(x^0) - \nabla_g(u^*)\big), \nabla g^*\big(\alpha_k\nabla_g(x^0) + (1-\alpha_k)\nabla_g(z^k)\big) - u^*\big\rangle\\
&= V_g\big(\alpha_k \nabla_g(u^*) + (1-\alpha_k)\nabla_g(z^k), u^*\big) + \alpha_k\big\langle \nabla_g(x^0) - \nabla_g(u^*), x^{k+1} - u^*\big\rangle\\
&= D_g\big(u^*, \nabla g^*\big[\alpha_k\nabla_g(x^0) + (1-\alpha_k)\nabla_g(z^k)\big]\big)\\
&\quad + \alpha_k\big\langle \nabla_g(x^0) - \nabla_g(u^*), x^{k+1} - u^*\big\rangle.\\
&\le (1-\alpha_k) D_g(u^*, z^k) + \alpha_k\big\langle \nabla_g(x^0) - \nabla_g(u^*), x^{k+1} - u^*\big\rangle\\
&\le (1-\alpha_k) D_g(u^*, x^k) - (1-\alpha_k)\,\theta_k\big(D_g(u^*, x^k) - D_g(u^*, x^{k-1})\big)\\
&\quad + \alpha_k\big\langle \nabla_g(x^0) - \nabla_g(u^*), x^{k+1} - u^*\big\rangle\\
&= \big[1 - \alpha_k - (1-\alpha_k)\theta_k\big] D_g(u^*, x^k) + (1-\alpha_k)\,\theta_k D_g(u^*, x^{k-1})\\
&\quad + \alpha_k\big\langle \nabla_g(x^0) - \nabla_g(u^*), x^{k+1} - u^*\big\rangle.
\end{aligned}$$

Now, we consider the following two possible cases to prove $\lim\limits_{k\to\infty} D_g(u^*, x^k) = 0$.

Case 1. There exists an integer $k_1 \in \mathbb{N}$ such that $D_g(u, x^{k+1}) \le D_g(u, x^k)$ for all $k \ge k_1$, which gives that $\{D_g(u^*, x^k)\}$ is convergent and

$$\lim_{k\to\infty} \big(D_g(u, x^k) - D_g(u, x^{k+1})\big) = 0. \tag{3.21}$$

Using (3.19), (3.21), and condition (A2), we get

$$\lim_{k\to\infty} D_g(z^k, w^k) = \lim_{k\to\infty} D_{g_j}(P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^k, \mathcal{F}_j w^k) = 0,\ \forall j = 0, 1, \ldots, N. \tag{3.22}$$

By using the definition of $\{w^k\}$ and combining (2.7) with (3.1), one has

(3.23)
$$D_g(x^k, w^k) \le (1-\theta_k) D_g(x^k, x^k) + \theta_k D_g(x^k, x^{k-1}) = \theta_k D_g(x^k, x^{k-1}) \le \eta_k.$$

By using condition (C2), it follows from (3.23) that

$$\lim_{k\to\infty} D_g(x^k, w^k) = 0. \tag{3.24}$$

By using the definition of x^{k+1}, (2.7), and $\lim\limits_{k\to\infty}\alpha_k = 0$, we have

$$D_g(z^k, x^{k+1}) \le \alpha_k D_g(z^k, x^0) + (1-\alpha_k) D_g(z^k, z^k) \to 0 \text{ as } k \to \infty. \tag{3.25}$$

It follows from (2.6), (3.22), (3.24), and (3.25), we have

$$\lim_{k\to\infty}\|z^k - w^k\| = \lim_{k\to\infty}\|x^k - w^k\| = \lim_{k\to\infty}\|z^k - x^{k+1}\| = 0. \tag{3.26}$$

It follows from (3.26) that

$$\|x^k - x^{k+1}\| \le \|x^k - w^k\| + \|w^k - z^k\| + \|z^k - x^{k+1}\| \to 0 \text{ as } k \to \infty. \tag{3.27}$$

Since $\{x^k\}$ is bounded, there exists a subsequence $\{x^{k_l}\}$ of $\{x^k\}$ such that converges weakly to $\widehat{u} \in \mathcal{H}_0$ and

$$\limsup_{k\to\infty}\langle x^k - u^*, \nabla g(x^0) - \nabla g(u^*)\rangle = \lim_{l\to\infty}\langle x^{k_l} - u^*, \nabla g(x^0) - \nabla g(u^*)\rangle. \tag{3.28}$$

By (3.24), one finds that the subsequence $\{w^{k_l}\}$ of $\{w^k\}$ also converges weakly to $\widehat{u} \in \mathcal{H}_0$. This together with (3.22) implies that $\mathcal{F}_j\widehat{u} \in \mathcal{C}_j$ for all $j = 0, 1, \ldots, N$, that is $\widehat{u} \in \Omega$. Since $u^* = P^g_\Omega(x^0)$, by applying (2.4) with (3.28), one finds that

$$\limsup_{k\to\infty}\langle x^k - u^*, \nabla g(x^0) - \nabla g(u^*)\rangle = \langle \widehat{u} - u^*, \nabla g(x^0) - \nabla g(u^*)\rangle \le 0. \tag{3.29}$$

It follows from (3.27) and (3.29) that

$$\begin{aligned}\limsup_{k\to\infty}\langle x^{k+1} - u^*, \nabla g(x^0) - \nabla g(u^*)\rangle &\le \limsup_{k\to\infty}\langle x^{k+1} - x^k, \nabla g(x^0) - \nabla g(u^*)\rangle \\ &\quad + \limsup_{k\to\infty}\langle x^k - u^*, \nabla g(x^0) - \nabla g(u^*)\rangle.\end{aligned} \tag{3.30}$$

By using Lemma 2.1, (C3), (3.20), and (3.29), one has that $\lim_{k\to\infty} D_g(u^*, x^k) = 0$. This together with (2.6) gives that $\lim_{k\to\infty}\|x^k - u^*\| = 0$.

Case 2. There exists a subsequence $\{D_g(u^*, x^{k_l})\}$ of $\{D_g(u^*, x^k)\}$ such that $D_g(u^*, x^{k_l}) \le D_g(u^*, x^{k_l+1})$ for all $l \in \mathbb{N}$. By applying Lemma 2.2, we see that there exists an increasing sequence $\{\phi(l)\} \subset \mathbb{N}$ such that $\lim_{l\to\infty}\phi(l) = \infty$ and the following inequalities hold, for any $l \in \mathbb{N}$

$$D_g(u^*, x_{\phi(l)}) \le D_g(u^*, x_{\phi(l)+1}) \text{ and } D_g(u^*, x_l) \le D_g(u^*, x_{\phi(l)+1}). \tag{3.31}$$

In view of (3.19), we obtain that

$$\begin{aligned}&(1 - \alpha_{\phi(l)})\big[(1-\tau)D_g(z^{\phi(l)}, w^{\phi(l)}) + \gamma_{\phi(l)}\sum_{j=0}^N \big(D_{g_j}(\mathcal{F}_j z^{\phi(l)}, P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^{\phi(l)}) \\ &\qquad + D_{g_j}(P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^{\phi(l)}, \mathcal{F}_j w^{\phi(l)})\big)\big] \le \alpha_k D_g(u, x^0) + D_g(u, x^{\phi(l)}) \\ &\qquad - D_g(u, x^{\phi(l)+1}) - \theta_k\big(D_g(u, x^{\phi(l)}) - D_g(u, x^{\phi(l)-1})\big).\end{aligned} \tag{3.32}$$

It follows from (3.32) that

$$\lim_{l\to\infty} D_{g_j}(P^{g_j}_{\mathcal{C}_j}\mathcal{F}_j w^{\phi(l)}, \mathcal{F}_j w^{\phi(l)}),\ \forall j = 0,1,\dots,N. \tag{3.33}$$

By repeating the same arguments as in the proof of Case 1, we conclude that there exists a subsequensce of $\{x^{\phi(l)}\}$ converges weakly to some $\widehat{u} \in \Omega$ and

$$\limsup_{k\to\infty}\langle x^{\phi(l)+1} - u^*, \nabla g(x^0) - \nabla g(u^*)\rangle \le 0. \tag{3.34}$$

It follows from (3.20) and (3.31) that

$$\begin{aligned} D_g(u^*, x^{\phi(l)+1}) & \\ \le (1-\alpha_{\phi(l)})D_g(u^*, x^{\phi(l)+1}) &+ \alpha_{\phi(l)}\langle x^{\phi(l)+1} - u^*, \nabla g(x^0) - \nabla g(u^*)\rangle. \end{aligned} \tag{3.35}$$

By combining (3.35) with (3.31), we find that

$$D_g(u^*, x^l) \le D_g(u^*, x^{\phi(l)}) \le \langle x^{\phi(l)+1} - u^*, \nabla g(x^0) - \nabla g(u^*)\rangle. \tag{3.36}$$

It follows from (3.34) and (3.36) that $\limsup_{l\to\infty} D_g(u^*, x^l) = 0$. Thus, $\lim_{l\to\infty}\|x^l - u^*\| = 0$. ∎

Remark 3.2. 1. If $N = 1$, then Algorithm 3.1 can be reduced to the more general form of solving the SFP: given parameters $\theta \in (0,\infty)$, $\mu \in (0,\infty)$, $\tau \in (0,1)$, $\nu \in (0,1)$, α_k satisfying (C1), η_k satisfying (C2), and $x^0, x^1 \in \mathcal{H}_0$, the iterative sequence $\{x^k\}$ is generated by the following:

$$\begin{aligned}
w^k &= \nabla g^*\left[\nabla g(x^k) + \theta_k\left(\nabla g(x^{k-1}) - \nabla g(x^k)\right)\right],\\
&\quad \text{where } \theta_k = \begin{cases} \min\left\{\frac{\eta_k}{\|\nabla g(x^{k-1}) - \nabla g(x^k)\|}, \theta\right\}, & \text{if } \nabla g(x^k) \ne \nabla g(x^{k-1}),\\ \theta, & \text{otherwise.}\end{cases}\\
y_0^k &= \nabla g^*\left[\mathcal{F}_0^*\left(\nabla g(\mathcal{F}_0 w^k) - \nabla g(P^g_{\mathcal{C}_0}\mathcal{F}_0 w^k)\right)\right],\\
y_1^k &= \nabla g_1^*\left[\mathcal{F}_1^*\left(\nabla g_1(\mathcal{F}_1 w^k) - \nabla g_1(P^{g_1}_{\mathcal{C}_1}\mathcal{F}_1 w^k)\right)\right].\\
z^k &= \nabla g^*\left(\nabla g(w^k) - \gamma_k\nabla g(y_0^k) - \gamma_k\nabla g_1(y_1^k)\right), \text{where } \gamma_k = \mu\nu^{\kappa_k},\\
&\quad \kappa_k \text{ being the smallest non-negative integer } \kappa \text{ satisfying}\\
&\quad \mu\nu^\kappa\left[D_g(\mathcal{F}_0 z^k, \mathcal{F}_0 w^k) + D_{g_1}(\mathcal{F}_1 z^k, \mathcal{F}_1 w^k)\right] \le \tau D_g(z^k, w^k).
\end{aligned}$$

If $w^k = z^k$, then stop; w^k is a solution of SFP-MOS.

$$x^{k+1} = \nabla g^*\left[\alpha_k\nabla g(x^0) + (1-\alpha_k)\nabla g(z^k)\right].$$

2. By setting $g_j(x) = \frac{1}{2}\|x\|^2$ for all $x \in \mathcal{H}_j$, $j = 0,\dots,N$, we obtain a special case of Algorithm 3.1., given parameters $\theta \in (0,\infty)$, $\mu \in (0,\infty)$,

$\tau \in (0,1)$, $\nu \in (0,1)$, α_k satisfying (C1), η_k satisfying (C2), and $x^0, x^1 \in \mathcal{H}_0$, the iterative sequence $\{x^k\}$ is generated by the following:

$$
\begin{aligned}
& w^k = x^k + \theta_k(x^{k-1} - x^k), \\
& \quad \text{where } \theta_k = \begin{cases} \min\left\{\frac{\eta_k}{\|x^{k-1}-x^k\|}, \theta\right\}, \text{if } x^{k-1} \neq x^k, \\ \theta, \qquad\qquad\qquad\qquad \text{otherwise.} \end{cases} \\
& y_j^k = \mathcal{F}_j^*\left(\mathcal{F}_j w^k - P_{\mathcal{C}_j}(\mathcal{F}_j w^k)\right), \forall j = 0, \ldots, N. \\
& z^k = w^k - \gamma_k \sum_{j=0}^{N} y_j^k, \text{where } \gamma_k = \mu\nu^{\kappa_k}, \\
& \quad \kappa_k \text{ being the smallest non-negative integer } \kappa \text{ satisfying} \\
& \quad \mu\nu^{\kappa} \sum_{j=0}^{N} \|\mathcal{F}_j z^k - \mathcal{F}_j w^k\|^2 \leq \tau\|z^k - w^k\|^2. \\
& \text{If } w^k = z^k, \text{ then stop; } w^k \text{is a solution of SFP-MOS.} \\
& x^{k+1} = \alpha_k x^0 + (1-\alpha_k) z^k.
\end{aligned}
$$

4. Numerical illustrations

In this section, an illustrative example is provided to show that choosing a suitable Bregman distance can significantly improve the performance of the algorithm for the SFP MOS problem. In the following experiment, we define

$$\text{TOL}(k) := \frac{1}{N+1}\sum_{j=0}^{N} \|\mathcal{F}_j x^k - P_{\mathcal{C}_j}(\mathcal{F}_j x^k)\|^2 \text{ for all } k \geq 1.$$

We use the stopping criterion $\text{TOL}(k) < \epsilon$ for the iterative process, where ϵ is the predetermined error. If $\text{TOL}(k) = 0$, then $w^k \in \Omega$. The source code was developed in MATLAB R2023a and executed on a laptop with an Intel(R) Core(TM) i7-13620H CPU @ 2.4GHz and 16 GB of RAM.

We consider an example in finite dimensional spaces.

Example 4.1. Let $\mathcal{H}_j = \mathbb{R}^2$. The closed convex set $\mathcal{C}_j \subseteq \mathcal{H}_j$ is defined as follows

$$\mathcal{C}_j = \{x = (x_1, x_2)^\top \in \mathbb{R}^2 \ \mid 1 \leq x_1, x_2 \leq 1 + \frac{1}{j+1}\}, \ j = 0, 1, \ldots, N.$$

The operator $\mathcal{F}_j : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $\mathcal{F}_j x = (x_1 + \frac{1}{j+1} x_2, x_2)^\top$ for all $x = (x_1, x_2)^\top \in \mathbb{R}^2$. This is a bounded linear operator. Moreover, we set $g_j(x) = g(x)$ for all $j = 0, 1, \ldots, 5$.

In this example, Algorithm 3.1 is employed for three specific cases of the Bregman distance as follows:

(1) Using the squared Euclidean distance is defined as

$$D_g^{\mathrm{SE}}(x, y) := \frac{1}{2} \sum_{i=1}^{2} (x_i - y_i)^2,$$

which is generated by $g^{\mathrm{SE}}(x) := \frac{1}{2} \sum_{i=1}^{2} x_i^2$ with dom $g^{\mathrm{SE}} = \mathbb{R}^2$. Therefore

$$\begin{aligned}
&\nabla g^{\mathrm{SE}}(x) = (x_1, x_2)^\top = x; \\
&\left(\nabla g^{\mathrm{SE}}\right)^* (y) = \left(\nabla g^{\mathrm{SE}}\right)^{-1} (y) = (y_1, y_2)^\top = y; \\
&P_{\mathcal{C}_j}^g x = \left(\max \left\{ 1, \min \left\{ x_1, 1 + \frac{1}{j+1} \right\} \right\}, \max \left\{ 1, \min \left\{ x_2, 1 + \frac{1}{j+1} \right\} \right\} \right)^\top .
\end{aligned}$$

(2) Using the Kullback–Leibler divergence is defined as

$$D_g^{\mathrm{KL}}(x, y) := \sum_{i=1}^{2} x_i \Big(\log \left(\frac{x_i}{y_i} \right) - 1 \Big) + \sum_{i=1}^{2} y_i,$$

generated by function $g^{\mathrm{KL}}(x) := \sum_{i=1}^{2} x_i \log x_i$ with its domain dom $g^{\mathrm{KL}} = \{x = (x_1, x_2)^\top \in \mathbb{R}^2,\ x_i > 0,\ i = 1, 2\}$.

Therefore

$$\begin{aligned}
&\nabla g^{\mathrm{KL}}(x) = (1 + \log x_1, 1 + \log x_2)^\top; \\
&\left(\nabla g^{\mathrm{KL}}\right)^* (y) = \left(\nabla g^{\mathrm{KL}}\right)^{-1} (y) = (\exp(y_1 - 1), \exp(y_2 - 1))^\top; \\
&P_{\mathcal{C}_j}^g x = \left(\max \left\{ 1, \min \left\{ x_1, 1 + \frac{1}{j+1} \right\} \right\}, \max \left\{ 1, \min \left\{ x_2, 1 + \frac{1}{j+1} \right\} \right\} \right)^\top .
\end{aligned}$$

(3) Using the squared Mahalanobis divergence is defined as

$$D_g^{\mathrm{SM}}(x, y) := \frac{1}{2} (x - y)^\top Q (x - y),$$

generated by $g^{\mathrm{SM}}(x) := \frac{1}{2}x^\top Qx$ with its domain dom $g^{\mathrm{SM}} = \mathbb{R}^2$, where Q is defined as $Q = \begin{bmatrix} 1 & 0 \\ 0 & 0.2 \end{bmatrix}$. Therefore

$$\nabla g^{\mathrm{SM}}(x) = Qx;$$

$$\left(\nabla g^{\mathrm{SM}}\right)^* (y) = \left(\nabla g^{\mathrm{SM}}\right)^{-1} (y) = Q^{-1}y;$$

$$P^g_{\mathcal{C}_j} x = \left(\max\left\{1, \min\left\{x_1, 1 + \frac{1}{j+1}\right\}\right\}, \max\left\{1, \min\left\{x_2, 1 + \frac{1}{j+1}\right\}\right\}\right)^\top .$$

For the numerical assessment of our approach, we employ the following parameters:

$$N = 5,\ \alpha_k = \frac{1}{(k+1)},\ \eta_k = \frac{1}{(k+1)^2},\ \mu = 0.3,\ \tau = 0.5,\ \nu = 0.6.$$

The numerical results are reported in Tables 1 and Figures 1 for $\epsilon = 10^{-6}$, where Iter. denotes the number of iterations and CPU Time denotes the computing time.

Table 1: Numerical results with different choices of Bregman projections

	Iter. (k)	CPU Time (s)
Squared Euclidean	13117	0.3968
Kullback–Leibler	11755	0.5269
Squared Mahalanobis	6167	0.2876

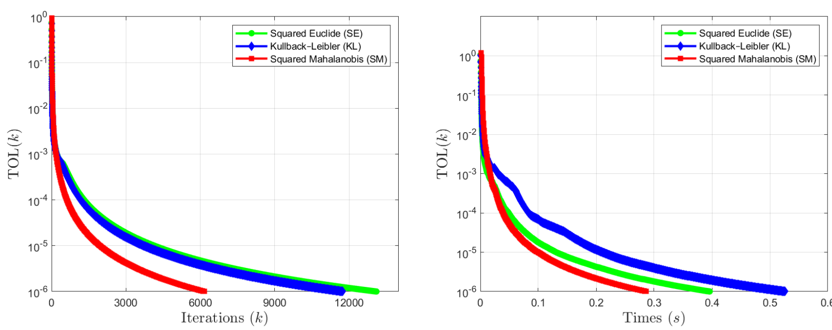

Figure 1: Performance of the algorithm with different choices of Bregman projections

The numerical results in Table 1 and Figure 1 reveal that the squared Mahalanobis (SM) divergence significantly outperforms the SE and KL divergences. From a geometric perspective, this is attributed to the SM divergence's ability to reshape the metric via the matrix Q to better align with the specific scaling of the operators F_j and sets C_j. This adaptation allows the algorithm to take more effective steps toward the solution set Ω.

5. Conclusion

Utilizing Bregman projections, we have proposed a new CQ inertial algorithm for solving the split feasibility problem with multiple output sets in a real Hilbert space. We have proven a theorem to show that the generated iterates by our scheme are strongly convergent. To show the efficiency of our algorithm, we have presented a comparative numerical example between some Bregman projections.

References

[1] **Bauschke, H.H., Combettes, P.L.,** Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Berlin: Springer (2011).

[2] **Bauschke, H.H., Wang, X., Ye, J., Yuan, X.,** *Bregman distances and Chebyshev sets,* J. Approx. Theory 159: 3–25 (2009).

[3] **Byrne, C.,** *Iterative oblique projection onto convex sets and the split feasibility problem,* Inverse Probl. 18: 441–453 (2002).

[4] **Byrne, C.,** *A unified treatment of some iterative algorithms in signal processing and image reconstruction,* Inverse Probl. 20: 103–120 (2004).

[5] **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,* USSR Comput. Math. Math. Phys. 7: 200–217 (1967).

[6] **Butnariu, D., Resmerita, E.,** *Bregman distances, totally convex functions, and a method for solving operator equations in Banach spaces,* Abstr. Appl. Anal. 2006, 84919 (2006).

[7] **Censor, Y., Elfving, T.,** *A multiprojection algorithm using Bregman projections in product space,* Numer. Algorithms. 8, 221–239 (1994).

[8] **Censor, Y., Elfving, T., Kopf, N., Bortfeld, T.,** *The multiple-sets split feasibility problem and its application,* Inverse Probl. 21(6):2071–2084 (2005).

[9] **Censor, Y., Bortfeld, T., Martin, B., Trofimov, A.,** *A unified approach for inversion problems in intensity-modulated radiation therapy,* Phys. Med. Biol. 51:2353–2365 (2006).

[10] **Chbani, Z., Riahi, H.,** *Weak and strong convergence of an inertial proximal method for solving Ky Fan minimax inequalities,* Optimization Letters, 7: 185–206 (2013).

[11] **Hayashi, M.,** *Bregman Divergence Based Em Algorithm and its Application to Classical and Quantum Rate Distortion Theory,* IEEE Transactions on Information Theory, 69(6), June (2023).

[12] **Hieu, D.V., Cholamjiak, P.,** *Modified extragradient method with Bregman distance for variational inequalities,* Appl. Anal. https://doi.org/10.1080/00036811.2020.1757078 (2020).

[13] **Hieu, D. V., Reich, S.,** *Two Bregman projection methods for solving variational inequalities,* Optimization 71, 1777–1802 (2022).

[14] **Landweber, L.,** *An iterative formula for Fredholm integral equations of the first kind,* Am. J. Math. 73: 615–624 (1951).

[15] **Liu, L., Cho, S.Y.,** *A Bregman Projection Algorithm with Self-Adaptive Step Sizes for Split Variational Inequality Problems Involving Non-Lipschitz Operators,* J. Nonlinear Var. Anal. 8, No. 3, 396-417 (2024).

[16] **Ly, L.X., Thuy, N.T.T., Anh, N.Q.,** *Relaxed two-step inertial method for solving a class of split variational inequality problems with application to traffic analysis,* Optimization. doi: 10.1080/02331934.2025.2459191 (2025).

[17] **Ly, L.X., Thuy, N.T.T., Tung, T.T.,** *Two novel two-step inertial algorithms for a class of bilevel variational inequalities,* Optimization, doi: 10.1080/02331934.2024.2345763 (2024).

[18] **Lohit, H., Kumar, D.,** *Modified total Bregman divergence driven picture fuzzy clustering with local information for brain MRI image segmentation,* Applied Soft Computing, 144: 110460 (2023).

[19] **Mainge, P. E.,** *Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization,* Set-valued Anal. 16, 899–912 (2008).

[20] **Marquez, V. M., Reich, S., Sabach, S.,** *Iterative methods for approximating fixed points of Bregman nonexpansive operators,* Discret. Contin. Dyn. Syst. Ser. S. 6, 1043–1063 (2013).

[21] **Polyak, B.T.,** *Some methods of speeding up the convergence of iteration methods,* USSR Comput. Math. Math. Phys. 4:1–17 (1964).

[22] **Reich, S., Truong, M.T., Mai, T.N.H.,** *The split feasibility problem with multiple output sets in Hilbert spaces,* Optim. Lett. 14, 2335–2353 (2020).

[23] **Sunthrayuth, P., Jolaoso, L. O., Cholamjiak, P.,** *New Bregman projection methods for solving pseudo-monotone variational inequality problem,* J. Appl. Math. Comput. 68, 1565–1589 (2022).

[24] **Thuy, N.T.T., Nghia, N.T.,** *A parallel algorithm for generalized multiple-set split feasibility with application to optimal control problems,* Taiwan J Math.; 26(5): 1069–1092 (2022).

[25] **Thuy, N.T.T., Nghia, N.T.,** *A new iterative method for solving the multiple-set split variational inequality problem in Hilbert spaces,* Optimization, 72: 1549–1575 (2023).

[26] **Thuy, N.T.T., Nghia, N.T.,** *A hybrid projection method for solving the multiple-sets split feasibility problem,* Comput. Appl. Math.; 42(6): 292 (2023).

[27] **Thuy, N.T.T., Nghia, N.T.,** *An inertial-type algorithm for a class of bilevel variational inequalities with the split variational inequality problem constraints,* Optimization, doi: 10.1080/02331934 (2023).

[28] **Thuy, N.T.T., Nghia, N.T.,** *Some novel inertial ball-relaxed CQ algorithms for solving the split feasibility problem with multiple output sets,* Journal of Applied Analysis and Computation ;14(3): 1485–1507 (2023).

[29] **Thuy, N.T.T., Tung, T.T.,** *A self adaptive inertial algorithm for solving variational inequalities over the solution set of the split variational inequality problem,* Optimization Letters. 18, 1619–1645 (2024).

[30] **Thuy, N.T.T., Tung, T.T.,** *Two relaxed CQ methods for the split feasibility problem with multiple output sets,* Bulletin of the Malaysian Mathematical Sciences Society ; 47(68) (2024).

[31] **Thuy, N.T.T., Tung, T.T.,** *Strong convergence of one-step inertial algorithm for a class of bilevel variational inequalities,* Optimization, doi: 10.1080/02331934.2024.2448737 (2025).

[32] **Thuy, N.T.T., Tung, T.T., Ly, L.X.,** *Strongly convergent two-step inertial algorithm for a class of bilevel variational inequalities,* Comp. Appl. Math. 44:117 (2025).

[33] **Wega, G.B., Zegeye, H.,** *Convergence results of forward-backward method for a zero of the sum of maximally monotone mappings in Banach spaces,* Comput. Appl. Math. 39(3): 223. doi: 10.1007/s40314-020-01246-z (2020).

Le Xuan Ly
Faculty of Mathematics and Informatics,
Hanoi University of Science and Technology
Hanoi
Vietnam
ly.lexuan@hust.edu.vn