Random projection method for stochastic split feasibility problems

Le Hai Yen
Institute of Mathematics, VAST
18 Hoang Quoc Viet Road, 10307 Hanoi, Vietnam
Email: lhyen@math.ac.vn

August 2, 2019

Abstract: We focus on the multiple-sets split feasibility problem of two arbitrary (possibly infinite) collections of closed convex sets. Under some conditions, it can be reformulated as a stochastic optimization problem. We propose a class of random projection algorithms and prove the almost sure convergence of these algorithms. We also provided convergence rates and some numerical experiments to illustrate the behavior of the algorithms.

1 Introduction

Consider the classical multiple-sets split feasibility problem (MSF)

$$\text{Find } x \in \cap_{i=1}^{t} C_i \text{ such that } Ax \in \cap_{j=1}^{r} Q_j, \qquad (MSF)$$

where A is a given real $m \times n$ matrix, $C_1, \dots, C_t$, $Q_1, \dots, Q_r$ are closed convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. The MSFP was firstly defined by Censor et al in [6] for modeling many pratical applications especially intensity- modulated radiation therapy. It also generalizes both the convex feasibility problem and the split feasibility problem. When $Q_j \equiv \mathbb{R}^m$ for all j, the (MSF) problem becomes the convex feasibility problem ([1, 7])

$$\text{Find } x \in \cap_{i=1}^{t} C_i,$$

and when $t = r = 1$, it becomes the split feasibility problem ([3, 5, 9])

$$\text{Find } x \in C \text{ such that } Ax \in Q.$$

Optimization problems involving a large number of constraints appear more and more in the pratical applications such as inverse problems, computer science, machine learning and statistics (see [11] and the references therein). The convex feasibility problem of a (possibly infinite) collections of closed convex sets also called stochastic feasibility problem was firstly considered in [2] and then formulated as a stochastic optimization problem in [10]. Also in [10], the authour proposed a random projection algorithm and studied its convergence rate for stochastic feasibility problem . In [12], the authours proposed several stochastic reformulations and develop a general projection algorithm for the stochastic convex feasibility problem that can be paralleled. Recently, the stochastic fixed point problem has been investigated in [8]. Motivated by these works, we are interested in the stochastic split feasiblity problem (SSF)

$$\text{Find } x \in \cap_{i\in\mathcal{I}} C_i \text{ such that } Ax \in \cap_{j\in\mathcal{J}} Q_j, \qquad (SSF)$$

where A is a given real $m \times n$ matrix and $\{C_i\}_{i\in\mathcal{I}}, \{Q_j\}_{j\in\mathcal{J}}$ are arbitrary collections of closed convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Comparing to the clasical (MSF) problem, in (SSF), the sets I and J may be infinite. In the next section, we reformulate the (SSF) problem as a stochastic optimization problem and study the equivalence of these problems. In Section 3, we propose a random projection algorithm for solving the (SSF) and study its convergence analysis. Numerical experimental results are provided in Section 4.

2 Problem formulation

Let C be a closed convex set in $\mathbb{R}^n$. We denote by P_C the projection on C. In the following lemma, we recall some important properties of P_C that will be useful for the next part of the paper.

Lemma 2.1. *(see for example [4, 6])*

(i) P_C is firmly non-expansive,i.e. for all $x, y \in \mathbb{R}^n$

$$\|P_C(x) - P_C(y)\|^2 \le \langle x - y, P_C(x) - P_C(y)\rangle.$$

(ii) $I - P_C$ is firmly non-expansive,i.e. for all $x, y \in \mathbb{R}^n$

$$\|(I - P_C)(x) - (I - P_C)(y)\|^2 \le \langle x - y, (I - P_C)(x) - (I - P_C)(y)\rangle.$$

Now, we consider the stochastic split feasibility problem:

$$\text{Find } x \in \cap_{i\in\mathcal{I}} C_i \text{ such that } Ax \in \cap_{j\in\mathcal{J}} Q_j, \tag{SSF}$$

where A is a given real $m \times n$ matrix and $\{C_i\}_{i\in\mathcal{I}}, \{Q_j\}_{j\in\mathcal{J}}$ are finite or infinite collections of closed convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively.

Problem (SSF) can be reformulated as the following stochastic optimization problem:

$$\min_{x\in\mathbb{R}^n} \mathcal{F}(x) = \frac{1}{2}\mathbf{E}\left[\|x - P_{C_\omega}(x)\|^2 + \beta\|Ax - P_{Q_\theta}(Ax)\|^2\right], \tag{SOP}$$

where β is an arbitrary positive number, $\omega \sim \mathcal{P}$, $\theta \sim \mathcal{Q}$, $\mathcal{P}$ is a probability distribution over $\mathcal{I}$, $\mathcal{Q}$ is a probability distribution over $\mathcal{J}$ and the expectation is taken with respect to ω, θ.

We denote the solution set of Problem (SSF) by S, and the solution set of Problem (SOP) by S_1. It is clear that a solution of (SSF) is also a solution of (SOP), *i.e.* $S \subset S_1$, but the inverse inclusion is not always true, for example, when the random variable ω takes only one value in the set $\mathcal{I}$ or the random variable θ takes only one value in the set $\mathcal{J}$.

Lemma 2.2. *Assume that $S \neq \emptyset$, then (SSF) and (SOP) are equivalent, i.e. the solution set of (SSF) equals the solution set of (SOP) if one of the following conditions holds:*

(i) $\mathcal{P}\{\omega = i\} > 0$ for any $i \in \mathcal{I}$ and $\mathcal{Q}\{\theta = j\} > 0$ for any $j \in \mathcal{J}$.

(ii) Linear regularity condition: There exists $\kappa < \infty$ such that

$$dist_S^2(x) \leq \kappa\mathcal{F}(x) \quad \forall x \in \mathbb{R}^n. \tag{1}$$

Proof. (i) Let $x \in S$, then $x \in C_i$ and $Ax \in Q_j$ for any $i \in \mathcal{I}, j \in \mathcal{J}$. Since ω and θ are random variables taking values in the set $\mathcal{I}, \mathcal{J}$, respectively, we have $x = P_{C_\omega}(x), Ax = P_{Q_\theta}(Ax)$ or $\mathcal{F}(x) = 0$. Therefore, $S \subset S_1$.

Now, let $x \in S_1$, we have $\mathcal{F}(x) = 0$. For any $i \in I$,

$$0 = \mathcal{F}(x) \geq \|x - P_{C_i}(x)\|^2\mathcal{P}\{\omega = i\}.$$

But $\mathcal{P}\{\omega = i\} > 0$, then $x = P_{C_i}(x)$ or $x \in C_i$. Similarly, we have $Ax \in Q_j$ for any $j \in \mathcal{J}$. It means that $x \in S$.

(ii) As proved in (i), $S \subset S_1$. If $x \in S_1$, we have

$$dist_S^2(x) \leq \kappa\mathcal{F}(x) = 0.$$

So, $x \in S$.

□

Remark 2.3. *Condition (i) is similar to the condition used in [10] and Condition (ii) (Linear regularity condition) was used in several works ([8, 10, 12]). Note that the linear regularity condition is quite conservative and does not hold for any collection of closed convex sets (see Example 1, [12]).*

Let

$$F(x,\omega,\theta) = \frac{1}{2}\left[\|x - P_{C_\omega}(x)\|^2 + \beta\|Ax - P_{Q_\theta}(Ax)\|^2\right], \tag{2}$$

then

$$\mathcal{F}(x) = \mathbf{E}\left[F(x,\omega,\theta)\right], \tag{3}$$

and

$$\begin{aligned} \nabla\mathcal{F}(x) &= \mathbf{E}(\nabla_x F(x,\omega,\theta)) & (4)\\ &= x - \mathbf{E}\left[P_{C_\omega}(x)\right] + \beta A^T Ax - \beta A^T\mathbf{E}\left[P_{Q_\theta}(Ax)\right]. & (5) \end{aligned}$$

Lemma 2.4. *(i) For each $\omega \in I, \theta \in J$, $F(x,\omega,\theta)$ has Lipschitz gradient with constant $L = 1+\beta\lambda(A^TA)$, where $\lambda(A^TA)$ is the largest eigenvalue of A^TA.*

(ii) The function $\mathcal{F}(x)$ also has Lipschitz gradient with constant L.

Proof. (i) It is easy to see that

$$\nabla_x F(x,\omega,\theta) = x - P_{C_\omega}(x) + \beta A^T(Ax - P_{Q_\theta}(Ax)).$$

For fixed $\omega \in \mathcal{E}, \theta \in \mathcal{F}$,

$$\begin{aligned} & \|\nabla_x F(x,\omega,\theta) - \nabla_x F(y,\omega,\theta)\| \\ \le\ & \|x - P_{C_\omega}(x) - y + P_{C_\omega}(y)\| \\ & +\beta\left\|A^T\left[Ax - P_{Q_\theta}(Ax) - Ay + P_{Q_\theta}(Ay)\right]\right\| \\ \le\ & (1+\beta\lambda(A^TA))\|x-y\|. \end{aligned}$$

The last inequality follows from the firmly non-expansive property of $I - P_{C_\omega}$ and $I - P_{Q_\theta}$.

(ii) Since $\nabla\mathcal{F}(x) = \mathbf{E}\left[\nabla_x F(x,\omega,\theta)\right]$,

$$\begin{aligned} & \|\nabla\mathcal{F}(x) - \nabla\mathcal{F}(y)\| \\ =\ & \|\mathbf{E}\left[\nabla_x F(x,\omega,\theta)\right] - \mathbf{E}\left[\nabla_x F(y,\omega,\theta)\right]\| \\ \le\ & \mathbf{E}\left[\|\nabla_x F(x,\omega,\theta) - \nabla_x F(y,\omega,\theta)\|\right] \\ \le\ & L\|x-y\|. \end{aligned}$$

□

Lemma 2.5.

$$\boldsymbol{E}\left[\|\nabla_x F(x,\omega,\theta)\|^2\right] \le 2L\mathcal{F}(x). \tag{6}$$

Proof. We have

$$\begin{aligned} & \|\nabla_x F(x,\omega,\theta)\|^2 \\ = & \|x - P_{C_\omega}(x) + \beta A^T(Ax - P_{Q_\theta}(Ax))\|^2 \\ = & \|x - P_{C_\omega}(x)\|^2 + \beta^2\|A^T(Ax - P_{Q_\theta}(Ax))\|^2 \\ & +2\beta\langle x - P_{C_\omega}(x), A^T(Ax - P_{Q_\theta}(Ax))\rangle. \end{aligned}$$

By Cauchy-Schwarz inequality,

$$\begin{aligned} & \langle x - P_{C_\omega}(x), A^T(Ax - P_{Q_\theta}(Ax))\rangle \\ \le & \lambda(A^TA)\|x - P_{C_\omega}(x)\|^2 + \frac{1}{\lambda(A^TA)}\|A^T(Ax - P_{Q_\theta}(Ax))\|^2. \end{aligned}$$

Therefore,

$$\|\nabla_x F(x,\omega,\theta)\|^2 \le 2LF(x,\omega,\theta).$$

By taking expectation with respect to ω and θ, we obtain

$$\mathbf{E}\left[\|\nabla_x F(x,\omega,\theta)\|^2\right] \le 2L\mathcal{F}(x).$$

□

Lemma 2.6. *(Supermartingale convergence lemma[10, 13])*
Let $\{v_k\}, \{u_k\}, \{a_k\}$ and $\{b_k\}$ be sequences of nonnegative random variables such that

$$\boldsymbol{E}[v_{k+1}|\mathcal{V}_k] \le (1+a_k)v_k - u_k + b_k \quad a.s. \text{ for all } k \ge 0,$$

$$\sum_{k=0}^{\infty} a_k < \infty \quad a.s., \quad \sum_{k=0}^{\infty} b_k < \infty \quad a.s.,$$

where $\mathcal{V}_k$ denotes the σ-algebraic generated by random variables $v_0, \dots, v_k$, $u_0, \dots, u_k$, $a_0, \dots, a_k$, $b_0, \dots, b_k$. Then, we have $\lim_{k\to\infty} v_k = v$ for a random variable $v \ge 0 \quad a.s.$, and $\sum_{k=0}^{\infty} u_k < \infty \quad a.s.$

3 Algorithm and it convergence analysis

Algorithm:

Take a mini-batch size $N \ge 1$, and a positive sequence $\{\alpha_k\}_{k\ge1}$

Iter 0**:** Let x_0 be arbitrary.

Iter k: Draw $2N$ independent samples $\omega_k^1, \omega_k^2, \ldots, \omega_k^N \sim \mathcal{P}$, $\theta_k^1, \theta_k^2, \ldots, \theta_k^N \sim \mathcal{Q}$.
Compute

$$x_k \quad = \quad x^{k-1} - \frac{\alpha_k}{N} \sum_{i=1}^{N} \nabla_x F(x^{k-1}, \omega_k^i, \theta_k^i).$$

We denote by $\mathcal{X}_k$ the history of the method up to time $k \geq 1$

$$\mathcal{X}_k = \left\{ x_0, (\omega_t^i, 1 \leq i \leq N, 1 \leq t \leq k), (\theta_t^i, 1 \leq i \leq N, 1 \leq t \leq k) \right\}.$$

Proposition 3.1. *If the sequence $\{\alpha_k\}$ satisfy the following condition*

$$\sum_{k=1}^{\infty} \alpha_k^2 < \infty, \sum_{k=1}^{\infty} \alpha_k = \infty,$$

then there exists a nonnegative random variable c such that

$$\lim_{k \to \infty} \mathcal{F}(x^k) = c \quad a.s., \tag{7}$$

and

$$\liminf_{k \to \infty} \nabla \mathcal{F}(x^k) = 0 \quad a.s. \tag{8}$$

Proof. As proved in Lemma 2.4, the function $\mathcal{F}(x)$ has Lipschitz gradient with constant $L = 1 + \lambda(A^T A)$. Therefore,

$$\begin{aligned} \mathcal{F}(x^k) &\leq \mathcal{F}(x^{k-1}) + \nabla \mathcal{F}(x^{k-1})^T (x^k - x^{k-1}) + \frac{L}{2} \|x^k - x^{k-1}\|^2 \\ &= \mathcal{F}(\S^{\|-\infty}) - \frac{\alpha_k}{N} \sum_{i=1}^{N} \nabla \mathcal{F}(x^{k-1})^T \nabla_x F(x^{k-1}, \omega_k^i, \theta_k^i) \\ &\quad + \frac{L\alpha_k^2}{2N^2} \| \sum_{i=1}^{N} \nabla_x F(x^{k-1}, \omega_k^i, \theta_k^i) \|^2. \end{aligned}$$

Take the expectation on $\mathcal{X}_{k-1}$ and note that ω_k^i and θ_k^i are independent of the past $\mathcal{X}_{k-1}$ when x^{k-1} is determined by $\mathcal{X}_{k-1}$, we have

$$\begin{aligned} \mathbf{E}\left[\mathcal{F}(x^k) | \mathcal{X}_{k-1}\right] &\leq F(x^{k-1}) - \frac{\alpha_k}{N} \sum_{i=1}^{N} \nabla \mathcal{F}(x^{k-1})^T \mathbf{E}\left[\nabla_x F(x^{k-1}, \omega_k^i, \theta_k^i)\right] \\ &\quad + \frac{L\alpha_k^2}{2N^2} \mathbf{E}\left[\| \sum_{i=1}^{N} \nabla_x F(x^{k-1}, \omega_k^i, \theta_k^i) \|^2 \right] \quad a.s. \end{aligned} \tag{9}$$

By definitions of F and $\mathcal{F}$, it is easy to see that

$$\mathbf{E}\left[\nabla_x F(x^{k-1},\omega_k^i,\theta_k^i)\right] = \nabla\mathcal{F}(x^{k-1}). \tag{10}$$

In addition, thanks to Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathbf{E}\left[\|\sum_{i=1}^{N}\nabla_x F(x^{k-1},\omega_k^i,\theta_k^i)\|^2\right] &\leq \mathbf{E}\left[N\sum_{i=1}^{N}\|\nabla_x F(x^{k-1},\omega_k^i,\theta_k^i)\|^2\right] \\ &= N\sum_{i=1}^{N}\mathbf{E}\left[\|\nabla_x F(x^{k-1},\omega_k^i,\theta_k^i\|^2\right].\end{aligned}$$

By using Lemma 2.5, we obtain

$$\mathbf{E}\left[\|\sum_{i=1}^{N}\nabla_x F(x^{k-1},\omega_k^i,\theta_k^i)\|^2\right] \leq 2N^2L\mathcal{F}(x^{k-1}). \tag{11}$$

Combinning (9), (10), (11), we have

$$\mathbf{E}\left[\mathcal{F}(x)|\mathcal{X}_{k-1}\right] \leq (1+L^2\alpha_k^2)\mathcal{F}(x^{k-1}) - \alpha_k\|\nabla\mathcal{F}(x^{k-1})\|^2 \quad a.s. \tag{12}$$

Now, thanks to the supermartingale convergence Lemma 2.6, we can conclude that

$$\mathcal{F}(x^k) \to c \quad a.s.,$$

for some nonegative random variable c and

$$\sum_{k=1}^{\infty}\alpha_k\|\nabla\mathcal{F}(x^{k-1})\|^2 < \infty \quad a.s.$$

But by assumption, $\sum_{k=1}^{\infty}\alpha_k = \infty$. It implies that

$$\liminf_{k\to\infty}\|\nabla\mathcal{F}(x^{k-1})\|^2 = 0 \quad a.s.$$

Hence,

$$\liminf_{k\to\infty}\nabla\mathcal{F}(x^k) = 0 \quad a.s.$$

□

Theorem 3.2. *Assume that the solution set S_1 of (SOP) is nonempty. Then if $\sum_{k=1}^{\infty}\alpha_k^2 < \infty$ and $\sum_{i=1}^{\infty}\alpha_k = \infty$ the sequence $\{x^k\}$ generated by the Algorithm converges almost surely to a random point in the solution set S_1.*

Proof. Let z belong to S_1 and $\mathcal{F}^* = \mathcal{F}(z)$ be the optimal value of (SOP). We have

$$\begin{aligned} \|x_{k+1} - z\|^2 &= \|x^{k-1} - z\|^2 + 2\langle x^{k-1} - z, x^k - x^{k-1}\rangle + \|x^k - x^{k-1}\|^2. \\ &= \|x^{k-1} - z\|^2 - 2\frac{\alpha_k}{N}\sum_{i=1}^{N}\langle x^{k-1} - z, \nabla_x F(x^{k-1}, \omega_k^i, \theta_k^i)\rangle \\ &\quad + \frac{\alpha_k^2}{N^2}\|\sum_{i=1}^{N}\nabla_x F(x^{k-1}, \omega_k^i, \theta_k^i)\|^2. \end{aligned}$$

Taking the conditional expectation on $\mathcal{X}_{k-1}$ and using

$$\mathbf{E}\left[\nabla_x F(x^{k-1}, \omega_k^i, \theta_k^i)\right] = \nabla \mathcal{F}(x^{k-1}),$$

$$\mathbf{E}\left[\|\sum_{i=1}^{N}\nabla_x F(x^{k-1}, \omega_k^i, \theta_k^i)\|^2\right] \leq 2N^2 L\mathcal{F}(x^{k-1}),$$

we obtain

$$\begin{aligned} &\mathbf{E}\left[\|x^k - z\|^2 | \mathcal{X}_{k-1}\right] \\ \leq\ & \|x^{k-1} - z\|^2 - 2\alpha_k \nabla\mathcal{F}(x^{k-1})^T(x^{k-1} - z) + 2\alpha_k^2 L\mathcal{F}(x^{k-1}). \end{aligned}$$

Since $\mathcal{F}$ is convex, we have

$$\nabla\mathcal{F}(x^{k-1})^T(x^{k-1} - z) \geq \mathcal{F}(x^{k-1}) - \mathcal{F}(z) = \mathcal{F}(x^{k-1}) - \mathcal{F}^*.$$

So,

$$\begin{aligned} \mathbf{E}\left[\|x^k - z\|^2 \| \mathcal{X}_{k-1}\right] &\leq \|x^{k-1} - z\|^2 - 2\alpha_k(\mathcal{F}(x^{k-1}) - \mathcal{F}^*) \\ &\quad + 2\alpha_k^2 L\mathcal{F}(x^{k-1}). \end{aligned} \tag{13}$$

By Proposition 3.1, the sequence $\mathcal{F}(x^{k-1})$ converge almost surely, hence it is bounded almost surely. Combining this with the condition $\sum_{k=1}^{\infty}\alpha_k^2 < \infty$, we imply that

$$\sum_{k=1}^{\infty}\alpha_k^2\mathcal{F}(x^{k-1}) < \infty \quad a.s.$$

Clearly, $\mathcal{F}(x^{k-1}) \geq \mathcal{F}^* \geq 0$. Thanks to the supermartigale convergence lemma, we have the sequence $\left\{\|x^k - z\|\right\}$ is convergent almost surely for z arbitrary in S_1. Moreover,

$$\sum_{k=1}^{\infty}\alpha_k(\mathcal{F}(x^{k-1} - \mathcal{F}^*) < \infty \quad a.s.$$

Therefore, $\liminf_{k\to\infty}(\mathcal{F}(x^{k-1} - \mathcal{F}^*) = 0 \quad a.s.$ or

$$\liminf_{k\to\infty} \mathcal{F}(x^{k-1}) = \mathcal{F}^* \quad a.s. \tag{14}$$

On the other hand, the sequence $\{\|x^k - z\|\}$ is convergent almost surely. Therefore, almost surely, $\{x^k\}$ is bounded and has limit points. By using (14) and the continuity of $\mathcal{F}$, we can conclude that $\{x^k\}$ converges and its limit point is in S_1 almost surely. □

Theorem 3.3. *Assume that the solution set S of (MSFP) is nonempty. Then if there exist positive numbers $\underline{\alpha}$ and overlineα such that*

$$0 < \underline{\alpha} \leq \alpha_k \leq \overline{\alpha} < \frac{1}{L},$$

then the sequence $\{x^k\}$ generated by the Algorithm converges almost surely to a random point in the solution set S.

Proof. Let z be a solution of (SSF) then $z \in S_1$ and the optimal value of (SOP) is $\mathcal{F}^* = \mathcal{F}(z) = 0$. So (13) becomes

$$\mathbf{E}\left[\|x^k - z\|^2 | \mathcal{X}_{k-1}\right] \leq \|x^{k-1} - z\|^2 - 2\alpha_k(1 - \alpha_k L)\mathcal{F}(x^{k-1}). \tag{15}$$

Note that $\mathcal{F}(x^{k-1}) \geq 0$ and $\alpha_k(1 - \alpha_k L) > 0$. By using the supermartingle convergence lemma, we obtain that the sequence $\{\|x^k - z\|\}$ is convergent almost surely for any $z \in S$. In addition,

$$\sum_{k=1}^{\infty} \alpha_k(1 - \alpha_k L)\mathcal{F}(x^{k-1}) < \infty \quad a.s.$$

Since $\alpha_k(1 - \alpha_k L) \geq \underline{\alpha}(1 - \overline{\alpha} L) > 0$, it implies that

$$\liminf_{k\to\infty} \mathcal{F}(x^{k-1}) = 0.$$

By the same argument as in the proof of Theorem 3.2, we can conclude that $\{x^k\}$ converges and its limit point is in S almost surely. □

The following proposition provides the convergence rate of our algorithm.

Proposition 3.4. *Assume that S is nonempty and $0 < \underline{\alpha} \leq \alpha_k \leq \overline{\alpha} < \frac{1}{L}$ for every k.*

(i) Let $\hat{x}^k$ be the average point, i.e.

$$\hat{x}^k = \frac{1}{\sum_{i=0}^k \alpha_{i+1}} \sum_{i=0}^k \alpha_{i+1} x^i.$$

Then, we have

$$\boldsymbol{E}[\mathcal{F}(\hat{x}^k)] \leq \frac{dist_S^2(x_0)}{2(1-\overline{\alpha}L)\sum_{i=0}^k \alpha_{i+1}}.$$

(ii) If the linear regularity condition (1) holds, then we have

$$\boldsymbol{E}[dist_S(x^k)] \leq \left(1 - \frac{2\underline{\alpha}(1-\overline{\alpha}L)}{\kappa}\right) \boldsymbol{E}[dist_S(x^{k-1})],$$

and

$$\boldsymbol{E}[\mathcal{F}(x^k)] \leq \left(1 - \frac{2\underline{\alpha}(1-\overline{\alpha}L)}{\kappa}\right)^k \frac{dist_S(x^0)}{2}.$$

Proof. (i) Taking expectation on (15), we obtain, for any $z \in S$,

$$\mathbf{E}[\|x^k - z\|^2] \leq \mathbf{E}[\|x^{k-1} - z\|^2] - 2\alpha_k(1-\alpha_k L)\mathbf{E}[\mathcal{F}(x^{k-1})]. \quad (16)$$

For any α_k satisfied $0 < \underline{\alpha} \leq \alpha_k \leq \overline{\alpha} < \frac{1}{L}$, we have

$$\alpha_k(1-\alpha_k L) \geq \alpha_k(1-\overline{\alpha}L).$$

It implies that

$$2\alpha_k(1-\overline{\alpha}L)\mathbf{E}[\mathcal{F}(x^{k-1})] \leq \mathbf{E}[\|x^{k-1} - z\|^2] - \mathbf{E}[\|x^k - z\|^2]. \quad (17)$$

By taking the sum of (17) from 1 to $k+1$, we have

$$2(1-\overline{\alpha}L)\mathbf{E}[\sum_{i=0}^k \alpha_{i+1}\mathcal{F}(x^i)] \leq \|x_0 - z\|^2.$$

Thanks to the convexity of $\mathcal{F}$ and by taking $z = P_S(x_0)$, we can conclude that

$$\mathbf{E}[\mathcal{F}(\hat{x}^k)] \leq \frac{dist_S^2(x_0)}{2(1-\overline{\alpha}L)\sum_{i=0}^k \alpha_{i+1}}.$$

(ii) If the linear regularity condition holds then there exists $\kappa < \infty$ such that

$$dist_S^2(x) \leq \kappa \mathcal{F}(x) \quad \forall x \in \mathbb{R}^n.$$

From (15), we have

$$\mathbf{E}\left[\|x^k - z\|^2 | \mathcal{X}_{k-1}\right] \leq \|x^{k-1} - z\|^2 - \frac{2\alpha_k(1-\alpha_k L)}{\kappa} dist_S^2(x^{k-1}).$$

Taking expectation, we get

$$\mathbf{E}[\|x^k - z\|^2] \leq \mathbf{E}[\|x^{k-1} - z\|^2] - \frac{2\alpha_k(1-\alpha_k L)}{\kappa}\mathbf{E}[dist_S^2(x^{k-1})]. \quad (18)$$

We can choose $z = P_S(x^{k-1})$ and note that

$$\|x^k - P_S(x^{k-1})\|^2 \geq dist_S^2(x^k).$$

From (18), it implies that

$$\begin{aligned} \mathbf{E}[dist_S^2(x^k)] &\leq \left(1 - \frac{2\alpha_k(1-\alpha_k L)}{\kappa}\right)\mathbf{E}[dist_S^2(x^{k-1})] \\ &\leq \left(1 - \frac{2\underline{\alpha}(1-\overline{\alpha} L)}{\kappa}\right)\mathbf{E}[dist_S^2(x^{k-1})]. \end{aligned}$$

□

4 Numerical experiments

In this section, we report several numerical experimental results to illustrate the behavior of our algorithm. We implement the algorithm in Matlab on a Corei5 computer with 512Mb RAM.

Example 4.1. *In this example, we suppose A is an $m \times n$ matrix and*

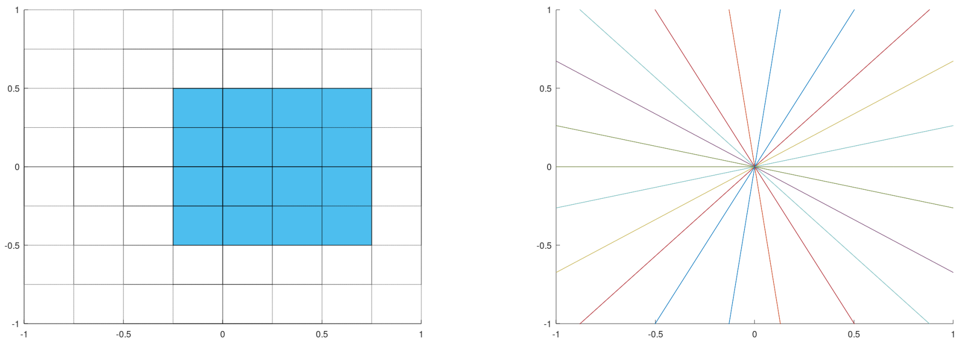
- *The random vectors $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ and $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ are the uniformly random vectors in $[0,1]^n$ and $[0,1]^m$, respectively;*
- *C_ω is a box in $\mathbb{R}^n$ defined by*

$$C_\omega = \{x \in \mathbb{R}^n | \quad -1 + \omega_i \leq x_i \leq \omega_i \quad \forall i\}.$$

- *Q_θ is a half-space in $\mathbb{R}^m$ defined by*

$$Q_\theta = \left\{y \in \mathbb{R}^m | \quad c_\theta^T y \leq 0\right\},$$

with $c_\theta = (-1, -1, \ldots, -1) + 2\theta$.



(a) The boxes C_ω (b) The half-spaces Q_θ

Figure 1: Boxes and half-spaces

Figure 1 illustrates the sets C_ω *and* Q_θ*. It is clear that*

$$\cap_\omega C_\omega = \{0\},$$

and

$$\cap_\theta Q_\theta = \{0\},$$

hence the (SSF) problem has unique solution that is the origin $(0, 0, \ldots, 0)$ *of* $\mathbb{R}^n$*. To test our algorithm, we take*

$$\beta = 1; \quad \alpha_k \equiv \frac{1}{1.5(\beta + \lambda(A^T A))} \quad \forall k,$$

and each entry of the matrix A *is uniformly generated in* $[0, 1]$*. For each size* (m, n) *of problem, we test the algorithm on* 100 *samples of* A *and report the average time and error* $\|x^k\|$ *corresponding to different values of mini-batch size* N *in Table 1 and 2. We stop the algorithm if* $\|x^k\| \leq 10^{-2}$ *or the number of iterations exceeds* 500*.*

Example 4.2. *Suppose that* A *is a* 2×2 *matrix and*

- ω *is uniformly distribution on the unit circle* $C(0, 1) \subset \mathbb{R}^2$ *and* C_ω *is the disc with center* ω *and radius* 2 *in* $\mathbb{R}^2$*.*
- θ *is uniformly distribution on the circle* $C(0, 2) \subset \mathbb{R}^2$ *and* Q_θ *is the disc with center* θ *and radius* 4 *in* $\mathbb{R}^2$*.*

Table 1: Average CPUs time corresponding to different problem and mini-batch sizes

(n, m)	$N = 1$	$N = 5$	$N = 10$
(5,5)	3.2953	15.631	31.286
(10,5)	3.8206	18.314	36.382
(20,5)	4.1570	20.735	41.819
(50,5)	4.5560	22.590	45.254
(100.5)	6.8337	33.558	67.650

Table 2: Average error corresponding to different problem and mini-batch sizes

(n, m)	$N = 1$	$N = 5$	$N = 10$
(5,5)	0.073978	0.073496	0.072076
(10,5)	0.23751	0.23911	0.23908
(20,5)	0.59016	0.58459	0.58973
(50,5)	1.5513	1.5517	1.5513
(100,5)	3.0151	3.0033	3.0245

We can easily check that

$$\cap_\omega C_\omega = \{x \in \mathbb{R}^2 | \|x\| \leq 1\} = B(0, 1),$$

and

$$\cap_\theta Q_\theta = \{y \in \mathbb{R}^2 | \|y\| \leq 2\} = B(0, 2).$$

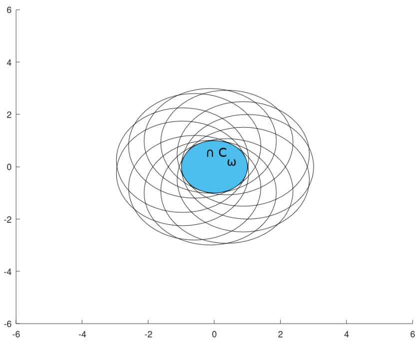
Figure 2 helps us understand the discs C_ω and Q_θ. The (SSF) problem has infinitely many solutions and x^* is a solution of (SSF) if and only if $\|x\| \leq 1$ and $\|Ax\| \leq 2$. Therefore, we set

$$er := \max(0, \|x\| - 1) + \max(0, \|Ax\| - 2),$$

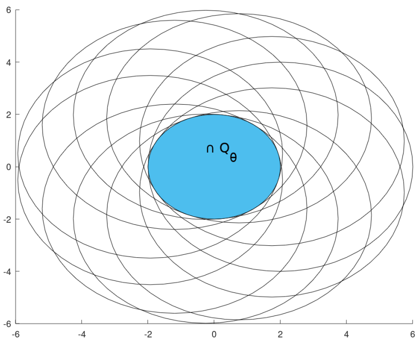
as the error for this example.

We take $\beta = 1$ and test the algorithm with different choices of mini-batch sizes $N = 1, 5, 10$. The stopping rules are $er < 10^{-4}$ and the number of iterations exceeds 500. The average errors over 100 problems corresponding to 100 randomly generated matrix A are plotted in Figure 3.

We also test the algorithm with different choices of α_k. The average times and errors are reported in Table 3. We observe that the choice of the parameter α_k plays an important role for the efficiency of the algorithm. The error goes to 0 quite quick when α_k close to $\frac{1}{L}$.



(a) The discs C_ω



(b) The discs Q_θ

Figure 2: Circles

Table 3: Average CPUs times and errors corresponding to different choices of α_k

	$\alpha_k = \frac{1}{1.1L}$	$\alpha_k = \frac{1}{1.5L}$	$\alpha_k = \frac{1}{k+1}$
Average CPUs time	0.33749	0.35064	2.8761
Average error	0.00031	0.00050	0.01419

5 Conclusion

We proposed a stochastic reformulation of the stochastic split feasibility problem and studied the equivalence between these problems. Then, we introduced a mini-batch random projection algorithm and proved the convergence in consistent and non-consistent cases. We also derived the linear convergence rates for this algorithm under linear regularity condition.

Acknowledgements

This work has been done when the author visited the Vietnam Institute for Advanced Study in Mathematics (VIASM). The author would like to thank VIASM for hospitality and financial support.

References

[1] H.H. Bauschke and J.M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Review **38(3)**, 367-426 (1996).

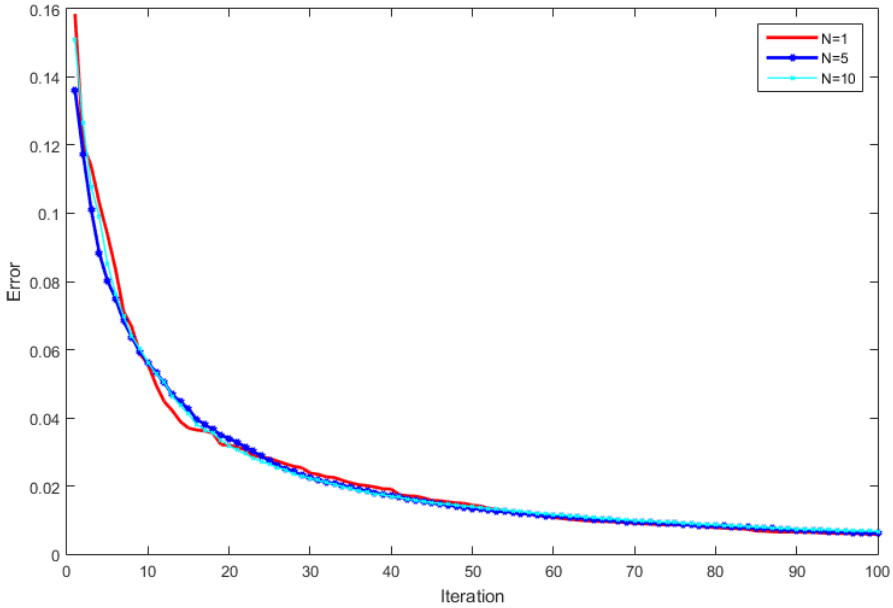


Figure 3: Convergence behavior corresponding to different choice of mini-batch size.

[2] D. Butnariu and S.D. Flam, *Strong convergence of expected-projection methods in Hilbert spaces*, Numerical Functional Analysis and Optimization **16(5)**, 601–636 (1995).

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

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

[5] Y. Censor and T. Elfving, *A multiprojections algorithm using Bregman projections in a product spaces*, Numerical Algorithms **8**, 221-239 (1994).

[6] Y. Censor, T. Elfving, N. Kopf and T. Bortfeld, *The multiple-sets split feasibility problem and its applications for inverse problems*, Inverse Problems **21**, 2071-2084 (2005).

[7] P.L. Combettes, *The convex feasibility problem in image recovery*, Advances in Imaging and Electron Physics **95** 155–270 (1996).

[8] N. Hermer. D. R. Luke and A.Sturm, *Random Function Iterations for Consistent Stochastic Feasibility*, https://arxiv.org/abs/1808.05426.

[9] A. Moudafi and B.S. Thakur, *Solving proximal split feasibility problems without prior knowledge of operator norms*. Optim. Lett. **8**, 2099-2110 (2014).

[10] A. Nedíc, *Random projection algorithms for convex set intersection problems*, 49th IEEE Conference on Decision and Control, 7655-7660 (2010).

[11] A. Nedíc and I. Necoara, *Random minibatch projection algorithms for convex problems with functional constraints*, https://arxiv.org/abs/1903.02117.

[12] I. Necoara, A. Patrascu and P. Richtarik, *Randomized projection methods for convex feasibility problems: conditioning and convergence rates*, https://arxiv.org/abs/1801.04873.

[13] H. Robbins and D. Siegmund, *A convergence theorem for non negative almost supermartingales and some applications*, Optimizing Methods in Statistics, 1971.

[14] M. Wamg and D.P. Bertsekas, *Stochastic first-order methods with random constraint projection*, SIAM Journal on Optimization **26(1)**, 681-717 (2016).