Random Projection Algorithms for Convex Set Intersection Problems

A. Nedić

Department of Industrial and Enterprise Systems Engineering
University of Illinois, Urbana, IL 61801
angelia@illinois.edu

Abstract— The focus of this paper is on the set intersection problem for closed convex sets admitting projection operation in a closed form. The objective is to investigate algorithms that would converge (in some sense) if and only if the problem has a solution. To do so, we view the set intersection problem as a stochastic optimization problem of minimizing the "average" residual error of the set collection. We consider a stochastic gradient method as a main tool for investigating the properties of the stochastic optimization problem. We show that the stochastic optimization problem has a solution if and only if the stochastic gradient method is convergent almost surely. We then consider a special case of the method, namely the random projection method, and we analyze its convergence. We show that a solution of the intersection problem exists if and only if the random projection method exhibits certain convergence behavior almost surely. In addition, we provide convergence rate results for the expected residual error.

Index Terms— Convex sets, intersection problem, random projection.

I. Introduction

In this paper, we deal with the set intersection problem of finding a common point to a (possibly infinite) collection of closed convex sets, where the projection on each of the sets can be done efficiently. Such a problem is also referred to as a feasibility problem. It arises in many applications including image restoration and tomography amongst others [13], [5], [12]. Our objective is to solve the problem by using an iterative algorithm that will, at the same time, enable us to determine whether the problem has a solution or not.

With this objective on mind, we at first consider a stochastic optimization problem of minimizing an "average residual error" associated with the set collection. To solve the problem, we employ the standard stochastic gradient method and study its convergence for diminishing stepsize. We discover that, unlike the standard stochastic gradient schemes, the convergence of the method *does not require any compactness assumptions* on the sets under consideration. This is possible due to some special properties of the problem. We establish that the method converges almost surely if and only if the problem of minimizing the average residual error has an optimal solution. We then focus on a random projection method, which is just a special case of the stochastic gradient method with a more aggressive stepsize (a constant stepsize). We establish that the feasibility problem has a solution if and only if the random projection method converges and

the residual error tends to zero almost surely. We also provide a convergence rate result showing that the expected residual error (along the averages of the iterates) diminishes to zero proportionally to $1/t$ in the number t of iterations. Furthermore, we show that the expected distances from the iterates to the intersection set converge with a geometric rate when either the sets are given by linear inequalities or the interiors of the sets have a common point.

The main contribution of this paper lies in the establishment of one-to-one relation between the almost sure convergence of the discussed algorithms and the existence of solutions for the corresponding problems to which the algorithms are applied. To the best of our knowledge, our convergence and convergence rate results for the random projection method are also new.

The present paper is closely related to [24] (see also [10]), where random subgradient algorithms have been studied for convex feasibility problems. In [24], under the nonempty interior intersection condition, a finite-time convergence is established for possibly infinite collection of sets. In contrast, in this paper we explore the behavior of the method in general including the possibility that the intersection set may be empty. The algorithms of this paper can be viewed as stochastic counterparts of the deterministic subgradient projection algorithms in the optimization literature for convex feasibility problems, starting with the early work in [1], [22], [8], and a more recent work in [18], [16], [21].

The work in this paper can also be viewed as a random implementation of the alternating projection method, which generates a sequence of iterates by projecting on the sets cyclically. The algorithm has been proposed by Von Neumann [26] for the intersection problem of two subspaces in a Hilbert space, and it has many generalization and extensions [3], [20], [19], [14], [15]. A nice survey of the work in this area is given in [4]. The most notable early work is that of Gubin *et al.* [19], who have established the first convergence rate result assuming that the intersection set has a nonempty interior. Unlike the alternate projection method (which is deterministic), the algorithms of this paper utilize random projections. On a broader scale, the work in this paper is related to convex feasibility problems resulting from random sampling [2], [11].

The paper is organized as follows. In Section II, we introduce the feasibility problem of our interest and we cast it as a stochastic optimization problem. We discuss the stochastic gradient method for solving the problem. In

Section III, we present some convergence results for the method, and we provide necessary and sufficient conditions for the existence of solutions for the optimization problem. These conditions are reflective of the convergence behavior of the method. In Section IV, we focus on a special case of the stochastic gradient method corresponding to the random projection method, and we study the convergence properties of the method. Further, we establish necessary and sufficient conditions for the existence of a common point for the original feasibility problem. The conditions are based on the behavior exhibited by the random projection method. In addition, we provide convergence rate estimates for the method. We conclude in Section V with some comments.

Notation. A vector is a column vector. We use x^T to denote the transpose of a vector x, and $\|x\|$ to denote the standard Euclidean norm. We write $\mathrm{dist}(\bar{x}, X)$ for the distance of a vector $\bar{x}$ from a closed convex set X, i.e., $\mathrm{dist}(\bar{x}, X) = \min_{x \in X} \|x - \bar{x}\|$. We use $\Pi_X[\bar{x}]$ to denote the projection of a vector $\bar{x}$ on the set X, i.e., $\Pi_X[\bar{x}] = \mathrm{argmin}_{x \in X} \|x - \bar{x}\|$. We write $\mathrm{Pr}\{\cdot\}$ to denote the probability of a random variable. We often abbreviate *almost surely* by *a.s.*

II. PROBLEM, ALGORITHM, ASSUMPTIONS

We are interested in the feasibility problem for a collection $\{X_i, i \in \mathscr{I}\}$ of convex closed sets in $\mathbb{R}^n$. We would like to use an iterative algorithm to find a common point for the sets when such a point exists. Also, we would like to determine if a solution does not exist based on the behavior of the algorithm. To do so, with the given collection $\{X_i, i \in \mathscr{I}\}$, we associate the following stochastic optimization problem:

$$\begin{aligned} \text{minimize} \quad & \mathscr{R}(x) = \frac{1}{2}\, \mathsf{E}\big[\|x - \Pi_{X_\omega}[x]\|^2\big] \\ \text{subject to} \quad & x \in \mathbb{R}^n, \end{aligned} \quad (1)$$

where ω is a random variable taking values in the set $\mathscr{I}$, and the expectation is taken with respect to ω. The function $\mathscr{R}(x)$ is the average residual of x with respect to the sets X_i. We assume that the projection on each set X_i is easy, i.e., it is available in a closed form Examples of such sets include hyperplanes, balls and the sets given by linear inequalities.

We let X^* and $\mathscr{R}^*$ denote the optimal set and the optimal value of problem (1), respectively. Note that $\mathscr{R}^* \geq 0$. We also let $X = \cap_{i \in \mathscr{I}} X_i$. When the intersection set X is nonempty, we have $X^* = X$. When the set X is empty, the optimal set X^* is either empty ($\mathscr{R}^*$ not attained) or it coincides with the set of stationary points of $\mathscr{R}(x)$, i.e., $X^* = \{x^* \mid \nabla \mathscr{R}(x^*) = 0\}$.

For problem (1), we consider a stochastic gradient algorithm, as follows. At time k, we have an iterate x_{k-1} and a realization of a random variable ω_k with values in the index set $\mathscr{I}$. The new iterate x_k is determined by

$$x_k = x_{k-1} - \alpha_k (x_{k-1} - \Pi_{\omega_k}[x_{k-1}]) \quad \text{for } k \geq 1, \quad (2)$$

where $\alpha_k > 0$ is a deterministic stepsize, while Π_{ω_k} denotes the projection on the set X_{ω_k}. The initial point $x_0 \in \mathbb{R}^n$ is selected randomly with an arbitrary distribution. The random

variable ω_k is a random sample of ω at time k, drawn from the same distribution as that of ω.

To see that algorithm (2) is the stochastic gradient approximation method for problem (1), note that the function $\mathscr{R}(x)$ can be written as:

$$\mathscr{R}(x) = \mathsf{E}[R(x,\omega)]$$

$$R(x,\omega) = \frac{1}{2}\|x - \Pi_{X_\omega}[x]\|^2 \qquad \text{for } x \in \mathbb{R}^n,\ \omega \in \mathscr{I}.$$

For each fixed $\omega \in \mathscr{I}$, the function $R(x,\omega)$ is convex and differentiable[1] in x over $\mathbb{R}^n$, and so is $\mathscr{R}(x)$. Furthermore,

$$\nabla \mathscr{R}(x) = \mathsf{E}[\nabla_x R(x,\omega)] = x - \mathsf{E}[\Pi_{X_\omega}[x]] \qquad \text{for } x \in \mathbb{R}^n.$$

In algorithm (2), the term $x_{k-1} - \Pi_{\omega_k}[x_{k-1}]$ is the gradient of the function $R(x,\omega)$ with respect to x, which is evaluated at $x = x_{k-1}$ and $\omega = \omega_k$, i.e.,

$$x_{k-1} - \Pi_{\omega_k}[x_{k-1}] = \nabla_x R(x_{k-1}, \omega_k).$$

Thus, method (2) can be written as

$$x_k = x_{k-1} - \alpha_k \nabla_x R(x_{k-1}, \omega_k) \quad \text{for } k \geq 1, \quad (3)$$

which is the stochastic gradient approximation method for problem (1).

Observe that for each $\omega \in \mathscr{I}$, the function $R(x,\omega)$ has Lipschitz gradients with constant $L = 1$. Thus, the function $\mathscr{R}(x)$ also has Lipschitz gradients with constant $L = 1$, i.e.,

$$\|\nabla \mathscr{R}(x) - \nabla \mathscr{R}(y)\| \leq \|x - y\| \qquad \text{for all } x, y \in \mathbb{R}^n.$$

Therefore, $\mathscr{R}$ satisfies the following relation[2]: for $x, y \in \mathbb{R}^n$,

$$\mathscr{R}(x) \leq \mathscr{R}(y) + \nabla \mathscr{R}(y)^T (x - y) + \frac{1}{2}\|x - y\|^2. \quad (4)$$

Furthermore, directly from the definitions of $\mathscr{R}(x)$ and $R(x,\omega)$, we can see that

$$\mathsf{E}\big[\|\nabla_x R(x,\omega)\|^2\big] = 2\mathscr{R}(x) \qquad \text{for any } x \in \mathbb{R}^n. \quad (5)$$

For now we do not make any assumptions on the cardinality of the set $\mathscr{I}$. We assume the following for the sets X_i.

Assumption 1: The set $X_i \subseteq \mathbb{R}^n$ is closed and convex for each $i \in \mathscr{I}$.

For the random sequence $\{\omega_k\}$, we assume the following.

Assumption 2: The sequence $\{\omega_k\}$ is i.i.d. and independent of the initial random point x_0.

In the subsequent development, we make use of the strictly nonexpansive property of the projection operation on a closed convex set $Y \subset \mathbb{R}^n$, given as follows[3]: for any $x \in \mathbb{R}^n$ and $z \in Y$,

$$\|\Pi_Y[x] - z\|^2 \leq \|x - z\|^2 - \|x - \Pi_Y[x]\|^2. \quad (6)$$

We also make use of the following supermartingale convergence result due to Robbins and Siegmund (see [23], Lemma 11 on page 50).

---

[1]See [17], Volume I, 1.5.5 Theorem on page 77.
[2]See [23], relation (15) on page 6.
[3]See for example [17], Volume II, 12.1.13 Lemma on page 1120.

Theorem 1: Let $\{v_k\}$, $\{u_k\}$, $\{a_k\}$ and $\{b_k\}$ be sequences of nonnegative random variables such that

$$\mathsf{E}[v_{k+1} \mid \mathcal{F}_k] \leq (1 + a_k)v_k - u_k + b_k \quad \text{for all } k \geq 0 \ \ a.s.,$$

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

where $\mathcal{F}_k$ denotes the collection $v_0, \ldots, v_k, \ u_0, \ldots, u_k,$ $a_0, \ldots, a_k, \ b_0, \ldots, b_k$. Then, we have $\lim_{k\to\infty} v_k = v$ for a random variable $v \geq 0$ *a.s.*, and $\sum_{k=0}^{\infty} u_k < \infty$ *a.s.*

III. EXISTENCE OF OPTIMAL SOLUTIONS

In this section, we relate the existence of solutions to problem (1) with the behavior of the algorithm (2). We consider the diminishing stepsize α_k that satisfies the following.

Assumption 3: The stepsize α_k is such that $\alpha_k > 0$ for all k, with $\sum_{k=1}^{\infty} \alpha_k = \infty$ and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$.

We let $\mathcal{F}_k$ denote the history of the method up to time k,

$$\mathcal{F}_k = \{x_0, (\omega_t, \ 1 \leq t \leq k)\} \qquad \text{for } k \geq 1,$$

with $\mathcal{F}_0 = \{x_0\}$. We have the following result.

Proposition 1: Let Assumptions 1–3 hold. Then, for the iterates $\{x_k\}$ generated by method (2), we have almost surely

$$\lim_{k\to\infty} \mathcal{R}(x_k) = c \qquad \text{for a random scalar } c \geq \mathcal{R}^*,$$

$$\liminf_{k\to\infty} \nabla\mathcal{R}(x_k) = 0.$$

Proof: Using relation (4) with $x = x_k$ and $y = x_{k-1}$, we obtain surely for all $k \geq 1$,

$$\mathcal{R}(x_k) \leq \mathcal{R}(x_{k-1}) + \nabla\mathcal{R}(x_{k-1})^T (x_k - x_{k-1}) + \frac{1}{2}\|x_k - x_{k-1}\|^2.$$

From (3) it surely follows for all $k \geq 1$,

$$\mathcal{R}(x_k) \leq \mathcal{R}(x_{k-1}) - \alpha_k \nabla\mathcal{R}(x_{k-1})^T \nabla_x R(x_{k-1}, \omega_k) + \frac{\alpha_k^2}{2}\|\nabla_x R(x_{k-1}, \omega_k)\|^2.$$

We now take the expectation conditioned on $\mathcal{F}_{k-1}$ and note that ω_k is independent of the past $\mathcal{F}_{k-1}$, while x_{k-1} is fully determined by $\mathcal{F}_{k-1}$. Using $\mathsf{E}[\nabla_x R(x_{k-1}, \omega_k) \mid \mathcal{F}_{k-1}] = \nabla\mathcal{R}(x_{k-1})$, we have almost surely for all $k \geq 1$,

$$\mathsf{E}[\mathcal{R}(x_k) \mid \mathcal{F}_{k-1}] \leq \mathcal{R}(x_{k-1}) - \alpha_k\|\nabla\mathcal{R}(x_{k-1})\|^2 + \frac{\alpha_k^2}{2}\mathsf{E}[\|\nabla_x R(x_{k-1}, \omega_k)\|^2 \mid \mathcal{F}_{k-1}].$$

Since $\mathsf{E}[\|\nabla_x R(x_{k-1}, \omega_k)\|^2 \mid \mathcal{F}_{k-1}] = 2\mathcal{R}(x_{k-1})$ (see (5)), it follows almost surely for all $k \geq 1$,

$$\mathsf{E}[\mathcal{R}(x_k) \mid \mathcal{F}_{k-1}] \leq (1 + \alpha_k^2)\mathcal{R}(x_{k-1}) - \alpha_k\|\nabla\mathcal{R}(x_{k-1})\|^2.$$

Since $\mathcal{R}(x) \geq 0$ for all x, by the supermartingale convergence (Theorem 1), we see that $\mathcal{R}(x_k) \to c$ almost surely for some random scalar $c \geq \mathcal{R}^*$. We also have $\sum_{k=1}^{\infty} \alpha_k\|\nabla\mathcal{R}(x_k)\|^2 < \infty$ *a.s.*, which in view of the condition $\sum_{k=1}^{\infty} \alpha_k = \infty$, implies that $\liminf_{k\to\infty} \|\nabla\mathcal{R}(x_k)\| = 0$ almost surely. ∎

When the optimal set X^* of (1) is nonempty, the sequence $\{x_k\}$ is bounded and converges almost surely to a random solution. We show this in the following.

Proposition 2: Let Assumptions 1–3 hold, and assume that the optimal set X^* is nonempty. Then, the sequence $\{x_k\}$ generated by method (2) converges almost surely to a random point in the set X^*.

Proof: Let $z \in X^*$ be arbitrary. We then surely have for any $k \geq 1$,

$$\|x_k - z\|^2 = \|x_{k-1} - z\|^2 + \alpha_k^2\|\nabla_x R(x_{k-1}, \omega_k)\|^2 - 2\alpha_k \nabla_x R(x_{k-1}, \omega_k)^T (x_{k-1} - z).$$

By taking the conditional expectation on the past $\mathcal{F}_{k-1}$, and using $\mathsf{E}[\nabla_x R(x_{k-1}, \omega_k) \mid \mathcal{F}_{k-1}] = \nabla\mathcal{R}(x_{k-1})$ and $\mathsf{E}[\|\nabla_x R(x_{k-1}, \omega_k)\|^2 \mid \mathcal{F}_{k-1}] = 2\mathcal{R}(x_{k-1})$, we obtain almost surely for all $k \geq 1$,

$$\mathsf{E}[\|x_k - z\|^2 \mid \mathcal{F}_{k-1}] = \|x_{k-1} - z\|^2 + 2\alpha_k^2 \mathcal{R}(x_{k-1}) - 2\alpha_k \nabla\mathcal{R}(x_{k-1})^T (x_{k-1} - z).$$

Since $\mathcal{R}(x)$ is convex, we have

$$\nabla\mathcal{R}(x_{k-1})^T (x_{k-1} - z) \geq \mathcal{R}(x_{k-1}) - \mathcal{R}(z).$$

It follows that almost surely for all $k \geq 1$,

$$\mathsf{E}[\|x_k - z\|^2 \mid \mathcal{F}_{k-1}] \leq \|x_{k-1} - z\|^2 + 2\alpha_k^2 \mathcal{R}(x_{k-1}) - 2\alpha_k (\mathcal{R}(x_{k-1}) - \mathcal{R}(z)). \qquad (7)$$

By Proposition 1, the values $\mathcal{R}(x_{k-1})$ converge almost surely and, hence, they are bounded a.s. This and the condition $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$ imply $\sum_{k=1}^{\infty} \alpha_k^2 \mathcal{R}(x_{k-1}) < \infty$ almost surely. Thus, for each $z \in X^*$, relation (7) satisfies the assumptions of the supermartingale convergence (Theorem 1). According to this theorem, we have: the sequence $\{\|x_k - z\|\}$ is convergent almost surely for any $z \in X^*$, and

$$\sum_{k=1}^{\infty} \alpha_k (\mathcal{R}(x_k) - \mathcal{R}(z)) < \infty \quad a.s.$$

Thus, $\liminf_{k\to\infty} \mathcal{R}(x_k) = \mathcal{R}(z)$ *a.s.*, where $\mathcal{R}(z) = \mathcal{R}^*$ by optimality of z.

Since the sequence $\{\|x_k - z\|\}$ is convergent almost surely for any $z \in X^*$, the sequence $\{x_k\}$ must be bounded. Thus, $\{x_k\}$ must have limit points almost surely. By the above liminf relation and continuity of $\mathcal{R}$, we conclude that $\{x_k\}$ has one (random) limit point in the set X^* almost surely. Using this and the almost sure convergence of $\{\|x_k - z\|\}$ for any $z \in X^*$, we can see that $\{x_k\}$ converges and its (random) limit point lies in X^* almost surely. ∎

The result of Proposition 2 is reminiscent of the standard convergence results in stochastic gradient approximation schemes with the stepsize satisfying Assumption 3. However, unlike the standard results, Proposition 2 does not require the assumption that the sequence $\{x_k\}$ is bounded (see for example [7], [25]). It neither requires the growth condition on the magnitude of the stochastic gradient error of the form:

$$\mathsf{E}[\|\nabla_x R(x, \omega) - \nabla\mathcal{R}(x)\|^2] \leq A(1 + \|\nabla\mathcal{R}(x)\|^2) \quad (8)$$

for all $x \in \mathbb{R}^n$ and some scalar $A \geq 0$ (see [7], [6]). Such conditions are unnecessary for the function $\mathscr{R}(x)$, since $\mathscr{R}(x)$ has many special properties. Specifically, the property given by $\mathsf{E}\big[\|\nabla_x R(x, \omega)\|^2\big] = 2\mathscr{R}(x)$ in a way plays the role of the growth condition in (8).

Based on Proposition 2, we have the following characterization of the case when X^* is nonempty in terms of the behavior of the sequence $\{x_k\}$.

Proposition 3: Let Assumptions 1–3 hold. Then, X^* is nonempty if and only if the sequence $\{x_k\}$ generated by method (2) is convergent almost surely.

Proof: If X^* is nonempty, then the almost sure convergence of $\{x_k\}$ follows by Proposition 2. The converse statement follows by Proposition 1. In particular, Proposition 1 guarantees that $\liminf_{k\to\infty} \nabla\mathscr{R}(x_k) = 0$ *a.s.* Let $\{\bar\omega_k\}$ be a sample path such that its corresponding iterate sequence $\{\bar x_k\}$ converges to some point $\bar x$ and $\liminf_{k\to\infty} \nabla\mathscr{R}(\bar x_k) = 0$. By continuity of the gradient $\nabla\mathscr{R}(x)$ it follows that $\nabla\mathscr{R}(\bar x_k) \to \nabla\mathscr{R}(\bar x)$. By $\liminf_{k\to\infty} \nabla\mathscr{R}(\bar x_k) = 0$, we have $\nabla\mathscr{R}(\bar x) = 0$. Thus, $\bar x$ is the stationary point of $\mathscr{R}$ and, therefore, a solution to problem (1). ∎

Proposition 3 shows that under Assumptions 1–2, the almost sure convergence of the iterates $\{x_k\}$ generated with the diminishing stepsize α_k (subject to Assumption 3) is both *necessary and sufficient* for the existence of an optimal solution of problem (1). Such a result is not typical in stochastic optimization. For our problem, the result is possible due to the special properties of the objective function.

IV. Existence of a Common Point

In this section, we focus on investigating the existence of a point in common to all $X_i, i \in \mathscr{I}$. To do so, at first we limit our attention to the case where the index set $\mathscr{I}$ is finite, i.e., $\mathscr{I} = \{1, \ldots, m\}$ for some $m \geq 2$. At second, we consider a more aggressive stepsize in stochastic gradient method (2).

A. Algorithm for Existence

We let the stepsize be given by $\alpha_k = 1$ for all $k \geq 1$, so that the stochastic gradient method reduces to

$$x_k = \Pi_{\omega_k}[x_{k-1}] \qquad \text{for all } k \geq 1. \tag{9}$$

We refer to this method as *random projection method*. The method is related to the following set intersection problem:

$$\text{determine a point } x \text{ such that } \quad x \in X \triangleq \bigcap_{i=1}^m X_i, \tag{10}$$

which is related to optimization problem (1). Note that when X is nonempty, for problem (1) we have $\mathscr{R}^* = 0$ and $\mathscr{R}(x) = 0$ for all $x \in X$.

When the sets X_i are known to us, method (9) is just a random method for solving deterministic problem (10). In this case, to an extent, we have a freedom in choosing the distribution of the i.i.d. variables ω_k.

When the sets $\{X_i, i \in \mathscr{I}\}$ are not a priori available to us, the set X_{ω_k} is the observed (learned) set at time k. In this case, the method could also be viewed as an on-line learning algorithm for problem (10).

We now discuss convergence properties of the algorithm. We at first formal state our assumption on finiteness of $\mathscr{I}$.

Assumption 4: We have $\mathsf{Pr}\{\omega = i\} > 0$ for each $i \in \mathscr{I}$, where $\mathscr{I} = \{1, \ldots, m\}$.

We next establish almost sure convergence of the method for the case when the set X is nonempty.

Proposition 4: Let Assumptions 1, 2 and 4 hold. Also, assume that the set X is nonempty. Then, for the iterates $\{x_k\}$ generated by random projection method (9), we have almost surely for all $z \in X$ and all $k \geq 1$,

$$\mathsf{E}\big[\|x_k - z\|^2 \mid \mathscr{F}_{k-1}\big] \leq \|x_{k-1} - z\|^2 - 2\mathscr{R}(x_{k-1}).$$

Furthermore, $\{x_k\}$ converges almost surely to a random point in the set X.

Proof: Let $z \in X$ and note that $z \in X_{\omega_k}$ for all $k \geq 1$. Therefore, by using the projection property given in relation (6) (where $C = X_{\omega_k}$), we surely have for all $z \in X$ and $k \geq 1$,

$$\|x_k - z\|^2 \leq \|x_{k-1} - z\|^2 - \|x_k - x_{k-1}\|^2.$$

By taking the conditional expectation on the past $\mathscr{F}_{k-1}$, we obtain almost surely for any $z \in X$ and for all $k \geq 1$,

$$\mathsf{E}\big[\|x_k - z\|^2 \mid \mathscr{F}_{k-1}\big] \leq \|x_{k-1} - z\|^2 - \mathsf{E}\big[\|x_k - x_{k-1}\|^2 \mid \mathscr{F}_{k-1}\big].$$

By the definition of x_k and the i.i.d. property of ω_k, using $\mathsf{E}\big[\|x_k - x_{k-1}\|^2 \mid \mathscr{F}_{k-1}\big] = 2\mathscr{R}(x_{k-1})$ we can re-rewrite the preceding relation as: almost surely for any $z \in X$ and for all $k \geq 1$,

$$\mathsf{E}\big[\|x_k - z\|^2 \mid \mathscr{F}_{k-1}\big] \leq \|x_{k-1} - z\|^2 - 2\mathscr{R}(x_{k-1}).$$

By the supermartingale convergence (Theorem 1) we almost surely have that $\{\|x_k - z\|^2\}$ converges for every $z \in X$ and $\sum_{k=0}^\infty \mathscr{R}(x_k) < \infty$.

Since $\{\|x_k - z\|^2\}$ converges almost surely for every $z \in X$, it follows that the sequence $\{x_k\}$ is bounded *a.s.* Thus, it has a limit point *a.s.* In view of $\sum_{k=0}^\infty \mathscr{R}(x_k) < \infty$ *a.s.*, it follows that $\lim_{k\to\infty} \mathscr{R}(x_k) = 0$ *a.s.* Therefore, for every accumulation point $\tilde x$ of $\{x_k\}$, we have $\mathscr{R}(\tilde x) = 0$ almost surely (by continuity of $\mathscr{R}(x)$). Since $\mathscr{R}(\tilde x) = \frac{1}{2}\mathsf{E}\big[\|\tilde x - \Pi_{X_\omega}[\tilde x]\|^2\big]$ and $\mathsf{Pr}\{\omega = i\} > 0$ for each $i \in \mathscr{I}$ (Assumption 4), it follows

$$\|\tilde x - \Pi_{X_i}[\tilde x]\|^2 = 0 \qquad \text{for each } i \in \mathscr{I} \text{ a.s.}$$

Thus, every accumulation point $\tilde x$ of $\{x_k\}$ almost surely lies in the set X_i for all $i \in \mathscr{I}$ and, hence, in the set X *a.s.* Since $\{\|x_k - z\|^2\}$ converges for every $z \in X$ *a.s.*, it follows that $\{x_k\}$ converges to a random point in the set X *a.s.* ∎

Using Proposition 4, we now provide a characterization of the case $X \neq \emptyset$ in terms of the convergence of $\{x_k\}$.

Proposition 5: Let Assumptions 1, 2 and 4 hold. Then, the set X is nonempty if and only if the sequence $\{x_k\}$ generated by random projection method (9) converges almost surely and $\lim_{k\to\infty} \mathscr{R}(x_k) = 0$ a.s.

Proof: If the set X is nonempty, then the almost sure convergence of the sequence $\{x_k\}$ follows by Proposition 4.
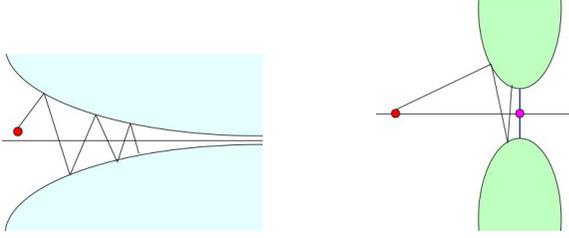
Fig. 1. The sets do not intersect, i.e., $X = \emptyset$. In the plot to the left, the distance between the sets is zero (i.e., $\mathscr{R}^* = 0$), which is reached asymptotically without attainment. The sequence $\{x_k\}$ does not converge, while $\lim_{k\to\infty} \mathscr{R}(x_k) = 0$. In the plot to the right, the average residual error $\mathscr{R}(x)$ is minimized at the midpoint between the sets, so $\mathscr{R}^* > 0$. The sequence $\{x_k\}$ converges to the midpoint and $\lim_{k\to\infty} \mathscr{R}(x_k) = \mathscr{R}^* > 0$.

Furthermore, by this proposition, the random limit point $\tilde{x}$ of $\{x_k\}$ is in the set X almost surely. Hence, $\lim_{k\to\infty} \mathscr{R}(x_k) = \mathscr{R}(\tilde{x}) = 0$ almost surely.

Assume now that the sequence $\{x_k\}$ is convergent and $\lim_{k\to\infty} \mathscr{R}(x_k) = 0$ almost surely. Let $\{\bar{\omega}_k\}$ be a sample path such that its corresponding iterate sequence $\{\bar{x}_k\}$ converges to some point $\bar{x}$. Then by continuity of $\mathscr{R}$ and the assumption, we have $\mathscr{R}(\bar{x}) = 0$. Since $\mathscr{R}(\bar{x}) = \frac{1}{2} \mathsf{E}\big[\|\bar{x} - \Pi_{X_\omega}[\bar{x}]\|^2\big]$, by Assumption 4 it follows

$$\|\bar{x} - \Pi_{X_i}[\bar{x}]\|^2 = 0 \qquad \text{for each } i \in \mathscr{I}.$$

Thus, $\bar{x}$ lies in the set X_i for each $i \in \mathscr{I}$, and hence in the set X a.s., thus showing that $X \neq \emptyset$. ∎

Figure 1 demonstrates the situations where the intersection set X is empty due to either non-convergence of $\{x_k\}$ or $\lim_{k\to\infty} \mathscr{R}(x_k) > 0$.

B. Convergence Rate

In our next result, we consider the averages of x_k and provide a convergence rate for the expected residual decrease with the number of iterations.

Proposition 6: Let Assumptions 1, 2 and 4 hold. Assume that the set X is nonempty. Let $\{x_k\}$ be the sequence generated by random projection method (9). Then, the averages $\hat{x}_t = \frac{1}{t} \sum_{k=1}^{t} x_{k-1}$ converge almost surely to a random point in the set X. Furthermore, we have

$$\mathsf{E}[\mathscr{R}(\hat{x}_t)] \leq \frac{1}{2t} \mathsf{E}\big[\operatorname{dist}^2(x_0, X)\big] \qquad \text{for all } t \geq 1.$$

Proof: By Proposition 4, $\{x_k\}$ converges almost surely to a random point in the set X. Therefore, the averages $\hat{x}_t = \frac{1}{t} \sum_{k=1}^{t} x_{k-1}$ also converge almost surely to the same random point in the set X.

By taking the expectation in the relation of Proposition 4 conditioned on $\mathscr{F}_0$, after rearranging the terms, we obtain almost surely for any $z \in X$ and $k \geq 1$,

$$2\mathsf{E}[\mathscr{R}(x_{k-1}) \mid \mathscr{F}_0] \leq \mathsf{E}\big[\|x_{k-1} - z\|^2 \mid \mathscr{F}_0\big] - \mathsf{E}\big[\|x_k - z\|^2 \mid \mathscr{F}_0\big].$$

By summing these relations from $k = 1$ to $k = t$ for some $t \geq 1$, we almost surely have for all $z \in X$ and $t \geq 1$,

$$2\sum_{k=1}^{t} \mathsf{E}[\mathscr{R}(x_{k-1}) \mid \mathscr{F}_0] \leq \|x_0 - z\|^2 - \mathsf{E}\big[\|x_t - z\|^2 \mid \mathscr{F}_0\big].$$

By discarding the last term in the preceding relation, letting $z = \Pi_X[x_0]$ and taking the total expectation, we obtain

$$2\sum_{k=1}^{t} \mathsf{E}[\mathscr{R}(x_{k-1})] \leq \mathsf{E}\big[\operatorname{dist}^2(x_0, X)\big] \qquad \text{for all } t \geq 1. \quad (11)$$

By dividing in relation (11) with $2t$, from the convexity of the function $\mathscr{R}(x)$ we have

$$\mathsf{E}\left[\mathscr{R}\left(\frac{1}{t}\sum_{k=1}^{t} x_{k-1}\right)\right] \leq \frac{1}{t}\sum_{k=1}^{t} \mathsf{E}[\mathscr{R}(x_{k-1})]$$
$$\leq \frac{1}{2t} \mathsf{E}\big[\operatorname{dist}^2(x_0, X)\big],$$

thus showing $\mathsf{E}[\mathscr{R}(\hat{x}_t)] \leq \frac{1}{2t} \mathsf{E}\big[\operatorname{dist}^2(x_0, X)\big]$. ∎

Using Proposition 6, we can provide a bound on the sum of the distances from the averages $\hat{x}_t$ to the sets X_i. We let

$$p_{\min} = \min_{1 \leq i \leq m} \Pr\{\omega = i\},$$

and note that $\mathscr{R}(x) = \frac{1}{2} \sum_{i=1}^{m} \Pr\{\omega = i\} \operatorname{dist}^2(x, X_\omega)$. Thus, we have for any $x \in \mathbb{R}^n$,

$$\max_{1 \leq i \leq m} \operatorname{dist}^2(x, X_i) \leq \sum_{i=1}^{m} \operatorname{dist}^2(x, X_i) \leq \frac{2}{p_{\min}} \mathscr{R}(x). \quad (12)$$

Using Proposition 6 and Markov inequality, we have the following probabilistic bound: for every $t \geq 1$,

$$\Pr\left\{\max_{i\in\mathscr{I}} \operatorname{dist}^2(\hat{x}_t, X_i) \geq \epsilon\right\} \leq \frac{1}{\epsilon}\frac{1}{p_{\min}}\frac{1}{t} \mathsf{E}\big[\operatorname{dist}^2(x_0, X)\big].$$

To see it, note that by Markov inequality, we have for $t \geq 1$,

$$\Pr\left\{\max_{i\in\mathscr{I}} \operatorname{dist}^2(\hat{x}_t, X_i) \geq \epsilon\right\} \leq \frac{1}{\epsilon} \mathsf{E}\left[\max_{i\in\mathscr{I}} \operatorname{dist}^2(\hat{x}_t, X_i)\right]. \quad (13)$$

From relation (12) with $x = \hat{x}_t$, we obtain for any $t \geq 1$,

$$\mathsf{E}\left[\max_{i\in\mathscr{I}} \operatorname{dist}^2(\hat{x}_t, X_i)\right] \leq \frac{2}{p_{\min}} \mathsf{E}[\mathscr{R}(\hat{x}_t)].$$

The result now follows by combining the preceding two relations with the estimate of Proposition 6.

We now consider the special case when each X_i is given by a linear constraint, and show a geometric convergence rate. The result is a stochastic counterpart of the similar result for the alternating projection algorithm [19], Theorem 5.

Proposition 7: Let Assumptions 2 and 4 hold. Assume that $X_i = \{x \in \mathbb{R}^n \mid a_i^T x \leq b_i\}$ with $a_i \neq 0$ for each i, and that the set X is nonempty. Let $\{x_k\}$ be the sequence generated by random projection method (9). Then, we have

$$\mathsf{E}\big[\operatorname{dist}^2(x_k, X)\big] \leq q^k \mathsf{E}\big[\operatorname{dist}^2(x_0, X)\big] \qquad \text{for all } k \geq 1,$$

with $q = 1 - p_{\min}\frac{c^2}{C^2}$, $C = \max_i \|a_i\|$ and $c > 0$ is a constant depending on the vectors a_i.

Proof: Letting $x = x_{k-1}$, $z = \Pi_X[x_{k-1}]$ and $Y = X_{\omega_k}$ in the projection relation (6) and using the fact $\operatorname{dist}(x_k, X) \leq \|x_k - \Pi_X[x_{k-1}]\|$, we obtain surely for all $k \geq 1$,

$$\operatorname{dist}^2(x_k, X) \leq \operatorname{dist}^2(x_{k-1}, X) - \|x_k - x_{k-1}\|^2. \quad (14)$$

Since each X_i is linear, we have

$$\|x_k - x_{k-1}\| = \frac{(a_{\omega_k}^T x - b_{\omega_k})^+}{\|a_{\omega_k}\|} \geq \frac{1}{C} \, (a_{\omega_k}^T x - b_{\omega_k})^+,$$

where $a^+ = \max\{0, a\}$ for $a \in \mathbb{R}$ and $C = \max_i \|a_i\|$. Therefore, we obtain

$$\mathsf{E}\big[\|x_k - x_{k-1}\|^2 \mid \mathscr{F}_{k-1}\big] \geq \frac{1}{C^2} \, p_{\min} \max_i \big((a_i^T x - b_i)^+\big)^2$$

$$\geq \frac{c^2}{C^2} \, p_{\min} \, \text{dist}^2(x_{k-1}, X).$$

The last inequality follows by weak sharp minima property of the function $\max_{1 \leq i \leq m}\{a_i^T x - b_i, 0\}$ with a scalar $c > 0$ that depends only on vectors a_i (see [9]). Taking the expectation in (14) conditioned on $\mathscr{F}_{k-1}$, using the preceding relation and then taking the total expectation, we obtain for all $k \geq 1$,

$$\mathsf{E}\big[\text{dist}^2(x_k, X)\big] \leq \left(1 - p_{\min} \frac{c^2}{C^2}\right) \mathsf{E}\big[\text{dist}^2(x_{k-1}, X)\big].$$

The desired estimate follows by recursion. ∎

A result similar to that of Proposition 7 can be shown assuming that the interior of the set X is nonempty. The proof uses some ideas from [19].

Proposition 8: Let Assumptions 1, 2 and 4 hold. Assume that the set X has a nonempty interior, i.e., there is a vector $\bar{x} \in X$ and $\delta > 0$ such that $\{z \in \mathbb{R}^n \mid \|z - \bar{x}\| \leq \delta\} \subset X$. Let $\{x_k\}$ be the sequence generated by random projection method (9) with a deterministic initial point x_0. Then,

$$\mathsf{E}\big[\text{dist}^2(x_k, X)\big] \leq q^k \, \text{dist}^2(x_0, X) \qquad \text{for all } k \geq 1,$$

with $q = 1 - p_{\min} \frac{\delta^2}{\|x_0 - \bar{x}\|^2}$.

Proof: By letting $x = x_{k-1}$, $Y = X_{\omega_k}$ and $z = \bar{x}$ in the projection relation (6), we obtain surely

$$\|x_k - \bar{x}\|^2 \leq \|x_0 - \bar{x}\|^2 \qquad \text{for all } k \geq 1. \quad (15)$$

Now using the projection relation (6) again with $x = x_{k-1}$, $Y = X_{\omega_k}$ and $z = \Pi_X[x_{k-1}]$, we can see that for all $k \geq 1$,

$$\text{dist}^2(x_k, X) \leq \text{dist}^2(x_{k-1}, X) - \|x_k - x_{k-1}\|^2.$$

Taking the conditional expectation and using

$$\mathsf{E}\big[\|x_k - x_{k-1}\|^2 \mid \mathscr{F}_{k-1}\big] \;=\; \mathsf{E}\big[\text{dist}^2(x_{k-1}, X_\omega\big]$$
$$\geq \; p_{\min} \max_i \text{dist}^2(x_{k-1}, X_i),$$

we have almost surely for all $k \geq 1$,

$$\mathsf{E}\big[\text{dist}^2(x_k, X) \mid \mathscr{F}_{k-1}\big] \leq \text{dist}^2(x_{k-1}, X)$$
$$- p_{\min} \max_i \text{dist}^2(x_{k-1}, X_i). \quad (16)$$

Let $x \in \mathbb{R}^n$ be arbitrary and define $\epsilon = \max_i \text{dist}^2(x, X_i)$. Consider the vector

$$y = \frac{\epsilon}{\epsilon + \delta} \, \bar{x} + \frac{\delta}{\epsilon + \delta} \, x.$$

We have $y \in X$. To see this note that we can write for each i,

$$y = \frac{\epsilon}{\epsilon + \delta} \, z + \frac{\delta}{\epsilon + \delta} \, \Pi_{X_i}[x] \text{ with } z = \bar{x} + \frac{\delta}{\epsilon}(x - \Pi_{X_i}[x]).$$

The vector $\Pi_{X_i}[x]$ belongs to the set X_i. The vector z satisfies $\|z - \bar{x}\| = \frac{\delta}{\epsilon}\|x - \Pi_{X_i}[x]\| \leq \delta$ where the last inequality follows by the definition of ϵ. By the interior point assumption on $\bar{x}$, it follows that $z \in X$ and therefore $z \in X_i$. Thus, for each i, the vector y can be written as a convex combination of two points in X_i, implying $y \in X_i$. Consequently $y \in X$, so that

$$\text{dist}(x, X) \leq \|x - y\| = \frac{\epsilon}{\epsilon + \delta} \, \|x - \bar{x}\|.$$

Using this and the definition of ϵ, we have

$$\text{dist}^2(x, X) \leq \frac{1}{\delta^2} \, \|x - \bar{x}\|^2 \, \max_i \text{dist}^2(x, X_i).$$

We now use the preceding relation with $x = x_{k-1}$ and combine it with inequality (15) to obtain for any $k \geq 1$,

$$\text{dist}^2(x_{k-1}, X) \leq \frac{1}{\delta^2} \, \|x_{k-1} - \bar{x}\|^2 \, \max_i \text{dist}^2(x_{k-1}, X_i)$$
$$\leq \frac{1}{\delta^2} \, \|x_0 - \bar{x}\|^2 \, \max_i \text{dist}^2(x_{k-1}, X_i).$$

Therefore, for all $k \geq 1$,

$$\max_i \text{dist}^2(x_{k-1}, X_i) \geq \frac{\delta^2}{\|x_0 - \bar{x}\|^2} \, \text{dist}^2(x_{k-1}, X).$$

Combining this relation with (16), we obtain

$$\mathsf{E}\big[\text{dist}^2(x_k, X) \mid \mathscr{F}_{k-1}\big] \leq \text{dist}^2(x_{k-1}, X)$$
$$- \frac{p_{\min}\delta^2}{\|x_0 - \bar{x}\|^2} \, \text{dist}^2(x_{k-1}, X).$$

Hence, for all $k \geq 1$,

$$\mathsf{E}\big[\text{dist}^2(x_k, X)\big] \leq \left(1 - \frac{p_{\min}\delta^2}{\|x_0 - \bar{x}\|^2}\right) \mathsf{E}\big[\text{dist}^2(x_{k-1}, X)\big],$$

and the desired relation follows. ∎

We can combine each of Propositions 7 and 8 with Markov inequality to obtain the following probabilistic bound:

$$\Pr\big\{\text{dist}^2(x_k, X) \geq \epsilon\big\} \leq \frac{q^k}{\epsilon} \, \mathsf{E}\big[\text{dist}^2(x_0, X)\big].$$

Let us note that this bound and the bound in (13) are optimized with respect to $p_{\min}$ when the distribution of ω is uniform over $\{1, \ldots, m\}$.

V. DISCUSSION

We have considered convex set intersection problem from the view point of stochastic optimization and we discussed a stochastic gradient algorithm for solving the problem. We studied the convergence of the method in relation with the existence of solutions to the optimization problem. We also investigated the convergence rate of the random projection algorithm and provided probabilistic bounds on the feasibility violation. Our immediate goal in future work is to consider extensions to a more general class of feasibility problems.

References

[1] S. Agmon, *The relaxation method for linear inequalities*, Canadian Journal of Mathematics **6** (1954), 382–393.

[2] T. Alamo, R. Tempo, and E.F. Camacho, *Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems*, IEEE Transactions on Automatic Control **54** (2009), no. 11, 2545–2559.

[3] N. Aronszajn, *Theory of reproducing kernels*, Transactions of the American Mathematical Society **68** (1950), no. 3, 337–404.

[4] H.H. Bauschke, *Projection algorithms: Results and open problems*, Inherently Parallel Algorithms in Feasibility and Optimization and their Applications (D. Butnariu, Y. Censor, and S. Reich, eds.), Elsevier, Amsterdam, Netherlands, 2001, pp. 11–22.

[5] H.H. Bauschke, P.L. Combettes, and D.R. Luke, *Hybrid projection-reflection method for phase retrieval*, Journal of the Optical Society of America A **20** (2003), no. 6, 1025–1034.

[6] D.P. Bertsekas and J.N. Tsitsiklis, *Gradient convergence in gradient methods*, SIAM Journal on Optimization **10** (2000), no. 3, 627–642.

[7] V.S. Borkar, *Stochastic approximation: A dynamical systems viewpoint*, Cambridge University Press, 2008.

[8] L.M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, Zh. Vychisl. Mat. & Mat. Fiz. **7** (1967), 620–631.

[9] J.V. Burke and M.C. Ferris, *Weak sharp minima in mathematical programming*, SIAM J. Control and Optimization **31** (1993), no. 6, 1340–1359.

[10] G. Calafiore and B.T Polyak, *Stochastic algorithms for exact and approximate feasibility of robust lmi's*, IEEE Transactions on Automatic Control **40** (2001), no. 11, 1755–1759.

[11] G.C. Calafiore, *Random convex programs*, http://www.optimization-online.org/DB-HTML/2009/12/2508.html, 2009.

[12] T.D. Capricelli and P.L. Combettes, *A convex programming algorithm for noisy discrete tomography*, Advances in Discrete Tomography and Its Applications (G.T. Herman and A. Kuba, eds.), Birkháuser, Boston, MA, 2007, pp. 207–226.

[13] P.L. Combettes, *Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections*, IEEE Transactions on Image Processing **6** (1997), no. 4, 493–506.

[14] F. Deutsch, *Rate of convergence of the method of alternating projections*, Parametric Optimization and Approximation (B. Brosowski and F. Deutsch, eds.), vol. 76, Birkhäuser, Basel, 1983, pp. 96–107.

[15] F. Deutsch and H. Hundal, *The rate of convergence for the cyclic projections algorithm I: Angles between convex sets*, Journal of Approximation Theory **142** (2006), 36–55.

[16] Z. Drezner, *The nested ball principle for the relaxation method*, Operations Research **31** (1983), no. 3, 587–590.

[17] F. Facchinei and J-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*, vol. I and II, Springer-Verlag New York, 2003.

[18] J.L. Goffin, *The relaxation method for solving systems of linear inequalities*, Mathematics of Operations Research **5** (1980), no. 3, 388–414.

[19] L.G. Gubin, B.T. Polyak, and E.V. Raik, *The method of projections for finding the common point of convex sets*, U.S.S.R. Computational Mathematics and Mathematical Physics **7** (1967), no. 6, 1211–1228.

[20] I. Halperin, *The product of projection operators*, Acta Scientiarum Mathematicarum **23** (1962), 96–99.

[21] K. Kiwiel, *Block-iterative surrogate projection methods for convex feasibility problems*, Linear Algebra and Its Applications **215** (1995), 225–259.

[22] T.S. Motzkin and I. Shoenberg, *The relaxation method for linear inequalities*, Canadian Journal of Mathematics **6** (1954), 393–404.

[23] B.T. Polyak, *Introduction to optimization*, Optimization Software, Inc., New York, 1987.

[24] _____, *Random algorithms for solving convex inequalities*, Inherently Parallel Algorithms in Feasibility and Optimization and their Applications (D. Butnariu, Y. Censor, and S. Reich, eds.), Elsevier, Amsterdam, Netherlands, 2001, pp. 409–422.

[25] S. Sundhar Ram, A. Nedić, and V.V. Veeravalli, *Incremental stochastic sub-gradient algorithms for convex optimization*, SIAM Journal on Optimization **20** (2009), no. 2, 691–717.

[26] J. von Neumann, *Functional operators*, Princeton University Press, 1950.