On the optimality of single projection variants of extragradient schemes for monotone stochastic variational inequality problems

Shisheng Cui and Uday V. Shanbhag*

April 26, 2019

Abstract

Classical extragradient schemes and their stochastic counterpart represent a cornerstone for resolving monotone variational inequality problems. Yet, such schemes have a per-iteration complexity of two projections on a convex set and two evaluations of the map, the former of which could be relatively expensive if $X$ is a complicated set. We consider two related avenues where the per-iteration complexity is significantly reduced: (i) A stochastic projected reflected gradient (SPRG) method requiring a single evaluation of the map and a single projection; and (ii) A stochastic subgradient extragradient (SSE) method that requires two evaluations of the map, a single projection, and a projection onto a halfspace (computable in closed form). Under suitable conditions, we prove almost sure (a.s.) convergence of the iterates to a random point in the solution set. Additionally, we show that under a variance-reduced framework, both schemes display a non-asymptotic rate of $O(1/K)$, matching their deterministic counterparts. To address constraints with a complex structure, we prove that random projection variants of both schemes also display a.s. convergence while displaying a rate of $O(1/\sqrt{K})$ in terms of the sub-optimality and infeasibility. Preliminary numerics support theoretical findings and the schemes outperform their standard extragradient counterparts in terms of the per-iteration complexity.

1 Introduction

This paper considers the solution of stochastic variational inequality problems, a stochastic generalization of the variational inequality problem. Given a set $X \subseteq \mathbb{R}^n$ and a map $F : \mathbb{R}^n \to \mathbb{R}^n$, the variational inequality problem $\text{VI}(X,F)$ requires finding a point $x^* \in X$ such that

$$F(x^*)^T(x - x^*) \geq 0, \quad \forall x \in X. \quad (\text{VI}(X,F))$$

In the stochastic generalization, the components of the map $F$ are expectation-valued; specifically $F_i(x) \triangleq \mathbb{E}[F_i(x,\xi(\omega))]$, where $\xi : \Omega \to \mathbb{R}^d$ is a random variable, $F_i : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ is a single-valued function, and the $\mathbb{E}[\cdot]$ denotes the expectation and the associated probability space being denoted by $(\Omega,F,\mathbb{P})$. In short, we are interested in a vector $x^* \in X$ such that

$$\mathbb{E}[F(x^*,\omega)]^T(x - x^*) \geq 0, \quad \forall x \in X, \quad (\text{SVI}(X,F))$$

* Cui and Shanbhag are with the Department of Indust. and Manuf. Engg., Penn. State Univ., University Park, PA 16802, USA. They are reachable at suc256, udaybag@psu.edu. The authors have been partly funded by NSF Grants 1246887 and 1400217 as well as the Gary and Sheila Bello Chair funds. A preliminary conference version of this work appeared in [10].
where \( \mathbb{E}[F(x, \omega)] = (\mathbb{E}[F_i(x, \omega)])_{i=1}^K \). The variational inequality problem is an immensely relevant problem that finds application in engineering, economics, and applied sciences (cf. \([8, 9, 12, 14, 29]\)). Increasingly, the stochastic generalization is of relevance and has found application in the study of a broad class of equilibrium problems under uncertainty. Of these, sample average approximation (SAA) scheme solves the expected value of the stochastic mapping which is approximated via the average over a large number of samples (cf. \([5, 7, 30, 34]\)). A counterpart to SAA schemes is the stochastic approximation (SA) methods where at each iteration, a sample of the stochastic mapping is used (cf. \([16, 18, 27]\)). Amongst the simplest of SA schemes are analogs of the standard projection-based schemes, which we review next.

### 1.1 Projection-based schemes and their variants

\[
x_{k+1} := \Pi_X(x_k - \gamma_k F(x_k)),
\]

(PG)

where \( \Pi_X(y) \) denotes the projection of \( y \) onto \( X \) and \( \gamma \) denotes the steplength. This method generally requires a strong monotonicity assumption on \( F \) to ensure convergence. An extension, suggested by Antipin \([1]\) and Korpelevich \([19]\), required that \( F \) be merely monotone:

\[
x_{k+\frac{1}{2}} := \Pi_X(x_k - \gamma F(x_k)),
\]

\[
x_{k+1} := \Pi_X(x_k - \gamma F(x_{k+\frac{1}{2}})).
\]

(EG)

In (EG) however, two projections were required at each iteration to obtain a new point and convergence was proved under the assumptions of Lipschitz continuity and monotonicity of the map \( F \). Naturally, when the set \( X \) is not necessarily a simple set, this projection operation by no means cheap. There have been several schemes in which merely monotone variational inequality problems can be addressed by taking a single projection operation and we consider two instances. In recent work, a *projected reflected gradient* (PRG) method was proposed by Malitsky \([21]\), requiring a single, rather than two, projections:

\[
x_{k+1} := \Pi_X(x_k - \gamma_k F(2x_k - x_{k-1})).
\]

(PRG)

Intuitively, this scheme has a similar structure to the projected gradient scheme taking a form with the following key distinction: Rather than evaluating the map at \( x_k \) (as in (PG)), the map is evaluated at the reflection of \( x_{k-1} \) in \( x_k \) which is \( x_k - (x_{k-1} - x_k) = 2x_k - x_{k-1} \). Remarkably, this simple modification allows for proving convergence of this scheme for merely monotone Lipschitz continuous maps \([21]\). An alternate modification of the extragradient method was proposed by Censor, Gibali and Reich and was referred to as the *subgradient extragradient method* (SE) \([6]\):

\[
x_{k+\frac{1}{2}} := \Pi_X(x_k - \gamma_k F(x_k)),
\]

\[
x_{k+1} := \Pi_{C_k}(x_k - \gamma_k F(x_{k+\frac{1}{2}})),
\]

(SE)

where \( C_k \triangleq \{ w \in \mathbb{R}^n \mid (x_k - \gamma_k F(x_k) - x_{k+\frac{1}{2}})^T(w - x_{k+\frac{1}{2}}) \leq 0 \} \). In (SE), the two projections are replaced by a projection onto the set and a second onto a halfspace (computable in closed form).
1.2 Stochastic variational inequality problems.

There have been schemes analogous to (PG) and (EG) in this regime with the key distinction that an evaluation of the map, namely $F(x_k)$, is replaced by $F(x_k, \omega_k)$, in the spirit of stochastic approximation [28]. Jiang and Xu [15] appear amongst the first who applied SA methods to solve stochastic variational inequalities. An extension to address merely monotone stochastic VIs was studied by Koshal et al. [20]. A regularized smoothing SA method to address stochastic VIs with non-Lipschitzian and merely monotone mappings was proposed in [35]. Recently, a class of prox generalization of SA methods were developed (cf. [24, 25, 36, 37]) for solving smooth and nonsmooth stochastic convex optimization problems and variational inequalities. For instance, a simple stochastic extension of the standard projection scheme for VI($X, F$) leads to a stochastic approximation scheme [28]:

$$x_{k+1} := \Pi_X(x_k - \gamma_k F(x_k, \omega_k)).$$  

(SP)

Similarly, an extragradient counterpart to (EG) is (SEG) and is defined below:

$$x_{k+\frac{1}{2}} := \Pi_X(x_k - \gamma_k F(x_k, \omega_k)),$$

$$x_{k+1} := \Pi_X(x_k - \gamma_k F(x_{k+\frac{1}{2}}), \omega_{k+\frac{1}{2}})).$$  

(SEG)

Fig. 1 illustrates (SEG) scheme. Extragradient-based schemes (and their stochastic mirror-prox counterparts) represent amongst the simplest of schemes for monotone SVIs (cf. [11, 16]). However, each iteration requires two projection steps, rather than one (as in (SP)). We summarize much of the prior results in Table 1. Given that this class of Monte-Carlo approximation schemes routinely requires 10s or 100s of thousands of steps, our interest lies in ascertaining whether projection-based schemes can be developed requiring a single projection step per iteration, reducing the per-iteration complexity by a factor of two. We consider two such schemes given a random point $x_0 \in X$: 

(i) **Stochastic projected reflected gradient schemes (SPRG)**:

$$x_{k+1} := \Pi_X(x_k - \gamma_k F(2x_k - x_{k-1}, \omega_k)).$$  

(SP)

and (ii) **Stochastic subgradient extragradient schemes (SSE)**.

$$x_{k+\frac{1}{2}} := \Pi_X(x_k - \gamma_k F(x_k, \omega_k)),$$

$$x_{k+1} := \Pi_{C_k}(x_k - \gamma_k F(x_{k+\frac{1}{2}}, \omega_{k+\frac{1}{2}})),$$  

(SSE)

where $C_k \triangleq \{ w \in \mathbb{R}^n | (x_k - \gamma_k F(x_k, \omega_k) - x_{k+\frac{1}{2}})^T(w - x_{k+\frac{1}{2}}) \leq 0 \}$. Clearly, the second projection is a simple optimization problem. Solving for $x_{k+1}$, we could obtain an equivalent scheme which requires a single projection (the proof is in appendix). Fig. 2 illustrate the steps of these schemes.
1.3 Incorporating variance reduction and random projections.

To mitigate computational complexity, we define two variable sample-size counterparts of (SPRG) and (SEG), where \( N_k \) samples of the map are utilized at iteration \( k \) to approximate the expectation. We define (i) Variable sample-size stochastic projected reflected gradient schemes:

\[
x_{k+1} := \Pi_X \left( x_k - \gamma_k \frac{\sum_{j=1}^{N_k} F(2x_k - x_{k-1}, \omega_{j,k})}{N_k} \right), \tag{v-SPRG}
\]

and (ii) Variable sample-size stochastic subgradient extragradient schemes:

\[
x_{k+\frac{1}{2}} := \Pi_X \left( x_k - \gamma_k \frac{\sum_{j=1}^{N_k} F(x_k, \omega_{j,k})}{N_k} \right),
\]

\[
x_{k+1} := \Pi_{C_k} \left( x_k - \gamma_k \frac{\sum_{j=1}^{N_k} F(x_{k+\frac{1}{2}}, \omega_{j,k+\frac{1}{2}})}{N_k} \right), \tag{v-SSE}
\]

where \( C_k \triangleq \{ w \in \mathbb{R}^n | (x_k - \gamma_k F(x_k, \omega_k) - x_{k+\frac{1}{2}})^T (w - x_{k+\frac{1}{2}}) \leq 0 \} \).

A difficulty arises when implementing such schemes on a complex set \( X \) when \( X \) is defined as the intersection of a large number of convex sets. Inspired by [32], we consider extending our work
to random projections when \( X \) is defined as the intersection of a finite number of sets:

\[
X = \bigcap_{i \in \mathcal{I}} X_i,
\]

where \( \mathcal{I} \) is a finite set and \( X_i \subseteq \mathbb{R}^n \) is closed and convex for all \( i \in \mathcal{I} \). The key distinction is that at each iteration, we project onto a random subset \( X_{l_k} \) rather than \( X \), where \( \{l_k\} \) is a sequence of random variables in the appropriate steps of (SPRG) and (SSE). In prior work, Nedić \[22, 23\] considered random projection algorithms for convex optimization problems with similarly defined sets and related schemes were subsequently considered for nonsmooth convex regimes \[3, 31, 33\]. Wang and Bertsekas \[32\] extended (SPG) to allow for projecting on a subset of constraints based on either a random projection technique on either a random or deterministic (such as cyclic projection) subset. We consider analogous generalizations to (SPRG) and (SEG):

(i) **Random projected stochastic projected reflected gradient schemes (r-SPRG):**

\[
x_{k+1} := \Pi_{l_k} (x_k - \gamma_k F(2x_k - x_{k-1}, \omega_k)),
\]

and (ii) **Random projected stochastic subgradient extragradient schemes (r-SSE).**

\[
x_{k+1} := \Pi_{C_k} (x_k - \gamma_k F(x_k, \omega_k)),
\]

\[
x_{k+1} := \Pi_{C_k} (x_k - \gamma_k F(x_{k+\frac{1}{2}}, \omega_{k+\frac{1}{2}})),
\]

where \( C_k \triangleq \{ w \in \mathbb{R}^n \mid (x_k - \gamma_k F(x_k, \omega_k) - x_{k+\frac{1}{2}})^T (w - x_{k+\frac{1}{2}}) \leq 0 \} \).

### 1.4 Contributions:

We summarize the key aspects of our schemes in Tables 2 and elaborate on these next:

| Variance-reduced schemes | Random projection |
|--------------------------|-------------------|
| Assump. | Rate | a.s. | Assump. | Rate, infeas. | a.s. |
| SPRG | monotone+weak-sharp | \( \mathcal{O}(1/K) \) | ✔ | monotone+weak-sharp | \( \mathcal{O}(1/\sqrt{K}) \), \( \mathcal{O}(1/\sqrt{K}) \) | ✔ |
| SSE | monotone | \( \mathcal{O}(1/K) \) | ✔ | monotone+weak-sharp | \( \mathcal{O}(1/\sqrt{K}) \), \( \mathcal{O}(1/\sqrt{K}) \) | ✔ |
| SEG | monotone | \( \mathcal{O}(1/K) \) | ✔ | monotone+weak-sharp | \( \mathcal{O}(1/\sqrt{K}) \), \( \mathcal{O}(1/\sqrt{K}) \) | ✔ |

(i) In Section II, we prove that in monotone regimes, the iterates produced by both (SPRG) and (SSE) converge almost surely (a.s.) to the solution for and the expectation of the distance function (for (SPRG)) or the gap function (for (SSE)) diminish at \( \mathcal{O}(1/K) \), matching the deterministic rate of convergence.

(ii) In Section IV, under merely monotone settings with a weak-sharpness requirement, random projection variants of (SPRG) and (SSE) are examined and we proceed to prove a.s. convergence of the iterates to the solution set. Additionally, we proceed to show that the expected distance to both the optimal solution set \( X^* \) and the feasible set \( X \) diminish at the rate of \( \mathcal{O}(1/\sqrt{K}) \).
(iii) In Section V, preliminary numerics are observed support our expectations based on the theoretical findings.

2 Background and Assumptions

We consider the schemes (SPRG) and (SSE) where \( x_0 \in X \) is a random initial point and \( \{ \gamma_k \} \) denotes the steplength sequence. We begin by imposing an assumption on the map \( F \) which will be valid throughout the remainder of this paper.

**Assumption 1.** The mapping \( F \) is L-Lipschitz continuous and monotone on \( \mathbb{R}^n \); i.e. for all \( x, y \in \mathbb{R}^n \),

\[
\| F(x) - F(y) \| \leq L \| x - y \| \text{ and } (F(x) - F(y))^T (x - y) \geq 0.
\]

We often impose a boundedness requirement on the set \( X \) and \( F(x^*) \).

**Assumption 2.** The set \( X \) is bounded, i.e., there exists a scalar \( B > 0 \) such that \( \| x - y \| \leq B \) for all \( x, y \in X \).

**Assumption 3.** There exists a constant \( C > 0 \) such that \( \| F(x^*) \| \leq C \).

In some instances, a weak-sharpness requirement is imposed on VI(\( X, F \)).

**Assumption 4 (Weak sharpness).** The variational inequality problem VI(\( X, F \)) satisfies the weak sharpness property implying that there exists an \( \alpha > 0 \) such that for all \( x \in X \), \( (x - x^*)^T F(x^*) \geq \alpha \text{dist}(x, X^*) \).

The following lemma is used in our analysis proofs and may be found in [2].

**Lemma 5.** Let \( X \) be nonempty closed convex set in \( \mathbb{R}^n \). Then for all \( y \in X \) and for any \( x \in \mathbb{R}^n \), we have that the following hold: (i) \( (\Pi_X(x) - x)^T (y - \Pi_X(x)) \geq 0 \); and (ii) \( \| \Pi_X(x) - y \|^2 \leq \| x - y \|^2 - \| x - \Pi_X(x) \|^2 \).

We assume the presence of a stochastic oracle that can provide a conditionally unbiased estimator of \( F(x) \), given by \( F(x, \omega) \) such that \( \mathbb{E}[F(x, \omega) \mid x] = F(x) \). Define \( w_k \triangleq F(x_k, \omega_k) - F(x_k), \bar{w}_k \triangleq \frac{\sum_{j=1}^{N_k} F(x_{k+j}, \omega_{j,k})}{N_k} - F(x_k), w_{k+1/2} \triangleq F(x_{k+1/2}, \omega_{k+1/2}) - F(x_{k+1/2}) \) and \( \bar{w}_{k+1/2} \triangleq \frac{\sum_{j=1}^{N_k} F(x_{k+j+1/2}, \omega_{j,k})}{N_k} - F(x_{k+1/2}), \) where \( N_k \) denotes the batch-size of sampled maps \( F(x, \omega, j, k) \) at iteration \( k \). Furthermore, let \( \mathcal{F}_k \) denote the history up to iteration \( k \), i.e., \( \mathcal{F}_k \triangleq \{ x_0, \omega_0, \omega_{1/2}, \omega_1, \cdots, \omega_{k-1}, \omega_{k-1/2} \} \) and \( \mathcal{F}_{k+1/2} \triangleq \mathcal{F}_k \cup \{ \omega_k \} \).

**Assumption 6.** At an iteration \( k \), the following hold in an a.s. sense: (i) The conditional means \( \mathbb{E}[\mathcal{F}_k] \) and \( \mathbb{E}[w_{k+1/2} \mid \mathcal{F}_{k+1/2}] \) are zero for all \( k \) in an a.s. sense; (ii) The conditional second moments are bounded in an a.s. sense or \( \mathbb{E}[\| w_k \|^2 \mid \mathcal{F}_k] \leq \nu^2 \) and \( \mathbb{E}[\| w_{k+1/2} \|^2 \mid \mathcal{F}_{k+1/2}] \leq \nu^2 \) for all \( k \) in an a.s. sense.

**Assumption 7.** The diminishing sequence \( \gamma_k \) is square-summable but non-summable: \( \sum_{k=0}^{\infty} \gamma_k^2 < \infty, \sum_{k=0}^{\infty} \gamma_k = \infty \).

The following super-martingale convergence Lemma is essential to our proof [26].
Lemma 8. Let $v_k$, $u_k$, $\delta_k$, $\psi_k$ be nonnegative random variables adapted to $\sigma$-algebra $F_k$, and let the following relations hold almost surely:

$$
\mathbb{E}[u_{k+1} \mid F_k] \leq (1 + u_k)v_k - \delta_k + \psi_k, \quad \forall k; \quad \sum_{k=0}^{\infty} u_k < \infty, \quad \text{and} \quad \sum_{k=0}^{\infty} \psi_k < \infty.
$$

Then a.s., we have that $\lim_{k \to \infty} u_k = v$ and $\sum_{k=0}^{\infty} \delta_k < \infty$, where $v \geq 0$ is a random variable.

3 Convergence Analysis for (v-SPRG) and (v-SSE)

3.1 Stochastic Projected Reflected Gradient Schemes

In this subsection, we prove the a.s. convergence of the iterates produced by (SPRG) when $F$ is a Lipschitz continuous and monotone map on $\mathbb{R}^n$, satisfying a weak-sharpness requirement. We begin with an intermediate lemma that relates the error in consecutive iterates.

Lemma 9. Let Assumptions \[\text{[1, 4, and 6]}\] hold and let $0 < \gamma_k = \gamma \leq \frac{1}{8L}$ for all $k$. Consider a sequence generated by (v-SPRG). For any $x_0 \in X$, the following holds for all $k \geq 0$:

$$
\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 + \frac{3}{4}\|x_{k+1} - y_k\|^2 + 2\gamma F(x^*)^T(x_k - x^*)
$$

Then a.s., we have that $\lim_{k \to \infty} u_k = v$ and $\sum_{k=0}^{\infty} \delta_k < \infty$, where $v \geq 0$ is a random variable.

Proof. Define $y_k \triangleq 2x_k - x_{k-1}$ for all $k \geq 1$ and $\bar{F}(y_k) \triangleq \sum_{j=1}^{N_k} F(y_k, \omega_{k,j}) \mathbb{I}_{N_k}$. By Lemma \[\text{[5(ii)]}\] and noting that $x_{k+1} = \Pi_X(x_k - \gamma_k \bar{F}(y_k))$ and $\bar{F}(y_k) = F(y_k) + \bar{w}_k$, the following holds for $x_{k+1}$ and any solution $x^*$.

$$
\|x_{k+1} - x^*\|^2 \leq \|x_k - \gamma_k \bar{F}(y_k) - x^*\|^2 - \|x_k - \gamma_k \bar{F}(y_k) - x_{k+1}\|^2
$$

$$
= \|x_k - x^*\|^2 - \|x_{k+1} - x_k\|^2 - 2\gamma_k(F(y_k) + \bar{w}_k)^T(x_{k+1} - x^*). \quad (1)
$$

Since $F$ is monotone over $\mathbb{R}^n$, by adding $2\gamma_k(F(y_k) - F(x^*))^T(y_k - x^*)$ to the rhs of (1), we obtain:

$$
\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \|x_{k+1} - x_k\|^2 + 2\gamma_k(F(y_k) - F(x^*))^T(y_k - x^*)
$$

$$
- 2\gamma_k(F(y_k) + \bar{w}_k)^T(x_{k+1} - x^*)
$$

$$
= \|x_k - x^*\|^2 - \|x_{k+1} - x_k\|^2 + 2\gamma_k F(y_k)^T(y_k - x_{k+1}) + 2\gamma_k F(y_k)^T(x_{k+1} - x^*)
$$

$$
- 2\gamma_k F(x^*)^T(y_k - x^*) - 2\gamma_k F(y_k)^T(x_{k+1} - x^*) + 2\gamma_k \bar{w}_k^T(y_k - x_{k+1}) - 2\gamma_k \bar{w}_k^T(y_k - x^*)
$$

$$
= \|x_k - x^*\|^2 - \|x_{k+1} - x_k\|^2 + 2\gamma_k F(y_k + \bar{w}_k)^T(y_k - x_{k+1}) - 2\gamma_k F(x^*)^T(y_k - x^*)
$$

$$
= \|x_k - x^*\|^2 - \|x_{k+1} - x_k\|^2 + 2\gamma_k (F(y_k) - F(y_{k-1}))^T(y_k - x_{k+1})
$$

$$
+ 2\gamma_k (F(y_{k-1}) + \bar{w}_k)^T(y_k - x_{k+1}) - 2\gamma_k (F(x^*) + \bar{w}_k)^T(y_k - x^*). \quad (2)
$$
Since \(x_{k+1}, x_{k-1} \in X\), by Lemma 5(i), we may conclude that
\[
(x_k - x_{k-1} + \gamma_{k-1}(F(y_{k-1}) + \bar{w}_{k-1}))^T(x_k - x_{k+1}) \leq 0 \quad \text{and} \quad (x_k - x_{k-1} + \gamma_{k-1}(F(y_{k-1}) + \bar{w}_{k-1}))^T(x_k - x_{k-1}) \leq 0.
\]
Adding these two inequalities yields the following:
\[
(x_k - x_{k-1} + \gamma_{k-1}(F(y_{k-1}) + \bar{w}_{k-1}))^T(y_k - x_{k+1}) \leq 0,
\]
since \(y_k = 2x_k - x_{k-1}\), leading to the following inequality:
\[
2\gamma_{k-1}(F(y_{k-1}) + \bar{w}_{k-1})^T(y_k - x_{k+1}) \leq 2(x_k - x_{k-1})^T(x_{k+1} - y_k) = 2(y_k - x_k)^T(x_{k+1} - y_k) = \|x_{k+1} - x_k\|^2 - \|x_k - y_k\|^2 - \|x_{k+1} - y_k\|^2,
\]
where the first equality follows from recalling that \(y_k = 2x_k - x_{k-1}\). Now, we may bound \(2\gamma_k(F(y_{k-1}) + \bar{w}_{k-1})^T(y_k - x_{k+1})\) as follows:
\[
\text{Term 2} = 2\gamma_k(F(y_{k-1}) + \bar{w}_{k-1})^T(y_k - x_{k+1}) = 2\gamma_k(F(y_{k-1}) + \bar{w}_{k-1})^T(y_k - x_{k+1})
\]
\[- 2\gamma_k(F(y_{k-1}) + \bar{w}_{k-1})^T(y_k - x_{k+1}) + 2\gamma_k(F(y_{k-1}) + \bar{w}_{k-1})^T(y_k - x_{k+1})
\]
\[= 2\gamma_k(\bar{w}_k - \bar{w}_{k-1})^T(y_k - x_{k+1}) + 2\left(\frac{\gamma_k}{\gamma_{k-1}}\right)\gamma_{k-1}(F(y_{k-1}) + \bar{w}_{k-1})^T(y_k - x_{k+1})
\]
\[\leq 8\gamma_k^2\|\bar{w}_k - \bar{w}_{k-1}\|^2 + \frac{1}{8}\|x_{k+1} - y_k\|^2 - \frac{\gamma_k}{\gamma_{k-1}}\|x_{k+1} - y_k\|^2 + \frac{2\gamma_k}{\gamma_{k-1}}\|x_k - x_{k+1}\|^2 - \frac{\gamma_k}{\gamma_{k-1}}\|x_k - y_k\|^2
\]
\[= 8\gamma_k^2\|\bar{w}_k - \bar{w}_{k-1}\|^2 + \left(\frac{1}{8} - \frac{\gamma_k}{\gamma_{k-1}}\right)\|x_{k+1} - y_k\|^2 + \frac{2\gamma_k}{\gamma_{k-1}}\|x_k - x_{k+1}\|^2 - \frac{\gamma_k}{\gamma_{k-1}}\|x_k - y_k\|^2,
\]
where \(2\gamma_k(\bar{w}_k - \bar{w}_{k-1})^T(y_k - x_{k+1}) \leq 8\gamma_k^2\|\bar{w}_k - \bar{w}_{k-1}\|^2 + \frac{1}{8}\|x_{k+1} - y_k\|^2\) and inequality (3) allows for bounding \(2\gamma_{k-1}(F(y_{k-1}) + \bar{w}_{k-1})^T(y_k - x_{k+1})\). Next we estimate \((F(y_k) - F(y_{k-1}))^T(y_k - x_{k+1})\). By the Cauchy-Schwarz inequality and the Lipschitz continuity of the map (Ass. 1), it follows that
\[
\text{Term 1} = 2\gamma_k(F(y_k) - F(y_{k-1}))^T(y_k - x_{k+1}) \leq 2\gamma_k\|F(y_k) - F(y_{k-1})\|\|y_k - x_{k+1}\|
\]
\[\leq 2\gamma_k L\|y_k - y_{k-1}\|\|y_k - x_{k+1}\| \leq 8\gamma_k^2 L^2\|y_k - y_{k-1}\|^2 + \frac{1}{8}\|x_{k+1} - y_k\|^2
\]
\[\leq 16\gamma_k^2 L^2\|x_k - y_{k-1}\|^2 + 16\gamma_k^2 L^2\|x_k - y_k\|^2 + \frac{1}{8}\|x_{k+1} - y_k\|^2,
\]
where (4) follows from \(\|u + v\|^2 \leq 2\|u\|^2 + 2\|v\|^2\). Using (4) and (5), we deduce from (2) that
\[
\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \left(1 - \frac{\gamma_k}{\gamma_{k-1}}\right)\|x_{k+1} - x_k\|^2 - \left(\frac{\gamma_k}{\gamma_{k-1}} - 16\gamma_k^2 L^2\right)\|x_k - y_k\|^2
\]
\[- \left(\frac{\gamma_k}{\gamma_{k-1}} - \frac{1}{4}\right)\|x_{k+1} - y_k\|^2 + 16\gamma_k^2 L^2\|x_k - y_{k-1}\|^2 + 8\gamma_k^2\|\bar{w}_k - \bar{w}_{k-1}\|^2
\]
\[- 2\gamma_k(F(x^*) + \bar{w}_k)^T(y_k - x^*),
\]
(7)
By assumption, $\gamma_k = \gamma \leq 1/(8L)$, for all $k$,

$$16\gamma_k^2 L^2 \leq \frac{1}{4} \leq \left(\frac{\gamma_k-1}{\gamma_k-2} - \frac{1}{4}\right).$$

Consequently, from (7) and by invoking (8), we may conclude the following:

$$\|x_{k+1} - x^*\|^2 + \left(\frac{\gamma_k}{\gamma_k-1} - \frac{1}{4}\right) \|x_{k+1} - y_k\|^2 \leq \|x_k - x^*\|^2 + \left(\frac{\gamma_k-1}{\gamma_k-2} - \frac{1}{4}\right) \|x_k - y_{k-1}\|^2$$

$$+ 8\gamma_k^2 \|w_k - w_{k-1}\|^2 - \left(\frac{\gamma_k}{\gamma_k-1} - 16\gamma_k^2 L^2\right) \|x_k - y_k\|^2 - 2\gamma_k F(x^*)^T(y_k - x^*) + \gamma_k \bar{w}_k^T(y_k - x^*).$$

We may bound $2\gamma_k F(x^*)^T(y_k - x^*)$ as follows:

$$- 2\gamma_k F(x^*)^T(y_k - x^*) = -2\gamma_k F(x^*)^T(x_k - x^*) - 2\gamma_k F(x^*)^T(x_{k-1} - x^*)$$

$$\leq -2\gamma_k F(x^*)^T(x_k - x^*) - 2\gamma_k F(x^*)^T(x_{k-1} - x^*) + 2\gamma_k F(x^*)^T(x_{k-1} - x^*)$$

By the weak sharpness property, we have that $F(x^*)^T(x_k - x^*) \geq \alpha \text{dist}(x_k, X^*)$, which together with (10), implies that

$$\|x_{k+1} - x^*\|^2 + \frac{3}{4} \|x_{k+1} - y_k\|^2 + 2\gamma_k F(x^*)^T(x_k - x^*)$$

$$\leq \|x_k - x^*\|^2 + \frac{3}{4} \|x_k - y_{k-1}\|^2 + 2\gamma_k F(x^*)^T(x_{k-1} - x^*)$$

$$+ 8\gamma_k^2 \|\bar{w}_k - \bar{w}_{k-1}\|^2 - (1 - 16\gamma_k^2 L^2) \|x_k - y_k\|^2 - 2\gamma_k \alpha \text{dist}(x_k, X^*) - 2\gamma_k \bar{w}_k^T(y_k - x^*).$$

\(\square\)

With this lemma, we now analyze convergence of (v-SPRG).

**Proposition 10** (a.s. convergence of (v-SPRG)). *Consider the scheme (v-SPRG). Let Assumptions \(4\) and \(4\) hold. Let $0 < \gamma_k = \gamma \leq \frac{1}{4}$ for all $k \geq 0$ and $\sum_{k=1}^{\infty} \frac{1}{\gamma_k} < \infty$. Then for any $x_0 \in X$, a sequence generated by (v-SPRG) converges to a solution $x^* \in X$ in an a.s. sense.*

**Proof.** Using (10) and taking expectations conditioned on $F_k$,

$$\mathbb{E}\left[\|x_{k+1} - x^*\|^2 + \frac{3}{4} \|x_{k+1} - y_k\|^2 + 2\gamma_k F(x^*)^T(x_k - x^*)|F_k\right]$$

$$\leq \|x_k - x^*\|^2 + \frac{3}{4} \|x_k - y_{k-1}\|^2 + 2\gamma_k F(x^*)^T(x_{k-1} - x^*) - 2\alpha \text{dist}(x_k, X^*)$$

$$+ 8\gamma_k^2 \mathbb{E}[\|\bar{w}_k - \bar{w}_{k-1}\|^2|F_k] - (1 - 16\gamma_k^2 L^2) \|x_k - y_k\|^2$$

$$\leq \|x_k - x^*\|^2 + \frac{3}{4} \|x_k - y_{k-1}\|^2 + 2\gamma_k F(x^*)^T(x_{k-1} - x^*) - 2\alpha \text{dist}(x_k, X^*)$$

$$+ 32\gamma_k^2 \frac{\nu^2}{N_k} - (1 - 16\gamma_k^2 L^2) \|x_k - y_k\|^2 = v_k - \delta_k + \psi_k,$$

where $v_k$, $\delta_k$, and $\psi_k$ are nonnegative random variables defined as

$$v_k \triangleq \|x_k - x^*\|^2 + \frac{3}{4} \|x_k - y_{k-1}\|^2 + 2\gamma_k F(x^*)^T(x_k - x^*),$$
\[ \delta_k \triangleq (1 - 16\gamma^2L^2) \| x_k - y_k \|^2 + 2\alpha\gamma \text{dist} (x_k, X^*) \quad \text{and} \quad \psi_k \triangleq 32\gamma^2 \nu^2 N_k. \]

We note that \( \sum_k \psi_k < \infty \) since \( \sum_k \frac{1}{N_k} < \infty \) and \( \delta_k \geq 0 \) since \( \text{dist}(x_k, X^*) \geq 0 \) for all \( k \) and
\[ (1 - 16\gamma^2L^2) \geq \frac{1}{4}. \]

We may now invoke Lemma 8 to claim that \( v_k \to \bar{v} \geq 0 \) and \( \sum_k \delta_k < \infty \) in an a.s. sense, implying the following holds a.s.:
\[ \infty > \sum_k \left( (1 - 16\gamma^2L^2) \| x_k - y_k \|^2 + 2\alpha\gamma \text{dist} (x_k, X^*) \right) \]
\[ \geq \sum_k \left( \left(1 - \frac{1}{4}\right) \| x_k - y_k \|^2 + 2\alpha\gamma \text{dist} (x_k, X^*) \right) = \sum_k \left( \frac{3}{4}\| x_k - y_k \|^2 + 2\alpha\gamma \text{dist} (x_k, X^*) \right), \]

where the second inequality follows from \( \gamma \leq 1/(8L) \). Consequently, we have that
\[ \infty > \sum_k \left( \frac{3}{4}\| x_k - y_k \|^2 + 2\alpha\gamma \text{dist} (y_k, X^*) \right). \]

It follows that in an a.s. sense,
\[ \infty > \sum_k \| x_k - y_k \|^2 = \sum_k \| x_k - x_{k-1} \|^2. \] (13)

From (13), \( x_k - y_k \to 0 \) as \( k \to \infty \) in an a.s. sense. Furthermore, in an a.s. sense, \( \sum_k \alpha \gamma \text{dist} (x_k, X^*) < \infty \) and in an a.s. sense, we have
\[ \lim_{k \to \infty} \text{dist}(x_k, X^*) = 0. \]

This implies that the entire sequence of \( \{x_k\} \) converges to a point in \( X^* \) in an a.s. sense. Since \( \{x_k\} \) and \( \{y_k\} \) have the same limit points almost surely, we have that \( \{y_k\} \) also converges to a point in \( X^* \) in an a.s. sense. \( \square \)

We are now in a position to derive a rate statement for the sequence of iterates. Importantly, we attain a rate of \( \mathcal{O}(1/K) \) in terms of the distance to the solution, an improvement over the rate of \( \mathcal{O}(1/\sqrt{K}) \) by using an increasing batch-size sequence \( \{N_k\} \).

**Proposition 11 (Rate statement for (SPRG)).** Consider the \( (\nu\text{-SPRG}) \) scheme. Let Assumptions 1, 2, and 6 hold. Let \( 0 < \gamma_k = \gamma \leq 1/8L \) for all \( k \geq 0 \), \( \sum_{k=1}^{\infty} \frac{1}{N_k} < M \), and \( \bar{x}_K \triangleq \sum_{k=1}^{K} x_k/K \).

(1) Then for any \( K \), \( \mathbb{E}[\text{dist}(\bar{x}_K, X^*)] \leq \mathcal{O} \left( \frac{1}{K} \right) \). (2) Suppose \( N_k = \lfloor k^a \rfloor \), for \( a > 1 \). The oracle complexity to obtain an \( x_K \) such that \( \mathbb{E}[\text{dist}(x_K, X^*)] \leq \epsilon \) is bounded as follows. \( \sum_{k=1}^{K} N_k \leq \mathcal{O} \left( \frac{1}{\epsilon^a} \right) \).

**Proof.** (1). From (12), taking expectations on both sides and by summing over \( k \) from 1 to \( K \), we have the following inequality:
\[ \sum_{k=1}^{K} 2\alpha\gamma \mathbb{E}[\text{dist}(\bar{x}_k, X^*)] \leq \mathbb{E}[\| x_1 - x^* \|^2] + \frac{3}{4} \mathbb{E}[\| x_1 - y_0 \|^2] + 2\gamma F(x^*)^T (x_1 - x^*) + 32\gamma^2 \nu^2 \sum_{k=1}^{K} \frac{1}{N_k}. \]
Dividing both sides by $2K\alpha\gamma$, we have the following sequence of inequalities:

$$\sum_{k=1}^{K} 2\alpha\gamma E[\text{dist}(x_k, x^*)] \leq \frac{\mathbb{E}[\|x_1 - x^*\|^2]}{2} + \frac{\sqrt{\frac{7}{2}} B^2 + 2\gamma BC}{2K\alpha\gamma} + \frac{16\gamma^2 \nu^2 \sum_{k=1}^{K} \frac{1}{N_k}}{K\alpha},$$

(14)

where the second inequality follows from the boundedness of $X$. By the convexity of the distance function, we have that

$$\mathbb{E}[\text{dist}(\bar{x}_K, x^*)] \leq \frac{\sum_{k=1}^{K} 2\alpha\gamma E[\text{dist}(x_k, x^*)]}{2} \frac{\mathbb{E}[\|x_1 - x^*\|^2]}{2\sum_{k=1}^{K} \alpha\gamma},$$

where $\bar{x}_K \triangleq \sum_{k=1}^{K} x_k / K$.

By choosing $N_k$ such that $\sum_{k=1}^{K} \frac{1}{N_k} < M < \infty$, we have

$$\mathbb{E}[\text{dist}(\bar{x}_K, x^*)] \leq \frac{1}{K} \left( \frac{\sqrt{\frac{7}{2}} B^2 + 2\gamma BC}{2\alpha\gamma} + \frac{16\gamma^2 \nu^2 M}{\alpha} \right) \leq O \left( \frac{1}{K} \right). \triangleq \hat{C}.$$

(2). It follows from Proposition III(1) that for $\epsilon$ sufficiently small,

$$\sum_{k=1}^{K} N_k \leq \sum_{k=1}^{K} \mathbb{E}[\|x_k - x^*\|^2] \leq \left\lfloor \frac{\hat{C}}{\epsilon^2} \right\rfloor \leq \frac{\hat{C}}{\epsilon^2} \leq \left( \frac{\hat{C}}{\epsilon^2} \right),$$

where the last inequality follows from $a > 1$.

3.2 Stochastic Subgradient Extragradient Schemes

We begin by proving the a.s. convergence of the iterates produced by (v-SSE). Unlike (v-SPRG), this scheme does not require an assumption of weak sharpness but mere monotonicity suffices.

**Proposition 12 (a.s. convergence of (v-SSE)).** Consider the scheme (v-SSE). Let Assumptions 1 and 2 hold. Suppose $0 < \gamma_k = \gamma \leq \frac{1}{2\alpha}$ and $\sum_{k=1}^{K} \frac{1}{N_k} < M$. Then any sequence generated by (v-SSE) converges to a solution $x^* \in X$ in an a.s. sense.

**Proof.** By Lemma B(i) we have

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - \gamma_k (F(x_k + \frac{1}{2}) + \bar{w}_{k+1}) - x^*\|^2 - \|x_k - \gamma_k (F(x_k + \frac{1}{2}) + \bar{w}_{k+1}) - x_k + 1\|^2$$

$$= \|x_k - x^*\|^2 - \|x_k - x_{k+1}\|^2 + 2\gamma_k (F(x_k + \frac{1}{2}) + \bar{w}_{k+1})^T (x^* - x_{k+1}). \quad (15)$$

It is clear that

$$F(x_{k+1} + \frac{1}{2})^T (x_{k+1} - x^*) = F(x_{k+1} + \frac{1}{2})^T (x_{k+1} - x_{k+1} + \frac{1}{2} + \frac{1}{2} - x^*)$$

$$= F(x_{k+1} + \frac{1}{2})^T (x_{k+1} - x_{k+1}) + F(x_{k+1} + \frac{1}{2})^T (x_{k+1} + \frac{1}{2} - x^*). \quad (16)$$
Substituting (16) in (15), we obtain
\[
\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \|x_k - x_{k+1}\|^2 + 2\gamma_k F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x_{k+1}) - F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) \\
+ 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+1}) \\
= \|x_k - x^*\|^2 - \|x_k - x_{k+\frac{1}{2}} + x_{k+\frac{1}{2}} - x_{k+1}\|^2 + 2\gamma_k F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x_{k+1}) \\
- F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+1}) \\
= \|x_k - x^*\|^2 - \|x_k - x_{k+\frac{1}{2}}\|^2 - \|x_{k+\frac{1}{2}} - x_{k+1}\|^2 - 2(x_k - x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x_{k+1}) \\
+ 2\gamma_k F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x_k) - F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+1}) \\
\leq \|x_k - x^*\|^2 - \|x_k - x_{k+\frac{1}{2}}\|^2 - \|x_{k+\frac{1}{2}} - x_{k+1}\|^2 - 2(x_k - x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) \\
+ 2\gamma_k \|x_k + x_{k+\frac{1}{2}}\|\|F(x_k) - F(x_{k+\frac{1}{2}})\| + 2\gamma_k (\bar{w}_k - \bar{w}_{k+\frac{1}{2}})^T(x_k - x_{k+\frac{1}{2}}) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) \\
\leq \|x_k - x^*\|^2 - \|x_k - x_{k+\frac{1}{2}}\|^2 - \|x_{k+\frac{1}{2}} - x_{k+1}\|^2 + \frac{1}{2}\|x_k - x_{k+\frac{1}{2}} + x_{k+\frac{1}{2}} - x_{k+1}\|^2 + 2\gamma_k^2 L^2\|x_k - x_{k+\frac{1}{2}}\|^2 \\
+ 2\gamma_k^2 \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 + \frac{1}{2}\|x_k + x_{k+\frac{1}{2}}\|^2 - F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) \\
= \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2)\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 \\
- F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) \\
= \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2)\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 \\
- F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}),
\]
By definition of $C_k$, we have
\[
(x_{k+1} - x_{k+\frac{1}{2}})^T(x_k - \gamma_k F(x_k) + \bar{w}_k) - x_{k+\frac{1}{2}} \leq 0.
\]

Substituting (18) in (17), we deduce that
\[
\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \|x_k - x_{k+\frac{1}{2}}\|^2 - \|x_{k+\frac{1}{2}} - x_{k+1}\|^2 + 2\gamma_k^2 L^2\|x_k - x_{k+\frac{1}{2}}\|^2 \\
+ 2\gamma_k^2 \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 + \frac{1}{2}\|x_k + x_{k+\frac{1}{2}}\|^2 - F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) \\
= \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2)\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 \\
- F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) \\
= \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2)\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 \\
- F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}),
\]
by noticing that $\gamma_k = \gamma$. Define $r_\gamma(x) \triangleq \|x - \Pi_X(x - \gamma F(x))\|$ as a residual function. We have
\[
r_\gamma^2(x_k) = \|x_k - \Pi_X(x_k - \gamma F(x_k))\|^2 \\
= \|x_k - x_{k+\frac{1}{2}} + \Pi_X(x_k - \gamma F(x_k) - \gamma \bar{w}_k) - \Pi_X(x_k - \gamma F(x_k))\|^2 \\
\leq 2\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma^2 \|\bar{w}_k\|^2.
\]
It follows that
\[ -\frac{1}{2} \| x_k - x_{k+h} \|^2 \leq -\frac{1}{4} \gamma^2(x_k) + \frac{1}{2} \gamma^2(\bar{w}_k)^2. \] (20)

Using (20) in (19), we obtain
\[
\| x_{k+1} - x^* \|^2 \leq \| x_k - x^* \|^2 - \left( \frac{1}{2} - 2\gamma^2L^2 \right) \| x_k - x_{k+\frac{1}{2}} \|^2 + 2\gamma^2 \| \bar{w}_k - \bar{w}_{k+\frac{1}{2}} \|^2 \\
- F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) - \frac{1}{2} \| x_k - x_{k+\frac{1}{2}} \|^2 \\
\leq \| x_k - x^* \|^2 - \left( \frac{1}{2} - 2\gamma^2L^2 \right) \| x_k - x_{k+\frac{1}{2}} \|^2 + 2\gamma^2 \| \bar{w}_k - \bar{w}_{k+\frac{1}{2}} \|^2 \\
- F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) - \frac{1}{4} \gamma^2(\bar{w}_k)^2 + 1 \\
\leq \| x_k - x^* \|^2 - \left( \frac{1}{2} - 2\gamma^2L^2 \right) \| x_k - x_{k+\frac{1}{2}} \|^2 + \frac{9}{2} \gamma^2 \| \bar{w}_k \|^2 + 4\gamma^2 \| \bar{w}_{k+\frac{1}{2}} \|^2 \\
- F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}} - x^*) + 2\gamma \bar{w}_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) - \frac{1}{4} \gamma^2(\bar{w}_k)^2.
\]

Taking expectations conditioned on \( F_k \) and leveraging \( \gamma \leq \frac{1}{2L} \), we obtain the following bound:
\[
\mathbb{E}[\| x_{k+1} - x^* \|^2 | F_k] \leq \| x_k - x^* \|^2 + \mathbb{E}[\mathbb{E}[4\gamma^2 \| \bar{w}_{k+\frac{1}{2}} \|^2 | F_{k+\frac{1}{2}}] | F_k] + \mathbb{E} \left[ \frac{9}{2} \gamma^2 \| \bar{w}_k \|^2 | F_k \right] \\
- \mathbb{E}[\mathbb{E}[2\gamma \bar{w}_{k+\frac{1}{2}}^T(x_{k+\frac{1}{2}} - x^*) | F_{k+\frac{1}{2}}] | F_k] - \frac{1}{4} \gamma^2(\bar{w}_k)^2 \\
\leq \| x_k - x^* \|^2 + \frac{17}{2} \gamma^2 \| \bar{w}_k \|^2 - \mathbb{E}[\mathbb{E}[2\gamma \bar{w}_{k+\frac{1}{2}}^T(x_{k+\frac{1}{2}} - x^*) | F_{k+\frac{1}{2}}] | F_k] - \frac{1}{4} \gamma^2(\bar{w}_k)^2 \\
= \| x_k - x^* \|^2 + \frac{17}{2} \gamma^2 \| \bar{w}_k \|^2 - \frac{1}{4} \gamma^2(\bar{w}_k)^2. \] (21)

We may now apply Lemma S which allows us to claim that \( \{\| x_k - x^* \|\} \) is convergent and \( \sum_k r_\gamma(x_k)^2 < \infty \) in an a.s. sense. Therefore, in an a.s. sense, we have
\[ \lim_{k \to \infty} r_\gamma(x_k)^2 = 0. \]

This implies that the entire sequence \( \{x_k\} \) converges to a point in \( X^* \) in an a.s. sense. \( \square \)

Next we derive rate statements for the averaged sequence in the mere monotonicity. Unlike in stochastic convex optimization where the function value represents a metric to ascertain progress of the algorithm, a similar metric is not immediately available for variational inequality problems. Instead, the progress of the scheme can be ascertained by using the gap function, defined next.

**Definition 3.1 (Gap function).** Given a nonempty closed set \( X \subseteq \mathbb{R}^n \) and a mapping \( F : \mathbb{R}^n \to \mathbb{R}^n \), then the gap function at \( x \) is denoted by \( G(x) \) and is defined as follows for any \( x \in X \).
\[ G(x) \triangleq \sup_{y \in X} F(y)^T(x - y). \]
The gap function is nonnegative for all \( x \in X \) and is zero if and only if \( x \) is a solution of SVI (cf. [12]). We establish the convergence rate for (v-SSE) by using the gap function.

**Proposition 13.** Consider the (v-SSE) scheme and let \( \{\bar{x}_K\} \) be defined as \( \bar{x}_K = \sum_{k=1}^{K} x_{k+\frac{1}{2}}/K \), where \( 0 < \gamma_k = \gamma \leq 1/(2L) \) for all \( k \geq 0 \) and \( \sum_{k=1}^{\infty} \frac{1}{\gamma_k} < M \). Let Assumptions \([12\, 13\, 14]\) hold. (1). Then we have \( \mathbb{E}[G(\bar{x}_K)] \leq O\left(\frac{1}{K}\right) \) for any \( K \). (2) Suppose \( N_k = \lfloor k^\alpha \rfloor \), for \( \alpha > 1 \). Then the oracle complexity to compute an \( \bar{x}_K \) such that \( \mathbb{E}[G(\bar{x}_K)] \leq \epsilon \) is bounded as follows: \( \sum_{k=1}^{K} N_k \leq O\left(\frac{1}{\epsilon}\right) \).

**Proof.** (1). From (19) and by replacing \( x^* \) by \( y \), we obtain

\[
F(y)^T(x_{k+\frac{1}{2}} - y) \leq \|x_k - y\|^2 - \|x_{k+1} - y\|^2 - (1 - 2\gamma^2 L^2)\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma^2 \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2
+ 2\gamma\bar{w}^T_{k+\frac{1}{2}}(x^* - x_{k+\frac{1}{2}}).
\]

Summing over \( k \), we obtain the following bound:

\[
\sum_{k=1}^{K} F(y)^T(x_{k+\frac{1}{2}} - y) \leq \frac{1}{K} \sum_{k=1}^{K} \|x_k - y\|^2 + \frac{2\gamma^2 \sum_{k=1}^{K} \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2}{K} + \frac{2\gamma \sum_{k=1}^{K} \bar{w}^T_{k+\frac{1}{2}}(x^* - x_{k+\frac{1}{2}})}{K}
\]

or \( F(y)^T(\bar{x}_K - y) \leq \frac{1}{K} \sum_{k=1}^{K} \|x_k - y\|^2 + \frac{2\gamma^2 \sum_{k=1}^{K} \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2}{K} + \frac{2\gamma \sum_{k=1}^{K} \bar{w}^T_{k+\frac{1}{2}}(x^* - x_{k+\frac{1}{2}})}{K}.
\]

By taking supremum over \( y \in X \), we obtain the following inequality:

\[
\sup_{y \in X} F(y)^T(\bar{x}_K - y) \leq \frac{1}{K} \sup_{y \in X} \|x_k - y\|^2 + \frac{2\gamma^2 \sum_{k=1}^{K} \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2}{K} + \frac{2\gamma \sum_{k=1}^{K} \bar{w}^T_{k+\frac{1}{2}}(x^* - x_{k+\frac{1}{2}})}{K}.
\]

\( \Rightarrow \) \( G(\bar{x}_K) \leq \frac{B^2}{K} + \frac{2\gamma \sum_{k=1}^{K} \bar{w}^T_{k+\frac{1}{2}}(x^* - x_{k+\frac{1}{2}})}{K} \).

Taking expectations on both sides, leads to the following inequality.

\[
\mathbb{E}[G(\bar{x}_K)] \leq \frac{B^2}{K} + \frac{2\gamma \sum_{k=1}^{K} \mathbb{E}[\|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2]}{K} + \frac{2\gamma \sum_{k=1}^{K} \mathbb{E}[\bar{w}^T_{k+\frac{1}{2}}(x^* - x_k)]}{K}
\]

\[
\leq \frac{B^2}{K} + \frac{2\gamma^2 \sum_{k=1}^{K} \mathbb{E}[\|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2]}{K} + \frac{2\gamma \sum_{k=1}^{K} \mathbb{E}[\bar{w}^T_{k+\frac{1}{2}}(x^* - x_k)]}{K}
\]

(2). We can use a same proof manner with Proposition [13](2). \( \square \)

**Remark:** While the statements display the similar rates for these three methods, the constants are naturally quite distinct. In particular, we note that the Lipschitz constant appears in the bounds defining the complexity of (SPRG) and lead to a somewhat poorer bound. Yet, as the numerics display, these distinctions are less evident in practice suggesting that the bounds are relatively weak.
4 Incorporating Random Projections in (SPRG) and (SSE)

In this section, we assume that even a single projection onto the feasible set $X$ is challenging. We assume that $X$ is given by an intersection of a collection of closed and convex sets $\{X_i\}_{i \in I}$ where $I$ is a finite set and consider a variants of (SPRG) and (SSE) where the projection onto $X$ is replaced by a projection onto a randomly selected set $X_i$. In Section 4.1, we review our main assumptions and any supporting results and proceed to derive asymptotic and rate guarantees in Sections 4.2 and 5.2 for the random projection variants of (SPRG) and (SSE), respectively.

4.1 Assumptions and Supporting Results

To establish the convergence, we need the following additional assumptions on the projection set $X = \bigcap_{i \in I} X_i$ and random projection process $\Pi_l$. The following assumption is known as linear regularity discussed in [32]. It indicates that this condition is a mild restriction in practice.

**Assumption 14.** There exists a positive scalar $\eta$ such that for any $x \in \mathbb{R}^n$
$$\|x - \Pi_X(x)\|^2 \leq \eta \max_{i \in I} \|x - \Pi_{X_i}(x)\|^2,$$
where $I$ is a finite set of indexes, $I = \{1, \ldots, m\}$.

The following assumption requires that each constraint is sampled with at least some probability and the random samples are nearly independent, which refers to [32].

**Assumption 15.** The random variables $l_k, k = 0, 1, \ldots$, are such that
$$\inf_{k \geq 0} P(l_k = X_i \mid F_k) \geq \frac{\rho_i}{m}, \quad i = 1, \ldots, m,$$
with probability 1, where for $i = 1, \ldots, m$, $\rho_i \in (0, 1)$ is a scalar.

The following lemma is essential to our proofs and it leverages basic properties of projection.

**Lemma 16.** Let $X$ be a closed convex subset of $\mathbb{R}^n$. We have
$$\|y - \Pi_X(y)\|^2 \leq 2\|x - \Pi_X(x)\|^2 + 8\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

**Proof.** Since $y - \Pi_X(y) = (x - \Pi_X(x)) - (x - y) + (\Pi_X(x) - \Pi_X(y))$, we have
$$\|y - \Pi_X(y)\| \leq \|x - \Pi_X(x)\| + \|x - y\| + \|\Pi_X(x) - \Pi_X(y)\| \leq \|x - \Pi_X(x)\| + 2\|x - y\|.$$ 
Thus,
$$\|y - \Pi_X(y)\|^2 \leq 2\|x - \Pi_X(x)\|^2 + 8\|x - y\|^2,$$
where the last inequality leverages $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$. \hfill \Box

The following lemma provides an inequality which is useful in deriving lower bound of $\|x_{k+1} - x^*\|^2$.

**Lemma 17.** Under Assumptions [3, 7] and 4, we have
$$F(x)^T(x - x^*) \geq \alpha \text{dist}(\Pi_X(x), X^*) - C \text{dist}(x, X), \quad \forall x \in \mathbb{R}^n.$$
Proof. We have
\[ F(x)^T(x - x^*) = (F(x) - F(x^*))^T(x - x^*) + F(x^*)^T(\Pi_X(x) - x^*) + F(x^T)(x - \Pi_X(x)). \tag{24} \]
From the monotonicity assumption on \( F \), we have
\[ (F(x) - F(x^*))^T(x - x^*) \geq 0. \tag{25} \]
Since \( x^* \) is a solution, it follows that from the weak sharpness property,
\[ F(x^*)^T(\Pi_X(x) - x^*) \geq o\text{dist}(\Pi_X(x), X^*). \tag{26} \]
Finally, \( F(x^*)^T(\Pi_X(x) - x) \leq \|F(x^*)\|\|x - \Pi_X(x)\| \) and \( \|F(x^*)\| \leq C \) (by Assumption 3),
\[ F(x^*)^T(x - \Pi_X(x)) \geq -\|F(x^*)\|\|x - \Pi_X(x)\| \geq -C\text{dist}(x, X). \tag{27} \]
By substituting (25) – (27) in (24), the result follows. \( \square \)

Lemma 18. Suppose Assumptions 1 and 3 hold. Then for any \( x \in \mathbb{R}^n \),
\[ \|F(x)\|^2 \leq 2L^2\|x - x^*\|^2 + 2C^2. \]
Proof. The result follows by using the triangle inequality \( \|F(x)\| \leq \|F(x - F(x^*))\| + \|F(x^*)\| \). \( \square \)

Lemma 19. Suppose Assumptions 14 and 15 hold. Then for any \( l_k \in I \) and any \( x \in \mathbb{R}^n \),
\[ \mathbb{E}[\|x - \Pi_{l_k}(x)\|^2 | \mathcal{F}_k] \geq \frac{\rho}{mn\eta}\text{dist}^2(x, X), \quad k \geq 0, \]
with probability 1, where \( \rho \triangleq \min_{i \in I}\{\rho_i\} \).
Proof. Following from Assumption 15 we have
\[ \mathbb{E}[\|x - \Pi_{l_k}(x)\|^2 | \mathcal{F}_k] = \sum_{i=1}^{m} P(l_k = i | \mathcal{F}_k)\|x - \Pi_i(x)\|^2 \geq \frac{\rho}{m}\|x - \Pi_j(x)\|^2, \quad \forall j = 1, \ldots, m \]
\[ \Rightarrow \mathbb{E}[\|x - \Pi_{l_k}(x)\|^2 | \mathcal{F}_k] \geq \frac{\rho}{m} \max_j \|x - \Pi_j(x)\|^2 \overset{(*)}{\geq} \frac{\rho}{mn\eta}\text{dist}^2(x, X). \]
\( \square \)

4.2 SPRG with random projections
We begin with an a.s. convergence claim for (r-SPRG).

Proposition 20. Let Assumptions 4, 5, 6, 7 hold. Then any sequence generated by (r-SPRG), where the projections are randomly generated, converges to a solution \( x^* \in X \) in an a.s. sense.
Proof. Define \( y_k = 2x_k - x_{k-1} \) for all \( k \geq 1 \). By Lemma 5(ii) and by noting that \( x_{k+1} = \Pi_X(x_k - \gamma_k F(2x_k - x_{k-1})) \) and \( F(x_k, \omega_k) = F(x_k) + w_k \), we have the following inequality:
\[ \|x_{k+1} - x^*\|^2 \leq \|x_k - \gamma_k F(y_k, \omega_k) - x^*\|^2 - \|x_k - \gamma_k F(y_k, \omega_k) - x_{k+1}\|^2 \]
\[ = \|x_k - x^*\|^2 - \|x_{k+1} - x_k\|^2 - 2\gamma_k (F(y_k) + w_k)^T (x_{k+1} - x^*) \]
\[ = \|x_k - x^*\|^2 - \|x_{k+1} - x_k\|^2 - 2\gamma_k F(y_k)^T (x_{k+1} - x^*) - 2\gamma_k w_k^T (x_{k+1} - x^*). \] (28)

Since
\[ \|y_k - x_{k+1}\|^2 = 2\|x_k - x_{k+1}\|^2 - \|x_{k-1} - x_{k+1}\|^2 + 2\|x_k - x_{k-1}\|^2, \]
We have
\[ \frac{1}{4}\|x_k - x_{k+1}\|^2 = \frac{1}{8}\|y_k - x_{k+1}\|^2 + \frac{1}{8}\|x_{k-1} - x_{k+1}\|^2 - \frac{1}{4}\|x_k - x_{k-1}\|^2. \] (29)

Using (29) in (28), we obtain
\[ \|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \frac{3}{4}\|x_{k+1} - x_k\|^2 - \frac{1}{8}\|y_k - x_{k+1}\|^2 - \frac{1}{8}\|x_{k-1} - x_{k+1}\|^2 \]
\[ + \frac{1}{4}\|x_k - x_{k-1}\|^2 - 2\gamma_k F(y_k)^T (x_{k+1} - x^*) - 2\gamma_k w_k^T (x_{k+1} - x^*) \]
\[ = \|x_k - x^*\|^2 - \frac{3}{4}\|x_{k+1} - x_k\|^2 - \frac{1}{8}\|y_k - x_{k+1}\|^2 - \frac{1}{8}\|x_{k-1} - x_{k+1}\|^2 \]
\[ + \frac{1}{4}\|x_k - x_{k-1}\|^2 - 2\gamma_k F(y_k)^T (y_k - x^*) - 2\gamma_k F(y_k)^T (x_{k+1} - y_k) - 2\gamma_k w_k^T (x_{k+1} - x^*) \]
\[ \leq \|x_k - x^*\|^2 - \frac{3}{4}\|x_{k+1} - x_k\|^2 - \frac{1}{8}\|y_k - x_{k+1}\|^2 - \frac{1}{8}\|x_{k-1} - x_{k+1}\|^2 + \frac{1}{4}\|x_k - x_{k-1}\|^2 \]
\[ - 2\gamma_k \alpha \text{dist} (\Pi_X(y_k), X^*) + 2\gamma_k C \text{dist}(y_k, X) - 2\gamma_k F(y_k)^T (x_{k+1} - y_k) - 2\gamma_k w_k^T (x_{k+1} - x^*), \] (30)

where the last inequality follows from Lemma 17. Since
\[ -2\gamma_k F(y_k)^T (x_{k+1} - y_k) \leq 16\gamma_k^2 \|F(y_k)\|^2 + \frac{1}{16}\|x_{k+1} - y_k\|^2 \] (31)

Using (31) in (30), we have
\[ \|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \frac{3}{4}\|x_{k+1} - x_k\|^2 - \frac{1}{16}\|y_k - x_{k+1}\|^2 - \frac{1}{8}\|x_{k-1} - x_{k+1}\|^2 + \frac{1}{4}\|x_k - x_{k-1}\|^2 \]
\[ - 2\gamma_k \alpha \text{dist} (\Pi_X(y_k), X^*) + 2\gamma_k C \text{dist}(y_k, X) + 16\gamma_k^2 \|F(y_k)\|^2 + 16\gamma_k^2 \|w_k\|^2 - 2\gamma_k w_k^T (y_k - x^*) \]
\[ \leq \|x_k - x^*\|^2 - \frac{3}{4}\|x_{k+1} - x_k\|^2 - \frac{1}{16}\|y_k - x_{k+1}\|^2 - \frac{1}{8}\|x_{k-1} - x_{k+1}\|^2 + \frac{1}{4}\|x_k - x_{k-1}\|^2 \]
\[ - 2\gamma_k \alpha \text{dist} (\Pi_X(y_k), X^*) + 2\gamma_k C \text{dist}(y_k, X) + 32\gamma_k^2 L^2 \|y_k - x^*\|^2 \]
\[ + 32\gamma_k^2 C^2 + 16\gamma_k^2 \|w_k\|^2 - 2\gamma_k w_k^T (y_k - x^*). \] (32)

Since
\[ -2\gamma_k \alpha \text{dist} (\Pi_X(y_k), X^*) \leq -2\gamma_k \alpha \text{dist} (x_k, X^*) + 2\gamma_k \alpha \|x_k - \Pi_X(y_k)\| \]
\[ \leq -2\gamma_k \alpha \text{dist} (x_k, X^*) + 2\gamma_k \alpha \|x_k - y_k\| + 2\gamma_k \alpha \|y_k - \Pi_X(y_k)\| \]
\[ = -2\gamma_k \alpha \text{dist} (x_k, X^*) + 2\gamma_k \alpha \|x_k - y_k\| + 2\gamma_k \alpha \text{dist}(y_k, X), \]
we have
\[ \|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2\gamma_k \alpha \text{dist} (x_k, X^*) - \frac{3}{4}\|x_{k+1} - x_k\|^2 - \frac{1}{16}\|y_k - x_{k+1}\|^2 \]

17
\[ -\frac{1}{8}\|x_{k-1} - x_{k+1}\|^2 + \frac{1}{4}\|x_k - x_{k-1}\|^2 + 2\gamma_k\alpha\|x_k - y_k\| + 2\gamma_k(C + \alpha)\text{dist}(y_k, X) + 64\gamma_k^2L^2\|x_k - x^*\|^2 + 64\gamma_k^2L^2\|x_k - x_{k-1}\|^2 + 32\gamma_k^2C^2 + 16\gamma_k^2\|w_k\|^2 - 2\gamma_kw_k^T(y_k - x^*). \]  

By Lemma 19

\[
\mathbb{E}[\|y_k - x_{k+1}\|^2 \mid \mathcal{F}_k] \geq \mathbb{E}[\|y_k - \Pi_k y_k\|^2 \mid \mathcal{F}_k] \geq \frac{\rho}{\eta}\|y_k\|^2.
\]

Taking expectations conditioned on \(\mathcal{F}_k\) and using (34) in (33), we have

\[ \mathbb{E}[\|x_{k+1} - x^*\|^2 + \frac{3}{4}\|x_{k+1} - x_k\|^2 \mid \mathcal{F}_k] \leq \|x_k - x^*\|^2 - 2\gamma_k\alpha\text{dist}(x_k, X^*) - \frac{1}{16}\mathbb{E}[\|y_k - x_{k+1}\|^2 \mid \mathcal{F}_k] \]

\[ - \frac{1}{8}\mathbb{E}[\|x_{k-1} - x_{k+1}\|^2 \mid \mathcal{F}_k] + \frac{1}{4}\|x_k - x_{k-1}\|^2 + 2\gamma_k\alpha\|x_k - y_k\| + 2\gamma_k(C + \alpha)\text{dist}(y_k, X) + 64\gamma_k^2L^2\|x_k - x^*\|^2 + 64\gamma_k^2L^2\|x_k - x_{k-1}\|^2 + 32\gamma_k^2C^2 + 16\gamma_k^2\|w_k\|^2 \]

\[ \leq \|x_k - x^*\|^2 - 2\gamma_k\alpha\text{dist}(x_k, X^*) - \frac{1}{16}\mathbb{E}[\|y_k - x_{k+1}\|^2 \mid \mathcal{F}_k] \]

\[ + 2\gamma_k(C + \alpha)\text{dist}(y_k, X) + 64\gamma_k^2L^2\|x_k - x^*\|^2 + 64\gamma_k^2L^2\|x_k - x_{k-1}\|^2 + 32\gamma_k^2C^2 + 16\gamma_k^2\|w_k\|^2 \]

\[ \leq (1 + 86\gamma_k^2L^2) \left( \|x_k - x^*\|^2 + \frac{3}{4}\|x_k - x_{k-1}\|^2 \right) - \left( \frac{1}{2}\|x_k - x_{k-1}\| + 2\gamma_k\alpha\|x_k - X^*\| + \frac{\rho}{16m\eta} \left( \text{dist}(y_k, X) - \frac{16m\eta\gamma_k(C + \alpha)}{\rho} \right)^2 \right) \]

\[ \beta_k \]

\[ + \left( \frac{2\gamma_k^2\alpha^2}{\rho} + \frac{16m\eta\gamma_k(C + \alpha)^2}{\rho} \gamma_k^2 + 32\gamma_k^2C^2 + 16\gamma_k^2\|w_k\|^2 \right). \]

In effect, we obtain the following recursion:

\[ \mathbb{E}[y_{k+1} \mid \mathcal{F}_k] \leq (1 - u_k)v_k - \beta_k + \eta_k, \]

where \(v_k = \|x_k - x^*\|^2 + \frac{3}{4}\|x_k - x_{k-1}\|^2\) and \(u_k = 86\gamma_k^2L^2\). Since \(\sum \gamma_k^2 < \infty\), it follows that \(u_k\) and \(\beta_k\) are summable. We may then invoke Lemma 8 and it follows that with probability one, the random sequence \(\{\|x_k - x^*\|^2 + \frac{3}{4}\|x_k - x_{k-1}\|^2\}\) is convergent and \(\sum \frac{1}{2}\|x_k - x_{k-1} - 2\gamma_k\alpha\|^2 + \)
$2\gamma_k \alpha \text{dist} (x_k, X^*) < \infty$ with probability one. We have that $\sum_k \frac{1}{B} \|x_k - x_{k-1} - 2\gamma_k \alpha\|^2 < \infty$ implying that $\|x_k - x_{k-1} - 2\gamma_k \alpha\| \to 0$ in a.s. sense. It follows that $\|y_k - x_{k-1} - 2\gamma_k \alpha\| \to 0$ a.s. Since $\gamma_k \to 0$, it follows that $y_k - x_k \to 0$ in an a.s. sense, which means $x_k - x_{k-1} \to 0$ in an a.s. sense. Thus $\{\|x_k - x^*\|\}$ is convergent in an a.s. sense. We may then conclude by contradiction that $\text{dist}(x_k, X^*) \to 0$ in an a.s. sense. If not, then with finite probability, every subsequence of $\{x_k\}$ satisfies $\text{dist}(x_k, X^*) \to h(\omega) \geq \bar{h} > 0$ implying that $\sum_{k=1}^{\infty} \gamma_k \alpha \text{dist}(x_k, X^*) = \infty$ with finite probability. This contradicts $\sum_k \beta_k < \infty$, implying that $x_k \xrightarrow{k \to \infty} x^*$ in an a.s. sense.

We now provide a rate and oracle complexity statement for this scheme.

**Proposition 21.** Let Assumptions 7 - 10, 14 - 17 hold and let $0 < \gamma_k = \gamma = \frac{\sqrt{B}}{2\sqrt{M_1K}}$, where $K$ is the pre-defined termination number of iterations and $M_1 = \frac{301}{2} L^2 B^2 + 2\alpha^2 + \frac{16m^2(C + \alpha)^2}{\rho} + 32C^2 + 16\nu^2$. Then the following holds for any sequence generated by (r-SPR) in an expected value sense, where $\bar{x}_k = \sum_{k=0}^{K-1} x_k / K$: (1) $\mathbb{E}[\text{dist}(\bar{x}_K, X^*)] \leq \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$; (2) The oracle complexity to compute an $\bar{x}_K$ such that $\mathbb{E}[\text{dist}(\bar{x}_K, X^*)]$ is bounded as follows: $\sum_{k=0}^{K-1} N_k \leq \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$, where $N_k = 1$ for all $k$.

**Proof.** (1) Taking expectations on both sides of (36), we have

$$2\gamma_k \alpha \mathbb{E}[\text{dist}(x_k, X^*)] \leq \mathbb{E}\left[\|x_k - x^*\|^2 + \frac{3}{4} \|x_k - x_{k-1}\|^2\right] - \mathbb{E}\left[\|x_{k+1} - x^*\|^2 + \frac{3}{4} \|x_{k+1} - x_k\|^2\right]$$

$$+ 86\gamma_k^2 L^2 \mathbb{E}\left[\|x_k - x^*\|^2 + \frac{3}{4} \|x_k - x_{k-1}\|^2\right] + 2\gamma_k^2 \alpha^2 + \frac{16m^2(C + \alpha)^2}{\rho} \gamma_k^2 + 32\gamma_k^2 C^2 + 16\gamma_k^2 \nu^2$$

$$\leq \mathbb{E}\left[\|x_k - x^*\|^2 + \frac{3}{4} \|x_k - x_{k-1}\|^2\right] - \mathbb{E}\left[\|x_{k+1} - x^*\|^2 + \frac{3}{4} \|x_{k+1} - x_k\|^2\right] + \gamma_k^2 M_1,$$

where $M_1 = \frac{301}{2} L^2 B^2 + 2\alpha^2 + \frac{16m^2(C + \alpha)^2}{\rho} + 32C^2 + 16\nu^2$. Summing over $k$ from $k = 0$ to $K - 1$, we have

$$2\gamma_k \alpha \sum_{k=0}^{K-1} \mathbb{E}[\text{dist}(x_k, X^*)] \leq \mathbb{E}\left[\|x_0 - x^*\|^2 + \frac{3}{4} \|x_0 - x_1\|^2\right]$$

$$- \mathbb{E}\left[\|x_K - x^*\|^2 + \frac{3}{4} \|x_K - x_{K-1}\|^2\right] + K \gamma^2 M_1$$

$$\leq \mathbb{E}\left[\|x_0 - x^*\|^2 + \frac{3}{4} \|x_0 - x_{1}\|^2\right] + K \gamma^2 M_1. \quad (38)$$

It follows that $2\gamma_k \alpha \mathbb{E}[\text{dist}(\bar{x}_K, X^*)] \leq \frac{7B^2}{8K \gamma^2} + \gamma^2 M_1$. Dividing both sides by $2\gamma_k \alpha$ and optimizing the right-hand side in $\gamma$, we obtain the following when $\gamma^* = \frac{\sqrt{B}}{2\sqrt{M_1K}}$,

$$\mathbb{E}[\text{dist}(\bar{x}_K, X^*)] \leq \frac{7B^2}{8K \gamma^*} + \frac{\gamma^* M_1}{2\alpha} = \frac{\sqrt{7M_1B}}{2\alpha \sqrt{K}} = \mathcal{O}\left(\frac{1}{\sqrt{K}}\right),$$

19
(2). From (1), we know that \( K = O(1/\varepsilon^2) \) and it follows that
\[
\sum_{k=1}^{K} N_k = \sum_{k=1}^{K} 1 = K = O\left(\frac{1}{\varepsilon^2}\right).
\]

The feasibility error arises because the random projection algorithms cannot guarantee \( \{x_k\} \) to be feasible. First we conduct almost-sure convergence analysis on the metric \( \{\text{dist}(x_k, X)\} \) for both randomly generated algorithms and then derive the optimal rate of convergence. To establish the rate of convergence, we need the following lemma.

**Lemma 22.** Let \( \{\delta_k\} \) and \( \{\alpha_k\} \) be sequences of nonnegative scalars such that
\[
\delta_{k+1} \leq (1 - \beta) \delta_k + K \alpha_k^2, \quad \forall k \geq 0,
\]
where \( \beta \in (0, 1) \) and \( K \leq 0 \) are constants. If there exists \( k \geq 0 \) such that \( \alpha_{k+1}^2 \geq (1 - \frac{\beta}{2}) \alpha_k^2 \) for all \( k \geq k \), we have
\[
\delta_k \leq \frac{2N}{\beta} \alpha_k^2 + \delta_0 (1 - \beta)^k + \left( K \sum_{t=0}^{k} \alpha_t^2 \right) (1 - \beta)^{k-k}.
\]

**Proof.** Please refer to [32].

**Proposition 23.** Let Assumptions 1 - 15 hold. Suppose \( \{x_k\} \) is generated by \( (r\text{-SPRG}) \), where the projections are randomly generated. Then \( \mathbb{E}[\text{dist}(\bar{x}_K, X)] \leq O\left(\frac{1}{\sqrt{K}}\right) \) for any \( K > 0 \).

**Proof.** Let \( z_k = x_k - \gamma_k F(2x_k - x_{k-1}, \omega_k) \). We have
\[
\text{dist}^2(x_{k+1}, X) \leq \|x_{k+1} - \Pi_X(z_k)\|^2 = \|\Pi_{l_k}(z_k) - \Pi_X(z_k)\|^2 \leq \|z_k - \Pi_X(z_k)\|^2 - \|\Pi_{l_k}(z_k) - z_k\|^2,
\]
where it follows from Lemma 5. By leveraging \( \|a + b\|^2 \leq \left(1 + \frac{4mn}{\rho}\right) \|a\|^2 + \left(1 + \frac{\rho}{4mn}\right) \|b\|^2 \), we obtain
\[
\|z_k - \Pi_X(z_k)\|^2 \leq \|z_k - \Pi_X(x_k)\|^2 = \|z_k - x_k + x_k - \Pi_X(x_k)\|^2 \\
\leq \left(1 + \frac{4mn}{\rho}\right) \|z_k - x_k\|^2 + \left(1 + \frac{\rho}{4mn}\right) \|x_k - \Pi_X(x_k)\|^2.
\]
Combining (39) and (40), we get
\[
\text{dist}^2(x_{k+1}, X) \leq \left(1 + \frac{4mn}{\rho}\right) \|z_k - x_k\|^2 + \left(1 + \frac{\rho}{4mn}\right) \text{dist}^2(x_k, X) - \|\Pi_{l_k}(z_k) - z_k\|^2.
\]
Following from Lemma 16 and 19, we have
\[
\mathbb{E}[\|z_k - \Pi_{l_k}(z_k)\|^2 | \mathcal{F}_k] \geq \frac{\rho}{mn} d^2(z_k) \geq \frac{\rho}{mn} \left(\frac{1}{2} \text{dist}^2(x_k, X) - 4 \|z_k - x_k\|^2\right)
\]
\[20\]
\begin{equation}
\geq \frac{\rho}{2m\eta} \text{dist}^2(x_k, X) - \frac{4\rho}{m\eta} \|z_k - x_k\|^2 \geq \frac{\rho}{2m\eta} \text{dist}^2(x_k, X) - 4\|z_k - x_k\|^2.
\tag{42}
\end{equation}

Applying (42) to (41), it follows that

\[ \mathbb{E}[\text{dist}^2(x_{k+1}, X) | \mathcal{F}_k] \leq \left( 1 - \frac{\rho}{4m\eta} \right) \text{dist}^2(x_k, X) + \left( 5 + \frac{4m\eta}{\rho} \right) \|z_k - x_k\|^2 \]

\[ \leq \left( 1 - \frac{\rho}{4m\eta} \right) \text{dist}^2(x_k, X) + \left( 5 + \frac{4m\eta}{\rho} \right) (4L^2B^2 + 4C^2 + 2\nu^2) \gamma_k^2. \tag{43} \]

It is clear that \( \gamma_{k+1}^2 \geq \left( 1 - \frac{\rho}{8m\eta} \right) \gamma_k^2 \) when \( k \) is sufficiently large. Leveraging Lemma 22, we have

\[ \mathbb{E}[\text{dist}^2(x_k, X)] \leq \left( \frac{40m\eta}{\rho} + \frac{32m^2\eta^2}{\rho^2} \right) (4L^2B^2 + 4C^2 + 2\nu^2) \gamma_k^2 + d(x_0) \left( 1 - \frac{\rho}{4m\eta} \right)^k \]

\[ + \left( 5 + \frac{4m\eta}{\rho} \right) (4L^2B^2 + 4C^2 + 2\nu^2) \sum_{t=0}^{k} \gamma_t^2 \left( 1 - \frac{\rho}{4m\eta} \right)^{k-t} \]

When \( k \) is sufficiently large, it satisfies that

\[ \mathbb{E}[\text{dist}^2(x_k, X)] \leq \left( \frac{40m\eta}{\rho} + \frac{32m^2\eta^2}{\rho^2} \right) (4L^2B^2 + 4C^2 + 2\nu^2) + U_1 \right) \gamma_k^2, \]

where \( U_1 \) is a large number. It follows that

\[ \mathbb{E}[\text{dist}^2(\bar{x}_K, X)] \leq \frac{\sum_{k=0}^{K-1} \mathbb{E}[\text{dist}^2(x_k, X)]}{K} \leq \mathcal{O} \left( \frac{\sum_{k=0}^{K-1} \gamma_k^2}{K} \right) = \mathcal{O} \left( \frac{1}{K} \right), \tag{44} \]

where we assume \( \sum_{k=0}^{K-1} \gamma_k^2 < M < \infty \). By Jensen’s inequality, we obtain

\[ \mathbb{E}[\text{dist}(\bar{x}_K, X)] \leq \sqrt{\mathbb{E}[\text{dist}^2(\bar{x}_K, X)]} = \mathcal{O} \left( \frac{1}{\sqrt{K}} \right). \]

\[ \square \]

4.3 SSE with random projections

We now proceed to provide an analogous set of statements for the SSE scheme with random projections.

Proposition 24. Let Assumptions 1, 3 – 15 hold and let \( \gamma_k \leq \frac{1}{2\pi} \). Then any sequence generated by (r-SSE), where the projections are randomly generated, converges to a solution \( x^* \in X \) in an a.s. sense.

Proof. By Lemma 5(ii), we have

\[ \|x_{k+1} - x^*\|^2 \leq \|x_k - \gamma_k (F(x_{k+\frac{1}{2}}) + w_{k+\frac{1}{2}}) - x^*\|^2 - \|x_k - \gamma_k (F(x_{k+\frac{1}{2}}) + w_{k+\frac{1}{2}}) - x_{k+1}\|^2 \]

\[ = \|x_k - x^*\|^2 - \|x_k - x_{k+1}\|^2 + 2\gamma_k (F(x_{k+\frac{1}{2}}) + w_{k+\frac{1}{2}})^T (x^* - x_{k+1}). \tag{45} \]
It is clear that

\[ F(x_{k+\frac{1}{2}})T(x_{k+1} - x^*) = F(x_{k+\frac{1}{2}})T(x_{k+1} - x_{k+\frac{1}{2}}) + F(x_{k+\frac{1}{2}})T(x_{k+\frac{1}{2}} - x^*). \] (46)

Using (46) in (45), we obtain

\[ \|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \|x_k - x_{k+1}\|^2 + 2\gamma_k F(x_{k+\frac{1}{2}})T(x_{k+\frac{1}{2}} - x_{k+1}) + 2\gamma_k w_{k+\frac{1}{2}}^T(x^* - x_{k+1}) \]
\[ - 2\gamma_k F(x_{k+\frac{1}{2}})T(x_{k+\frac{1}{2}} - x^*) \]
\[ = \|x_k - x^*\|^2 - \|x_k - x_{k+\frac{1}{2}} + x_{k+\frac{1}{2}} - x_{k+1}\|^2 + 2\gamma_k F(x_{k+\frac{1}{2}})T(x_{k+\frac{1}{2}} - x_{k+1}) \]
\[ + 2\gamma_k w_{k+\frac{1}{2}}^T(x^* - x_{k+1}) - 2\gamma_k F(x_{k+\frac{1}{2}})T(x_{k+\frac{1}{2}} - x^*) \]
\[ = \|x_k - x^*\|^2 - \|x_k - x_{k+\frac{1}{2}}\|^2 - \|x_{k+\frac{1}{2}} - x_{k+1}\|^2 + 2\gamma_k F(x_{k+\frac{1}{2}})T(x_{k+\frac{1}{2}} - x_{k+1}) \]
\[ + 2\gamma_k w_{k+\frac{1}{2}}^T(x^* - x_{k+1}) - 2\gamma_k F(x_{k+\frac{1}{2}})T(x_{k+\frac{1}{2}} - x^*) - 2\gamma_k w_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}})^T(x_k - \gamma_k F(x_{k+\frac{1}{2}}) - x_{k+\frac{1}{2}}) \]
\[ + 2\gamma_k w_{k+\frac{1}{2}}^T(x^* - x_{k+1}) - 2\gamma_k F(x_{k+\frac{1}{2}})T(x_{k+\frac{1}{2}} - x^*). \] (47)

With the similar approach in Proposition 12, we obtain

\[ \|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2)\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|w_k - w_{k+\frac{1}{2}}\|^2 \]
\[ + 2\gamma_k w_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) - 2\gamma_k F(x_{k+\frac{1}{2}})T(x_{k+\frac{1}{2}} - x^*) \]
\[ \leq \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2)\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|w_k - w_{k+\frac{1}{2}}\|^2 \]
\[ + 2\gamma_k w_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) - 2\gamma_k \text{dist} \left( \Pi_X(x_{k+\frac{1}{2}}), X^* \right) + 2\gamma_k Cd(x_{k+\frac{1}{2}}). \] (48)

Invoking weak sharpness property, we have

\[ -2\gamma_k \text{dist} \left( \Pi_X(x_{k+\frac{1}{2}}), X^* \right) \leq -2\gamma_k \text{dist} (x_k, X^*) + 2\gamma_k \alpha \|x_k - x_{k+\frac{1}{2}}\| + 2\gamma_k \alpha d(x_{k+\frac{1}{2}}) \] (49)

and

\[ 2\gamma_k (C + \alpha) d(x_{k+\frac{1}{2}}) \leq 2\gamma_k (C + \alpha) \text{dist} (x_k, X) + 2\gamma_k (C + \alpha) \|x_k - x_{k+\frac{1}{2}}\| \]
\[ \leq 2\gamma_k (C + \alpha) \text{dist} (x_k, X) + 4\gamma_k^2 (C + \alpha)^2 + \frac{1}{4} \|x_k - x_{k+\frac{1}{2}}\|^2; \] (50)

Using (49) and (50) in (48), we obtain

\[ \|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2)\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|w_k - w_{k+\frac{1}{2}}\|^2 \]
\[ + 2\gamma_k w_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) - 2\gamma_k \text{dist} \left( \Pi_X(x_{k+\frac{1}{2}}), X^* \right) + 2\gamma_k Cd(x_{k+\frac{1}{2}}) \]
\[ \leq \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2)\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|w_k - w_{k+\frac{1}{2}}\|^2 \]
\[ + 2\gamma_k w_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}) - 2\gamma_k \text{dist} (x_k, X^*) + 2\gamma_k \alpha \|x_k - x_{k+\frac{1}{2}}\| + 2\gamma_k (C + \alpha) d(x_{k+\frac{1}{2}}) \]
\[ \leq \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2)\|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|w_k - w_{k+\frac{1}{2}}\|^2 + 2\gamma_k w_{k+\frac{1}{2}}^T(x^* - x_{k+\frac{1}{2}}). \]
\[-2\gamma_k \text{dist} \left( x_k, X^* \right) + 2\gamma_k \alpha \| x_k - x_{k+\frac{1}{2}} \| + 2\gamma_k (C + \alpha) \text{dist} \left( x_k, X \right) + 4\gamma_k^2 (C + \alpha)^2 + \frac{1}{4} \| x_k - x_{k+\frac{1}{2}} \|^2 \leq \| x_k - x^* \|^2 - 2\gamma_k \text{dist} \left( x_k, X^* \right) - \left( \frac{5}{8} - 2\gamma_k^2 L^2 \right) \| x_k - x_{k+\frac{1}{2}} \|^2 - \frac{1}{8} \| x_k - x_{k+\frac{1}{2}} - 8\gamma_k \alpha \|^2 \]

\[+ 8\gamma_k^2 \alpha^2 + 4\gamma_k^2 (C + \alpha)^2 + 2\gamma_k (C + \alpha) \text{dist} \left( x_k, X \right) + 2\gamma_k^2 \| w_k - w_k^T \| - w_k \|^2 - 2\gamma_k w_k^T \left( x_{k+\frac{1}{2}} - x^* \right).\]

Taking expectations conditioned on \( \mathcal{F}_k \), we obtain

\[\mathbb{E}[\| x_{k+1} - x^* \|^2 \mid \mathcal{F}_k] \leq \| x_k - x^* \|^2 - 2\gamma_k \text{dist} \left( x_k, X^* \right) - \left( \frac{5}{8} - 2\gamma_k^2 L^2 \right) \mathbb{E}[\| x_k - x_{k+\frac{1}{2}} \|^2 \mid \mathcal{F}_k] \]

\[+ 8\gamma_k^2 \alpha^2 + 4\gamma_k^2 (C + \alpha)^2 + 2\gamma_k (C + \alpha) d(x_k) + 8\gamma_k^2 \nu^2.\] (51)

According to Lemma 19, we have

\[\mathbb{E}[\| x_k - x_{k+\frac{1}{2}} \|^2 \mid \mathcal{F}_k] = \mathbb{E}[\| x_k - \Pi_k (x_k - \gamma_k F(x_k, \omega_k)) \|^2 \mid \mathcal{F}_k] \]

\[\geq \mathbb{E}[\| x_k - \Pi_k (x_k) \|^2 \mid \mathcal{F}_k] \geq \frac{\rho}{m\eta} \text{dist}^2 (x_k, X).\] (52)

where the last inequality follows from Lemma 16. Multiplying (52) by \( \frac{1}{\alpha} \) and using it in (51), we have

\[\mathbb{E}[\| x_{k+1} - x^* \|^2 \mid \mathcal{F}_k] \leq \| x_k - x^* \|^2 - 2\gamma_k \text{dist} \left( x_k, X^* \right) - \left( \frac{3}{4} - 2\gamma_k^2 L^2 \right) \mathbb{E}[\| x_k - x_{k+\frac{1}{2}} \|^2 \mid \mathcal{F}_k] \]

\[+ 8\gamma_k^2 \alpha^2 + 4\gamma_k^2 (C + \alpha)^2 + 2\gamma_k (C + \alpha) d(x_k) - \frac{\rho}{8m\eta} \text{dist}^2 (x_k, X) + 8\gamma_k^2 \nu^2 \]

\[= \| x_k - x^* \|^2 - 2\gamma_k \text{dist} \left( x_k, X^* \right) - \left( \frac{3}{4} - 2\gamma_k^2 L^2 \right) \mathbb{E}[\| x_k - x_{k+\frac{1}{2}} \|^2 \mid \mathcal{F}_k] + 8\gamma_k^2 \alpha^2 \]

\[+ 4\gamma_k^2 (C + \alpha)^2 - \frac{\rho}{8m\eta} \left( \text{dist} (x_k, X) - \frac{8m\eta \gamma_k (C + \alpha)}{\rho} \right)^2 + \frac{8m\eta (C + \alpha)^2}{\rho} \gamma_k^2 + 8\gamma_k^2 \nu^2 \]

\[\leq \| x_k - x^* \|^2 - 2\gamma_k \text{dist} \left( x_k, X^* \right) + 8\gamma_k^2 \alpha^2 + 4\gamma_k^2 (C + \alpha)^2 + \frac{8m\eta (C + \alpha)^2}{\rho} \gamma_k^2 + 8\gamma_k^2 \nu^2.\] (53)

Now we may invoke Lemma 8. It follows that \( \{\| x_k - x^* \|^2\} \) is convergent a.s. and \( \sum 2\gamma_k \text{dist} \left( x_k, X^* \right) < \infty \). It remains to show that dist(\( x_k, X^* \)) \( \frac{k}{K} \rightarrow 0 \) a.s.. We proceed by contradiction and assume that with finite probability, \( \text{dist}(x_k, X^*) \rightarrow h(\omega) > 0 \). Since \( \sum \gamma_k \infty = \infty \), it follows that \( \sum \gamma_k \text{dist}(x_k, X^*) = \infty \) with finite probability. But this contradicts \( \sum 2\gamma_k \text{dist} \left( x_k, X^* \right) < \infty \) a.s.. Therefore, dist(\( x_k, X^* \)) \( \rightarrow 0 \) in an a.s. sense.

**Proposition 25.** Let Assumptions 7, 6, 14, 15 hold and let \( 0 < \gamma_k = \gamma = \frac{B}{\sqrt{M_2 R}} \), where \( K \) is the pre-defined termination number of iterations and \( M_2 = 8\alpha^2 + 4(C + \alpha)^2 + \frac{8m\eta(C + \alpha)^2}{\rho} + 8\nu^2 \). Then the following holds for any sequence generated by \( r \text{-SSE} \) in an expected value sense, where \( \bar{x}_k = \sum_{k=0}^{K-1} x_k/K: (1) \mathbb{E}[\text{dist}(\bar{x}_K, X^*)] \leq O \left( \frac{1}{\sqrt{K}} \right); (2) \text{The oracle complexity to compute an } \bar{x}_K \text{ such that } \mathbb{E}[\text{dist}(\bar{x}_K, X^*)] \text{ is bounded as follows: } \sum_{k=1}^{K} N_k \leq O \left( \frac{1}{\sqrt{K}} \right), \text{ where } N_k = 1 \text{ for all } k. \)

**Proof.** (1). Taking expectations on both sides of (53) and using a similar derivation with the proof
of Proposition 21 we have

\[ 2\gamma\alpha \mathbb{E}[\text{dist} (\bar{x}_K, X^*)] \leq \frac{B^2}{K} + \gamma^2 M_2, \]

where \( M_2 = 8\alpha^2 + 4(C + \alpha)^2 + \frac{8mn(C + \alpha)^2}{\rho^2} + 8\nu^2. \)

Dividing both sides by \( 2\gamma\alpha \) and minimizing the right-hand side in \( \gamma \), we obtain the following at the optimal \( \gamma = \frac{B}{\sqrt{M_2}K}. \)

\[
\mathbb{E}[\text{dist} (\bar{x}_K, X^*)] \leq \frac{B^2}{2K\gamma\alpha} + \frac{\gamma M_2}{2\alpha} = \frac{\sqrt{M_2}B}{\alpha\sqrt{K}} = O\left( \frac{1}{\sqrt{K}} \right). \tag{2}
\]

(2). The result follows using the same avenue as Proposition 21(2). \( \square \)

We conclude with an analysis of the infeasibility sequence.

**Proposition 26.** Let Assumptions 21 - 22 - 15 hold. Let \( \gamma_k \leq \frac{1}{2K}. \) Suppose \( \{x_k\} \) is generated by \((r\text{-SSE}), \) where the projections are randomly generated. Then the feasibility error satisfies

\[ \mathbb{E}[\text{dist}(\bar{x}_k, X)] \leq O\left( \frac{1}{\sqrt{K}} \right). \]

**Proof.** Let \( z_k = x_k - \gamma_k F(x_{k+\frac{1}{2}}, \omega_{k+\frac{1}{2}}). \) We have

\[
\text{dist}^2(x_{k+1}, X) \leq \|x_{k+1} - \Pi_X(x_{k+\frac{1}{2}})\|^2 = \|\Pi_{\mathcal{T}_h}(z_k) - x_{k+\frac{1}{2}} + x_{k+\frac{1}{2}} - \Pi_X(x_{k+\frac{1}{2}})\|^2 \\
\leq \left( 1 + \frac{4mn}{\rho} \right) \|\Pi_{\mathcal{T}_h}(z_k) - x_{k+\frac{1}{2}}\|^2 + \left( 1 + \frac{\rho}{4mn} \right) \|x_{k+\frac{1}{2}} - \Pi_X(x_{k+\frac{1}{2}})\|^2 \\
= \left( 1 + \frac{4mn}{\rho} \right) \|\Pi_{\mathcal{T}_h}(z_k) - \Pi_{\mathcal{T}_h}(x_k)\|^2 + \left( 1 + \frac{\rho}{4mn} \right) \|x_{k+\frac{1}{2}} - \Pi_X(x_{k+\frac{1}{2}})\|^2 \\
\leq \left( 1 + \frac{4mn}{\rho} \right) \|z_k - x_k\|^2 + \left( 1 + \frac{\rho}{4mn} \right) \|x_{k+\frac{1}{2}} - \Pi_X(x_{k+\frac{1}{2}})\|^2, \tag{54}
\]

where we leverage \( \|a + b\|^2 \leq \left( 1 + \frac{4mn}{\rho} \right) \|a\|^2 + \left( 1 + \frac{\rho}{4mn} \right) \|b\|^2. \) We have that

\[ \mathbb{E}[d^2(x_{k+\frac{1}{2}}, \mathcal{F}_k)] \leq \left( 1 - \frac{\rho}{4mn} \right) \text{dist}^2(x_k, X) + \left( 5 + \frac{4mn}{\rho} \right) (4L^2B^2 + 4C^2 + 2\nu^2)\gamma_k^2. \tag{55}\]

Using (55) in (54), we obtain

\[ \mathbb{E}[\text{dist}^2(x_{k+1}, X) | \mathcal{F}_k] \leq \left( 1 - \frac{\rho^2}{16m^2\eta^2} \right) \text{dist}^2(x_k, X) + \left( 8 + \frac{12mn}{\rho} + \frac{5\rho}{4mn} \right) (4L^2B^2 + 4C^2 + 2\nu^2)\gamma_k^2. \tag{56}\]

It is clear that \( \gamma_{k+1}^2 \geq \left( 1 - \frac{\rho^2}{32m^2\eta^2} \right) \gamma_k^2 \) when \( k \) is sufficiently large. Leveraging Lemma 22 we have

\[ \mathbb{E}[\text{dist}^2(x_k, X)] \leq \left( \frac{256m^2\eta^2}{\rho^2} + \frac{384m^3\eta^3}{\rho^3} + \frac{40mn}{\rho} \right) (4L^2B^2 + 4C^2 + 2\nu^2)\gamma_k^2 + d(x_0) \left( 1 - \frac{\rho^2}{16m^2\eta^2} \right)^k \]

24
By employing the same technique used in (44), we have
\[ E \text{errors and elapsed time are shown in Table 4. Table 4 shows the performance after 4000 iterations} \]

Recall that SEG has two projections onto the set, while the other two schemes just require one. We compare their performance under the same number of projections (Fig. 3). Next we change the size and parameters of the original game to ascertain parametric sensitivity. In Table 4 we consider test problems which are a set of 16 problems where the settings and their corresponding empirical errors and elapsed time are shown in Table 4. Table 4 shows the performance after 4000 iterations.
Table 3: Empirical and Theoretical errors under mere monotonicity

|          | SEG     | SPRG    | SSE     |
|----------|---------|---------|---------|
| Empirical| 9.8574e-3 | 9.2702e-3 | 9.1534e-3 |
| Theoretical| 2.259e3   | 3.373e3  | 2.259e3  |

Figure 3: Convergence based on projections under mere monotonicity

and find that while SEG has almost the same empirical error with the others but with significant computational cost. To check the performance of variance reduction, we enlarge the random set for random variable $a_j$ to [40, 60]. Fig. 4 shows comparison of variance reduction schemes with original ones under the same number of iterations. Table 5 shows the results generated from different nodes in the system. The number of iterations used is 4000. We note that all schemes show relatively similar sensitivity to the changes introduces.

**Key findings.** The key findings are that (SPRG) and (SSE) produce empirical errors but do so in approximately 65% of the time utilized by (SEG). Moreover, the presence of variance reduction allows for significant improvement in the empirical rates from the single-sample counterparts (See Table 5).

Table 5: Errors and elapsed time comparison of the schemes with different sizes under the same number of iterations

| Network Size | SEG  | Time  | SSE  | Time  | v-SSE | Time  | SPRG | Time  | v-SPRG | Time  |
|--------------|------|-------|------|-------|-------|-------|------|-------|--------|-------|
| 20           | 1.0e-1| 4.0e3s| 1.1e-1| 1.7e3s| 1.9e-3| 1.5e3s| 1.4e-1| 1.5e3s| 7.5e-3| 1.6e3s|
| 24           | 1.3e-1| 4.3e3s| 1.4e-1| 1.8e3s| 7.7e-3| 2.0e3s| 1.3e-1| 1.8e3s| 7.7e-3| 1.7e3s|
| 28           | 1.8e-1| 4.7e3s| 1.9e-1| 2.0e3s| 7.9e-3| 2.1e3s| 1.9e-1| 2.0e3s| 8.0e-3| 1.7e3s|
| 32           | 2.0e-1| 4.9e3s| 2.1e-1| 2.2e3s| 8.0e-3| 2.4e3s| 2.1e-1| 2.2e3s| 8.2e-3| 1.8e3s|
| 36           | 2.5e-1| 5.4e3s| 2.6e-1| 2.5e3s| 8.7e-3| 2.5e3s| 2.6e-1| 2.5e3s| 8.8e-3| 2.1e3s|
| 40           | 3.4e-1| 3.2e3s| 3.5e-1| 2.3e3s| 9.0e-3| 2.6e3s| 3.5e-1| 2.4e3s| 9.1e-3| 2.2e3s|
Table 4: Errors and elapsed time comparison of the three schemes with different parameters 
under mere monotonicity

| $c_{ij}$ | Seg | Time | SSE | Time | SPRG | Time |
|----------|-----|------|-----|------|------|------|
| 2 0.05  | 2.4e3 | 1.1e-3 | 2.4e3 | 1.1e-3 | 9.1e-3 | 1.6e3 |
| 2 0.05  | 2.4e3 | 1.1e-3 | 2.4e3 | 1.1e-3 | 9.1e-3 | 1.6e3 |
| 2 0.05  | 2.4e3 | 1.1e-3 | 2.4e3 | 1.1e-3 | 9.1e-3 | 1.6e3 |
| 2 0.05  | 2.4e3 | 1.1e-3 | 2.4e3 | 1.1e-3 | 9.1e-3 | 1.6e3 |
| 2 0.05  | 2.4e3 | 1.1e-3 | 2.4e3 | 1.1e-3 | 9.1e-3 | 1.6e3 |

Figure 4: Performance comparison between variance reduction schemes and original ones

5.2 Markov Invariant Distribution Approximation

We test the performance of the random projection schemes on an example from [32] which requires computing a low-dimensional approximation to the invariant distribution of a Markov chain. We denote its transition matrix by $P$ and its stationary distribution as $\pi$. The number of states is assumed to be 1000 and we want to approximate the states in a low-dimensional subspace of $\mathbb{R}^{20}$ with a transformation matrix $\Sigma$. Then we use a projection approach to approximate $\pi = P^T \pi$ as $\Sigma x = \Pi_X(P^T \Sigma x)$, where $X = \{ x \mid \Sigma x \geq 0, e^T \Sigma x = 1 \}$. It has been proved [4, 32] that the projected equation is equivalent to the VI:

$$(x - x^*)^T S x^* \geq 0, \quad \forall x \in \mathbb{R}^{20}, \Sigma x \geq 0, e^T \Sigma x = 1,$$

where $S = \Sigma^T (I - P^T) \Sigma$. We generate the transition matrix $P$ randomly in our experiment. The schemes are under strong monotone as well. Table 7 shows the empirical and theoretical errors of all extragradient-type schemes at the 10000th iteration. Figure 5 illustrates the convergence performance of the extragradient schemes considered.

We record the elapsed time and empirical errors of each scheme with 10 different transition matrices, as shown in Table 7 while the comparison between original stochastic schemes and the
Table 6: Empirical and Theoretical errors on random projections

|                  | r-SEG | r-SPRG | r-SSE |
|------------------|-------|--------|-------|
| Empirical        | 0.0776| 0.0758 | 0.0657|
| Theoretical      | 2.0616| 2.9183 | 2.0616|

Figure 5: Convergence based on projections on random projections

random projection variants are shown in Table 8.

**Key insights.** In random projection variants, the projection onto each random constraint is cheap. Thus, the run-time benefits of (r-SSE) are not obvious when compared with (r-SEG) while (r-SPRG) is still faster than others. This is because the second projection in (r-SSE), while computable in closed form, is almost as expensive as a (cheap) projection.

Table 7: Errors and elapsed time comparison of the three schemes with different transition matrices on random projections

| Matrix No. | r-SEG | Time | r-SEG | Time | r-SEG | Time |
|------------|-------|------|-------|------|-------|------|
| No.1       | 7.7e-2| 1.4e3| 6.5e-2| 1.4e3| 7.5e-2| 0.7e3|
| No.2       | 4.0e-2| 1.3e3| 3.9e-2| 1.4e3| 4.0e-2| 0.7e3|
| No.3       | 1.8e-2| 1.3e3| 1.7e-2| 1.4e3| 1.8e-2| 0.7e3|
| No.4       | 5.2e-2| 1.4e3| 4.8e-2| 1.4e3| 5.1e-2| 0.7e3|
| No.5       | 4.7e-2| 1.3e3| 4.4e-2| 1.4e3| 4.6e-2| 0.7e3|
| No.6       | 5.9e-2| 1.3e3| 5.5e-2| 1.4e3| 5.8e-2| 0.7e3|
| No.7       | 2.7e-2| 1.4e3| 2.6e-2| 1.4e3| 2.7e-2| 0.7e3|
| No.8       | 5.8e-2| 1.3e3| 5.3e-2| 1.4e3| 5.7e-2| 0.7e3|
| No.9       | 2.6e-2| 1.4e3| 2.3e-2| 1.4e3| 2.5e-2| 0.7e3|
| No.10      | 3.3e-2| 1.4e3| 3.1e-2| 1.4e3| 3.2e-2| 0.7e3|

Table 8: Errors and elapsed time comparison between the three schemes with random projections and original ones

|        | SEG  | r-SEG | SSE  | r-SSE | SRPG | r-SRPG |
|--------|------|-------|------|-------|------|--------|
| Error  | 4.3e-3| 7.1e-2| 3.7e-3| 6.9e-2| 4.2e-3| 7.9e-2|
| Time   | 2.8e4 | 1.4e3 | 1.6e4 | 1.4e3 | 1.5e4 | 0.7e3 |
6 Concluding remarks

Extragradient schemes and their sampling-based counterparts represent a key cornerstone of solving monotone deterministic and stochastic variational inequality problems. Yet, the per-iteration complexity of such schemes is twice as high as their single projection counterparts. We consider two avenues in which the two projections are replaced by exactly one projection (a projected reflected scheme) or a single projection onto the set and another onto a halfspace, the second of which is computable in closed form (a subgradient extragradient scheme). In both instances, we derive a.s. convergence statements and rate statements under variance reduction. Notably, the sequences achieve a non-asymptotic rate of $O(1/K)$, matching its deterministic counterpart. Furthermore, when this set is itself challenging to project onto, we develop a random projection variant for each scheme. Again, a.s. convergence and rate statements are provided. Empirical behavior of both schemes show significant benefits in terms of per-iteration complexity compared to extragradient counterparts.

References

[1] A. S. Antipin, Method of convex programming using a symmetric modification of lagrange function, Matekon, 14 (1978), pp. 23–38.

[2] H. H. Bauschke and P. L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces, Springer Science & Business Media, 2011.

[3] D. P. Bertsekas, Incremental proximal methods for large scale convex optimization, Mathematical programming, 129 (2011), pp. 163–195.

[4] ———, Temporal difference methods for general projected equations, IEEE Transactions on Automatic Control, 56 (2011), pp. 2128–2139.

[5] J. R Birge and F. Louveaux, Introduction to stochastic programming, Springer Science & Business Media, 2011.

[6] Yair Censor, Aviv Gibali, and Simeon Reich, The subgradient extragradient method for solving variational inequalities in hilbert space, Journal of Optimization Theory and Applications, 148 (2011), pp. 318–335.

[7] X. Chen, R. J-B Wets, and Y. Zhang, Stochastic variational inequalities: residual minimization smoothing sample average approximations, SIAM Journal on Optimization, 22 (2012), pp. 649–673.

[8] Y. Chen, G. Lan, and Y. Ouyang, Accelerated schemes for a class of variational inequalities, Mathematical Programming, (2014), pp. 1–37.

[9] ———, Accelerated schemes for a class of variational inequalities, Mathematical Programming, 165 (2017), pp. 113–149.

[10] S. Cui and U. V. Shanbhag, On the analysis of reflected gradient and splitting methods for monotone stochastic variational inequality problems, in 55th IEEE Conference on Decision and Control, CDC 2016, Las Vegas, NV, USA, December 12-14, 2016, IEEE, 2016, pp. 4510–4515.

[11] C. D Dang and G. Lan, On the convergence properties of non-euclidean extragradient methods for variational inequalities with generalized monotone operators, Computational Optimization and Applications, 60 (2015), pp. 277–310.

[12] F. Facchinei and J. S. Pang, Finite-dimensional variational inequalities and complementarity problems, Springer Science & Business Media, 2007.

[13] AN Iusem, A. Jofré, R. I Oliveira, and P. Thompson, Extragradient method with variance reduction for stochastic variational inequalities, SIAM Journal on Optimization, 27 (2017), pp. 686–724.

[14] A. N Iusem, A. Jofré, and P. Thompson, Incremental constraint projection methods for monotone stochastic variational inequalities, Mathematics of Operations Research, (2018).

[15] H. Jiang and H. Xu, Stochastic approximation approaches to the stochastic variational inequality problem, IEEE Transactions on Automatic Control, 53 (2008), pp. 1462–1475.
A. Juditsky, A. Nemirovski, and C. Tauvel, *Solving variational inequalities with stochastic mirror-prox algorithm*, Stochastic Systems, 1 (2011), pp. 17–58.

A. Kannan and U. V Shanbhag, *The pseudomonotone stochastic variational inequality problem: Analytical statements and stochastic extragradient schemes*, in American Control Conference (ACC), 2014, IEEE, 2014, pp. 2930–2935.

A. Kannan, U. V Shanbhag, and H. M Kim, *Addressing supply-side risk in uncertain power markets: stochastic Nash models, scalable algorithms and error analysis*, Optimization Methods and Software, 28 (2013), pp. 1095–1138.

GM Korpelevich, *The extragradient method for finding saddle points and other problems*, Matecon, 12 (1976), pp. 747–756.

J. Koshal, A. Nedić, and U. V Shanbhag, *Regularized iterative stochastic approximation methods for stochastic variational inequality problems*, Automatic Control, IEEE Transactions on, 58 (2013), pp. 594–609.

A. Kannan, U. V Shanbhag, and H. M Kim, *Addressing supply-side risk in uncertain power markets: stochastic Nash models, scalable algorithms and error analysis*, Optimization Methods and Software, 28 (2013), pp. 1095–1138.

GM Korpelevich, *The extragradient method for finding saddle points and other problems*, Matecon, 12 (1976), pp. 747–756.

A. Kannan and U. V Shanbhag, *The pseudomonotone stochastic variational inequality problem: Analytical statements and stochastic extragradient schemes*, in American Control Conference (ACC), 2014, IEEE, 2014, pp. 2930–2935.

A. Kannan, U. V Shanbhag, and H. M Kim, *Addressing supply-side risk in uncertain power markets: stochastic Nash models, scalable algorithms and error analysis*, Optimization Methods and Software, 28 (2013), pp. 1095–1138.

GM Korpelevich, *The extragradient method for finding saddle points and other problems*, Matecon, 12 (1976), pp. 747–756.

A. Kannan and U. V Shanbhag, *The pseudomonotone stochastic variational inequality problem: Analytical statements and stochastic extragradient schemes*, in American Control Conference (ACC), 2014, IEEE, 2014, pp. 2930–2935.