On the analysis of variance-reduced and randomized projection variants of single projection schemes for monotone stochastic variational inequality problems

Shisheng Cui and Uday V. Shanbhag*

December 22, 2020

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 onto a convex set and require 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 onto X, and a significantly cheaper projection (onto a halfspace) computable in closed form. Under a variance-reduced framework reliant on a sample-average of the map based on an increasing batch-size, we prove almost sure (a.s.) convergence of the iterates to a random point in the solution set for both schemes. Additionally, both schemes display a non-asymptotic rate of $\mathcal{O}(1/K)$ in terms of the gap function where K denotes the number of iterations; notably, both rates match those obtained in deterministic regimes. To address feasibility sets given by the intersection of a large number of convex constraints, we adapt both of the aforementioned schemes to a random projection framework. We then show that the random projection analogs of both schemes also display a.s. convergence under a weak-sharpness requirement; furthermore, without imposing the weak-sharpness requirement, both schemes are characterized by a provable rate of $\mathcal{O}(1/\sqrt{K})$ in terms of the gap function of the projection of the averaged sequence onto X as well as the infeasibility of this sequence. Preliminary numerics support theoretical findings and the schemes outperform standard extragradient schemes 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 VI(X, F) requires finding a point $x^* \in X$ such that

$$F(x^*)^T(x - x^*) \ge 0, \quad \forall x \in X.$$
 (VI(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 [9].

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 $(\Omega, \mathcal{F}, \mathbb{P})$. In short, we are interested in a vector $x^* \in X$ such that

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

where $\mathbb{E}[F(x,\omega)] = \left(\mathbb{E}[F_i(x,\omega)]\right)_{i=1}^K$. The variational inequality problem is an immensely relevant problem that finds application in engineering, economics, and applied sciences (cf. [8, 13, 17, 37]). 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, 38, 42]). A counterpart to SAA schemes is the stochastic approximation (SA) methods where at each iteration, a sample of the stochastic mapping is used (cf. [21, 24, 35]). 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

Given an $x_0 \in X$, the projection-based scheme (PG) generates a sequence $\{x_k\}$, where

$$x_{k+1} := \Pi_X(x_k - \gamma F(x_k)),$$
 (PG)

 $\Pi_X(y)$ denotes the projection of y onto X and γ denotes a suitably small steplength. This method generally requires a strong monotonicity assumption on F to ensure convergence. An extension referred to as the extragradient scheme, suggested by Antipin [1] and Korpelevich [26], required that F be merely monotone and Lipschitz continuous over the set X. However, this scheme requires two projection steps, as captured by (EG).

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

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

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 [28], 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 [28]. Malitsky [28] derived the rate of convergence of the sequence under a strong monotonicity assumption of the map. 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 $C_k \triangleq \{y \in \mathbb{R}^n \mid (x_k - \gamma_k F(x_k) - x_{k+\frac{1}{2}})^T (y - x_{k+\frac{1}{2}}) \leq 0\}$. In (SE), the two projections are replaced by a projection onto the set and a second projection onto a halfspace, the latter of which is computable in closed form. However, no rate of convergence has been provided in their analysis. A third scheme that employs a single projection to contend with merely monotone maps is the iterative Tikhonov regularization (ITR) scheme, a regularized variant of (PG) in which x_{k+1} is updated as per

$$x_{k+1} := \Pi_X(x_k - \gamma_k(F(x_k) + \epsilon_k x_k)), \tag{ITR}$$

where the steplength sequence $\{\gamma_k\}$ and the regularization sequence $\{\epsilon_k\}$ are suitably chosen positive diminishing sequences [22, 25, 43].

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 [36]. A simple stochastic extension of the standard projection scheme for VI(X, F) leads to a stochastic approximation scheme [36]:

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

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)

Jiang and Xu [19] appear amongst the first who applied SA methods to solve stochastic variational inequality problems. An extension of ITR to address merely monotone stochastic VIs was presented by Koshal, Nedić, and Shanbhag [27]. A regularized smoothing SA method to address nonsmooth stochastic Nash equilibrium problems that lead to stochastic VIs with possibly non-Lipschitzian and merely monotone mappings was proposed in [44]. There has also been a development of prox-based generalization of SA methods were developed (cf. [32, 33, 45, 46, 47]) for solving smooth and nonsmooth stochastic convex optimization problems and variational inequality problems. There has also been an effort to develop block-based schemes for Cartesian stochastic variational inequality problems [14]. Fig. 1 illustrates the (SEG) scheme. Extragradient-based schemes (and their stochastic mirror-prox counterparts) represent amongst the simplest of the non-regularized schemes for monotone SVIs (cf. [10, 21]). However, each iteration requires **two** projection steps, rather than one (as in (SPG)). 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

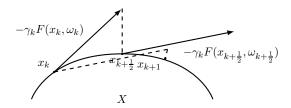


Figure 1: Stochastic extragradient scheme (SEG)

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).

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

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

$$\begin{aligned} x_{k+\frac{1}{2}} &\coloneqq \Pi_X(x_k - \gamma_k F(x_k, \omega_k)), \\ x_{k+1} &\coloneqq \Pi_{C_k}(x_k - \gamma_k F(x_{k+\frac{1}{2}}, \omega_{k+\frac{1}{2}})), \end{aligned} \tag{SSE}$$

where $C_k \triangleq \{y \in \mathbb{R}^n \mid (x_k - \gamma_k F(x_k, \omega_k) - x_{k+\frac{1}{2}})^T (y - x_{k+\frac{1}{2}}) \leq 0\}$. Clearly, the second projection is a simple optimization problem solvable in closed form. 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.

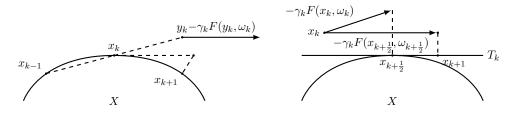


Figure 2: Left: (SPRG); Right: (SSE)

1.3 Incorporating variance reduction and random projections.

To reduce the overall 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 expected map: (i) **Variable sample-size (SPRG)**:

$$x_{k+1} \coloneqq \Pi_X \left(x_k - \gamma_k \frac{\sum_{j=1}^{N_k} F(2x_k - x_{k-1}, \omega_{j,k})}{N_k} \right) , \qquad (\mathbf{v}\text{-}\mathbf{SPRG})$$

and (ii) Variable sample-size (SSE).

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

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

Table 1: A review of stochastic approximation schemes for SVIs

Ref.	Applicability	Compact	Avg.	Metric	Rate	A.s.	# proj.
[19]	Strongly monotone, Lipschitz	N	N	Iterates	-	Y	1
[27]	Monotone, Lipschitz	N	N	Iterates	-	Y	1
[44]	Monotone, non-Lip.	N	N	Iterates	-	Y	1
[21]	Monotone, non-Lip.	Y	Y	Gap fn.	$\mathcal{O}(1/\sqrt{K})$	N	1
[15]	Strongly monotone, Lip.	N	N	Iterates	$\mathcal{O}(1/K)$	N	1
[45, 47]	Monotone, non-Lip.	Y	Y	Gap fn.	$\mathcal{O}(1/\sqrt{K})$	Y	1
[23]	Strongly pseudo/monotone+weak-sharp	Y	N	MSE	$\mathcal{O}(1/K)$	Y	2
[39]	Strongly monotone, Lip., random proj.	N	N	Iterates	$\mathcal{O}(1/\sqrt{K})$	Y	1
[16]	Pseudomonotone, Lip., var. reduction	N	N	Iterates	$\mathcal{O}(1/K)$	Y	2
[17]	Monotone+weak-sharp, Lip., random proj.	N	Y	Dist. fn.	$\mathcal{O}(1/\sqrt{K})$	Y	2
[17]	Monotone, non-Lip., random proj.	Y	Y	Gap fn.	$\mathcal{O}(K^{\delta} \ln K / \sqrt{K})$ where $\delta > 0$	Y	1
v-SPRG	Monotone+weak-sharp, Lip., var. reduction Noise: $\mathbb{E}[\ w(x)\ ^2 \mid x] \leq \nu_1^2 \ x\ ^2 + \nu_2^2$ a.s.	Y	Y	Gap fn.	$\mathcal{O}(1/K)$	Y	1
v-SSE	Monotone, Lip., var. reduction Noise: $\mathbb{E}[\ w(x)\ ^2 \mid x] \leq \nu_1^2 \ x\ ^2 + \nu_2^2$ a.s.	Y	Y	Gap fn.	$\mathcal{O}(1/K)$	Y	1
r-SPRG	Monotone+weak-sharp, Lip., random proj. Noise: $\mathbb{E}[\ w(x)\ ^2 \mid x] \leq \nu_1^2 \ x\ ^2 + \nu_2^2$ a.s.	Y	Y	Gap fn.	$\mathcal{O}(1/\sqrt{K})$	Y	1
r-SSE	Monotone+weak-sharp, Lip., random proj. Noise: $\mathbb{E}[\ w(x)\ ^2 \mid x] \leq \nu_1^2 \ x\ ^2 + \nu_2^2$ a.s.	Y	Y	Gap fn.	$\mathcal{O}(1/\sqrt{K})$	Y	1

where
$$C_k \triangleq \left\{ y \in \mathbb{R}^n \mid \left(x_k - \gamma_k \frac{\sum_{j=1}^{N_k} F(x_k, \omega_{j,k})}{N_k} - x_{k+\frac{1}{2}} \right)^T (y - x_{k+\frac{1}{2}}) \le 0 \right\}.$$

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 [39], 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 [30, 31] considered random projection algorithms for convex optimization problems with similarly defined sets and related schemes were subsequently considered for nonsmooth convex regimes [3, 40, 41]. Wang and Bertsekas [39] applied this avenue to strongly monotone stochastic variational inequality problems by extending (SPG) to allow for projecting on a subset of constraints via random projection technique while Iusem, Jofré, and Thompson [17] extended this framework by incorporating iterative regularization. We consider analogous generalizations to (SPRG) and (SSE):

(i) Random projections SPRG schemes (r-SPRG).

$$x_{k+1} := \Pi_{l_k}(x_k - \gamma_k F(2x_k - x_{k-1}, \omega_k)), \qquad (\mathbf{r}\text{-}\mathbf{SPRG})$$

where Π_{l_k} is defined as projection onto a random subset X_{l_k} and

(ii) Random projections SSE schemes (r-SSE).

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

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

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

1.4 Jutification and relation to other variance-reduced schemes

- (i) Terminology and applicability. The term "variance-reduced" reflects the usage of increasing accurate approximations of the expectation-valued map, as opposed to noisy sampled variants that are used in single sample schemes. The resulting schemes are often referred to as mini-batch SA schemes and often achieve deterministic rates of convergence. Schemes such as SVRG [20] and SAGA [11] also achieve deterministic rates of convergence but are customized for finite sum problems unlike mini-batch schemes that can process expectations over general probability spaces. Unlike in mini-batch schemes where increasing batch-sizes are employed, in schemes such as SVRG, the entire set of samples is periodically employed for computing a step.
- (ii) Weaker assumptions and stronger statements. The proposed variance-reduced framework has several crucial benefits that cannot be reaped in the single-sample regime: (i) Under suitable assumptions, both (v-SPRG) and (v-SSE) achieve optimal deterministic rates in terms of major iterations (projection steps) while achieving near-optimal sample complexity, i.e. $\mathcal{O}(1/\epsilon^{2+\delta})$. (ii) In addition, both sets of schemes are equipped with a.s. convergence guarantees, statements which are seldom obtained for single-sample extragradient schemes (to the best of our knowledge).
- (iii) Sampling requirements. Naturally, variance-reduced schemes can generally be employed only when sampling is relatively cheap compared to the main computational step (such as computing a projection or a prox.) In terms of overall sample-complexity, the proposed schemes are near optimal. As k becomes large, one might question how one might contend with N_k tending to $+\infty$. This issue does not arise since most schemes of this form are meant to provide ϵ -approximations. For instance, if $\epsilon = 1\mathrm{e}-3$, then such a scheme requires approximately $\mathcal{O}(1\mathrm{e}3)$ steps. Since $N_k \approx \lceil k^a \rceil$ and a > 1, we require approximately $(\mathcal{O}(1\mathrm{e}3))^a$ samples. In a setting where multi-core architecture is ubiquitous, such requirements are not terribly onerous particularly since computational costs have been reduced from $\mathcal{O}(1\mathrm{e}6)$ (single-sample) to $\mathcal{O}(1\mathrm{e}3)$. It is worth noting that competing schemes such as SVRG would require taking the full batch-size intermittently and finite-sum problems routingely have 1e9 or more samples.

1.5 Contributions

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

(i) In Section 3, we prove that in settings where the maps are monotone and Lipschitz continuous, the iterates produced by both variance reduced variants of (v-SPRG) and v-(SSE) converge almost surely (a.s.) to a solution, where (v-SPRG) requires an additional weak sharpness requirement. However, without a weak-sharpness requirement, the gap function for an averaged sequence for both schemes diminishes at the rate of $\mathcal{O}(1/K)$. We emphasize that our findings for (v-SPRG) match the best known deterministic rate of convergence while we weaken the assumption for the convergence of sequences generated by (v-SSE) from strong monotonicity to mere monotonicity. To the best of our knowledge, no convergence rate for a deterministic version of (SSE) is available

Table 2: (SRP)	G) and (SSE) schemes comparison,	, $\mathbb{E}[\ w(x)\ ^2$	$ x \le \nu_1^2 x ^2 + \nu_2^2$
----------------	-------------	-----------------------	---------------------------	-------------------------------------

	Variance-reduc	ed schemes	Random projection		
	Assump.	Result	Assump.	Result	
(SPRG)	(v-SPRG): mono.+Lip., weak-sharpness	$ x_k - x^* \xrightarrow{k \to \infty} 0$	(r-SPRG): mono.+Lip., weak-sharpness	$ x_k - x^* \xrightarrow{k \to \infty} 0$	
	(v-SPRG): mono.+Lip. +compactness	$\mathbb{E}[(G(\bar{x}_K))] \le \mathcal{O}\left(\frac{1}{K}\right)$	(r-SPRG): mono.+Lip. +compactness	$\mathbb{E}[(G(\Pi_X(\bar{x}_K)))] \le \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$ $\mathbb{E}[\operatorname{dist}(\bar{x}_K, X)] \le \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$	
(SSE)	(v-SSE): mono.+Lip.	$ x_k - x^* \xrightarrow{k \to \infty} 0$	(r-SSE): mono.+Lip., weak-sharpness	$ x_k - x^* \xrightarrow{k \to \infty} 0$	
	(v-SSE): mono.+Lip. +compactness	$\mathbb{E}[(G(\bar{x}_K))] \le \mathcal{O}\left(\frac{1}{K}\right)$	(r-SSE): mono.+Lip. +compactness	$\mathbb{E}[(G(\Pi_X(\bar{x}_K)))] \le \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$ $\mathbb{E}[\operatorname{dist}(\bar{x}_K, X)] \le \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$	

in the literature.

- (ii) In Section 4, under a weak-sharpness requirement, the sequences produced by random projection variants (r-SPRG) and (r-SSE) are shown to converge a.s. to the solution set of the original problem. Additionally, without weak sharpness, the gap function of the projection of the averaged sequence on X as well as the infeasibility of the sequence with respect to the feasible set X diminish at the rate of $\mathcal{O}(1/\sqrt{K})$.
- (iii) In Section 5, 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 suitable Lipschitzian and monotonicity assumptions on the map F which will be valid through the remainder of this paper.

Assumption 1 (Monotone and Lipschitz maps). The mapping F is L-Lipschitz continuous and monotone on \mathbb{R}^n , i.e. $\forall x, y \in \mathbb{R}^n$, $||F(x) - F(y)|| \le L||x - y||$ and $(F(x) - F(y))^T(x - y) \ge 0$. \square

Since F is a monotone map, VI(X, F) may have multiple solutions. We assume that the set of solutions of VI(X, F), denoted by X^* , is compact and nonempty.

Assumption 2 (Compactness of X^* and Boundedness of F). The set X^* is compact and nonempty where X^* denotes the set of solutions of VI(X,F), i.e. $X^* \triangleq \{x^* \mid x^* \text{ solves } VI(X,F)\}$. There exists a constant C > 0 such that $||F(x^*)|| \leq C$ for all $x^* \in X^*$.

A sufficiency condition for the boundedness of X^* (Assumption 2) is a suitable coercivity property of F over the set X [13, Prop. 2.2.7]. This then allows for claiming the boundedness of F over X^* . For proving almost sure convergence of the iterates, we often impose a weak-sharpness requirement on VI(X, F), which requires utilizing the distance between a point x and a set X, denoted by dist(x, X) and defined as $dist(x, X) \triangleq \min_{y \in X} ||x - y||$.

Assumption 3 (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 $x \in X$, $(x - x^*)^T F(x^*) \ge \alpha \text{dist}(x, X^*)$.

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 $w_k\triangleq F(x_k,\omega_k)-F(x_k)$, $\bar{w}_k\triangleq \frac{\sum_{j=1}^{N_k}F(x_k,\omega_{j,k})}{N_k}-F(x_k)$, $w_{k+1/2}\triangleq F(x_{k+1/2},\omega_{k+1/2})-F(x_{k+1/2})$ and $\bar{w}_{k+1/2}\triangleq \frac{\sum_{j=1}^{N_k}F(x_{k+1/2},\omega_{j,k})}{N_k}-F(x_{k+1/2})$, 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.,

$$\mathcal{F}_k \triangleq \left\{x_0, \{F(x_0, \omega_{j,0})\}_{j=1}^{N_0}, \{F(x_{1/2}, \omega_{j,1/2})\}_{j=1}^{N_0}, \cdots \{F(x_{k-1}, \omega_{j,k-1})\}_{j=1}^{N_{k-1}}, \{F(x_{k-1/2}, \omega_{j,k-1/2})\}_{j=1}^{N_{k-1}}\right\}$$

and $\mathcal{F}_{k+\frac{1}{2}} \triangleq \mathcal{F}_k \cup \{F(x_k, \omega_{j,k})\}_{j=1}^{N_k}$. In settings where the set X may be unbounded, the assumption that the conditional second moment w_k is uniformly bounded a.s. is often a stringent requirement. Instead, we impose a state-dependent assumption on w_k .

Assumption 4 (State-dependent bound on noise). At iteration k, the following hold in an a.s. sense: (i) The conditional means $\mathbb{E}[w_k \mid \mathcal{F}_k]$ and $\mathbb{E}[w_{k+\frac{1}{2}} \mid \mathcal{F}_{k+\frac{1}{2}}]$ are zero for all k in an a.s. sense; (ii) The conditional second moments are bounded in an a.s. sense as follows. $\mathbb{E}[\|w_k\|^2 \mid \mathcal{F}_k] \leq \nu_1^2 \|x_k\|^2 + \nu_2^2$ and $\mathbb{E}[\|w_{k+\frac{1}{2}}\|^2 \mid \mathcal{F}_{k+\frac{1}{2}}] \leq \nu_1^2 \|x_{k+\frac{1}{2}}\|^2 + \nu_2^2$ for all k in an a.s. sense.

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

Lemma 1. Let X be a 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) $(\Pi_X(x) - x)^T (y - \Pi_X(x)) \ge 0$; and (ii) $\|\Pi_X(x) - y\|^2 \le \|x - y\|^2 - \|x - \Pi_X(x)\|^2$.

The Robbins-Siegmund (super-martingale convergence) lemma and its variant are also employed in our analysis (see [34]).

Lemma 2. Let v_k , u_k , δ_k , ψ_k be nonnegative random variables adapted to σ -algebra \mathcal{F}_k , and let the following relations hold almost surely.

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

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

Lemma 3. Let v_k be nonnegative random variables adapted to σ -algebra \mathcal{F}_k where $\mathbb{E}[v_0] < \infty$. Suppose

$$\mathbb{E}[v_{k+1} \mid \mathcal{F}_k] \le (1 - \alpha_k)v_k + \beta_k, \text{ a.s. for all } k \ge 0.$$

In addition, suppose $0 \le \alpha_k \le 1$ and $\beta_k \ge 0$ for $k \ge 0$, $\sum_{k=0}^{\infty} \alpha_k = \infty$, $\sum_{k=0}^{\infty} \beta_k < \infty$, and $\frac{\beta_k}{\alpha_k} \to 0$ as $k \to \infty$. Then $v_k \to 0$ as $k \to \infty$ in an a.s. sense. Furthermore, $\mathbb{E}[v_k] \to 0$ as $k \to \infty$.

We need the following Lemma to prove the a.s. convergence of (v-SPRG).

Lemma 4. Suppose the mapping F is monotone on \mathbb{R}^n and the solution set of VI(X, F) is given by X^* . Then for any $x^*, z^* \in X^*$, we have that

$$F(x^*)^T(z^* - x^*) = F(z^*)^T(x^* - z^*) = 0.$$

Proof. Suppose a limit point of a subsequence of $\{x_k\}$ is given by $z^* \in X^*$. By the definition of X^* , we have for any $x^* \in X^*$ that

$$F(x^*)^T(z^* - x^*) \ge 0, (1)$$

$$F(z^*)^T(x^* - z^*) \ge 0. (2)$$

Combining these two inequalities, we obtain

$$(F(x^*) - F(z^*))^T (x^* - z^*) \le 0.$$

Since the mapping F is monotone, we also have

$$(F(x^*) - F(z^*))^T (x^* - z^*) \ge 0.$$

It follows that $F(z^*)^T(x^*-z^*) = F(x^*)^T(x^*-z^*)$, which by invoking (2) implies that $F(x^*)^T(x^*-z^*) \ge 0$. However, by recalling (1), we have that $F(x^*)^T(z^*-x^*) = 0$. Consequently, $F(z^*)^T(x^*-z^*) = F(x^*)^T(z^*-x^*) = 0$. Thus, the conclusion follows.

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

In this section, we analyze the convergence properties of (v-SPRG) and (v-SSE) in Sections 3.1 and 3.2, respectively.

3.1 Stochastic Projected Reflected Gradient Schemes

In this subsection, we prove the a.s. convergence of the iterates produced by (**v-SPRG**) when F is a Lipschitz continuous and monotone map on \mathbb{R}^n under a weak-sharpness requirement. We then relax the weak-sharpness assumption in deriving a rate statement in terms of the gap function for the averaged sequence. We begin with a lemma that relates the error in consecutive iterates.

Lemma 5. Consider a sequence generated by (v-SPRG). Suppose Assumption 1 holds and $0 < \gamma_k = \gamma \le \frac{1}{8\tilde{L}}$ for all k where $\tilde{L}^2 \triangleq (L^2 + \frac{10\nu_1^2}{N_0})$. Then for any $x_0 \in X$ and any $x^* \in X^*$, the following holds for all $k \ge 0$.

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

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

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

Proof. Define $y_k \triangleq 2x_k - x_{k-1}$ for all $k \geq 1$ and $\bar{F}(y_k) \triangleq \frac{\sum_{j=1}^{N_k} F(y_k, \omega_{k,j})}{N_k}$. We reuse the notation of \bar{w}_k and define $\bar{w}_k = \bar{F}(y_k) - F(y_k)$ in this proof. By Lemma 1(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^* .

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

$$= ||x_k - x^*||^2 - ||x_{k+1} - x_k||^2 - 2\gamma_k (F(y_k) + \bar{w}_k)^T (x_{k+1} - x^*).$$
(3)

Since F is monotone over \mathbb{R}^n , by adding $2\gamma_k(F(y_k) - F(x^*))^T(y_k - x^*)$ to the right hand side (rhs)

of (3), we obtain:

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

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

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

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

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

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

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

$$(4)$$

Since $x_{k+1}, x_{k-1} \in X$, by Lemma 1(i), we may conclude that

$$(x_k - x_{k-1} + \gamma_{k-1}(F(y_{k-1}) + \bar{w}_{k-1}))^T (x_k - x_{k+1}) \le 0$$
 and $(x_k - x_{k-1} + \gamma_{k-1}(F(y_{k-1}) + \bar{w}_{k-1}))^T (x_k - x_{k-1}) \le 0$.

Adding these two inequalities yields the following:

$$(x_k - x_{k-1} + \gamma_{k-1}(F(y_{k-1}) + \bar{w}_{k-1}))^T(y_k - x_{k+1}) \le 0,$$

since $y_k = 2x_k - x_{k-1}$, leading to the following inequality:

$$2\gamma_{k-1}(F(y_{k-1}) + \bar{w}_{k-1})^T (y_k - x_{k+1}) \le 2(x_k - x_{k-1})^T (x_{k+1} - y_k)$$

$$= 2(y_k - x_k)^T (x_{k+1} - y_k) = ||x_{k+1} - x_k||^2 - ||x_k - y_k||^2 - ||x_{k+1} - y_k||^2,$$
(5)

where the first equality follows from recalling that $y_k = 2x_k - x_{k-1}$. Now, we may bound $2\gamma_k(F(y_{k-1}) + \bar{w}_k)^T(y_k - x_{k+1})$ as follows:

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

 $- 2\gamma_k (F(y_{k-1}) + \bar{w}_{k-1})^T (y_k - x_{k+1}) + 2\gamma_k (F(y_{k-1}) + \bar{w}_{k-1})^T (y_k - x_{k+1})$
 $= 2\gamma_k (\bar{w}_k - \bar{w}_{k-1})^T (y_k - x_{k+1}) + 2\left(\frac{\gamma_k}{\gamma_{k-1}}\right) \gamma_{k-1} (F(y_{k-1}) + \bar{w}_{k-1})^T (y_k - x_{k+1})$
 $\leq 8\gamma_k^2 \|\bar{w}_k - \bar{w}_{k-1}\|^2 + \frac{1}{8} \|x_{k+1} - y_k\|^2 - \frac{\gamma_k}{\gamma_{k-1}} \|x_{k+1} - y_k\|^2 + \frac{\gamma_k}{\gamma_{k-1}} \|x_{k+1} - x_k\|^2 - \frac{\gamma_k}{\gamma_{k-1}} \|x_k - y_k\|^2$
 $= 8\gamma_k^2 \|\bar{w}_k - \bar{w}_{k-1}\|^2 + \left(\frac{1}{8} - \frac{\gamma_k}{\gamma_{k-1}}\right) \|x_{k+1} - y_k\|^2 + \frac{\gamma_k}{\gamma_{k-1}} \|x_{k+1} - x_k\|^2 - \frac{\gamma_k}{\gamma_{k-1}} \|x_k - y_k\|^2,$ (6)

where $2\gamma_k(w_k - w_{k-1})^T(y_k - x_{k+1}) \leq 8\gamma_k^2 ||w_k - w_{k-1}||^2 + \frac{1}{8} ||x_{k+1} - y_k||^2$ and inequality (5) 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

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

$$\le 2\gamma_k L ||y_k - y_{k-1}|| ||y_k - x_{k+1}|| \le 8\gamma_k^2 L^2 ||y_k - y_{k-1}||^2 + \frac{1}{8} ||x_{k+1} - y_k||^2$$
(7)

$$\leq 16\gamma_k^2 L^2 \|x_k - y_{k-1}\|^2 + 16\gamma_k^2 L^2 \|x_k - y_k\|^2 + \frac{1}{8} \|x_{k+1} - y_k\|^2, \tag{8}$$

where (8) follows from $||u+v||^2 \le 2||u||^2 + 2||v||^2$. Using (6) and (8), we deduce from (4) that

$$||x_{k+1} - x^*||^2 \le ||x_k - x^*||^2 - \left(1 - \frac{\gamma_k}{\gamma_{k-1}}\right) ||x_{k+1} - x_k||^2 - \left(\frac{\gamma_k}{\gamma_{k-1}} - 16\gamma_k^2 L^2\right) ||x_k - y_k||^2 - \left(\frac{\gamma_k}{\gamma_{k-1}} - \frac{1}{4}\right) ||x_{k+1} - y_k||^2 + 16\gamma_k^2 L^2 ||x_k - y_{k-1}||^2 + 8\gamma_k^2 ||\bar{w}_k - \bar{w}_{k-1}||^2 - 2\gamma_k (F(x^*) + \bar{w}_k)^T (y_k - x^*).$$

$$(9)$$

By assumption, $0 \le \frac{1}{8\tilde{L}}$ for all k,

$$16\gamma_k^2 L^2 \le 16\gamma_k^2 \tilde{L}^2 \le \frac{1}{4} \le \left(\frac{\gamma_{k-1}}{\gamma_{k-2}} - \frac{1}{4}\right). \tag{10}$$

Consequently, from (9) and by invoking (10), we may conclude the following:

$$||x_{k+1} - x^*||^2 + \left(\frac{\gamma_k}{\gamma_{k-1}} - \frac{1}{4}\right) ||x_{k+1} - y_k||^2 \le ||x_k - x^*||^2 + \left(\frac{\gamma_{k-1}}{\gamma_{k-2}} - \frac{1}{4}\right) ||x_k - y_{k-1}||^2 + 8\gamma_k^2 ||w_k - w_{k-1}||^2 - \left(\frac{\gamma_k}{\gamma_{k-1}} - 16\gamma_k^2 L^2\right) ||x_k - y_k||^2 - 2\gamma_k F(x^*)^T (y_k - x^*) + \gamma_k \bar{w}_k^T (y_k - x^*).$$
 (11)

We may bound $2\gamma_k F(x^*)^T (y_k - x^*)$ as follows when $\gamma_{k-1} \ge \gamma_k$:

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

$$\leq -2\gamma_k F(x^*)^T (x_k - x^*) - 2\gamma_k F(x^*)^T (x_k - x^*) + 2\gamma_{k-1} F(x^*)^T (x_{k-1} - x^*).$$
(12)

By substituting $\gamma_k = \gamma$ for every k in (12), we obtain the required result.

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

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

$$+ 8\gamma^2 ||\bar{w}_k - \bar{w}_{k-1}||^2 - (1 - 16\gamma^2 L^2) ||x_k - y_k||^2 - 2\gamma F(x^*)^T (x_k - x^*) - 2\gamma \bar{w}_k^T (y_k - x^*). \tag{13}$$

By leveraging this lemma, we prove a.s. convergence of the sequence produced by (v-SPRG).

Theorem 1 (a.s. convergence of (v-SPRG)). Consider a sequence generated by (v-SPRG). Let Assumptions 1 – 4 hold. Suppose $0 < \gamma \le \frac{1}{8\tilde{L}}$ where $\tilde{L}^2 \triangleq (L^2 + \frac{10\nu_1^2}{N_0})$ and $\{N_k\}$ is a non-decreasing sequence satisfying $\sum_{k=1}^{\infty} \frac{1}{N_k} < M$. Then for any $x_0 \in X$, the sequence generated by (v-SPRG) converges to a point in X^* in an a.s. sense.

Proof. Using (13), taking expectations conditioned on \mathcal{F}_k , and invoking Assumption 3 and 4, we obtain the following.

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

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

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

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

$$+ \frac{16\gamma^{2}(\nu_{1}^{2}(\|y_{k}\|^{2} + \|y_{k-1}\|^{2}) + 2\nu_{2}^{2})}{N_{k}} - \left(1 - 16\gamma^{2}L^{2}\right) \|x_{k} - y_{k}\|^{2} \\
\leq \|x_{k} - x^{*}\|^{2} + \frac{3}{4}\|x_{k} - y_{k-1}\|^{2} + 2\gamma F(x^{*})^{T}(x_{k-1} - x^{*}) - 2\alpha\gamma \operatorname{dist}(x_{k}, X^{*}) \\
+ \frac{16\gamma^{2}(\nu_{1}^{2}(3\|y_{k}\|^{2} + 2\|y_{k-1} - y_{k}\|^{2}) + 2\nu_{2}^{2})}{N_{k}} - \left(1 - 16\gamma^{2}L^{2}\right) \|x_{k} - y_{k}\|^{2} \\
\leq \|x_{k} - x^{*}\|^{2} + \frac{3}{4}\|x_{k} - y_{k-1}\|^{2} + 2\gamma F(x^{*})^{T}(x_{k-1} - x^{*}) - 2\alpha\gamma \operatorname{dist}(x_{k}, X^{*}) \\
+ \frac{16\gamma^{2}(\nu_{1}^{2}(6\|y_{k} - x_{k}\|^{2} + 6\|x_{k}\|^{2} + 4\|y_{k-1} - x_{k}\|^{2} + 4\|y_{k} - x_{k}\|^{2}) + 2\nu_{2}^{2}}{N_{k}} - \left(1 - 16\gamma^{2}L^{2}\right) \|x_{k} - y_{k}\|^{2} \\
\leq \|x_{k} - x^{*}\|^{2} + \frac{3}{4}\|x_{k} - y_{k-1}\|^{2} + 2\gamma F(x^{*})^{T}(x_{k-1} - x^{*}) - 2\alpha\gamma \operatorname{dist}(x_{k}, X^{*}) \\
+ \frac{16\gamma^{2}(\nu_{1}^{2}(10\|y_{k} - x_{k}\|^{2} + 12\|x_{k} - x^{*}\|^{2} + 12\|x^{*}\|^{2} + 4\|y_{k-1} - x_{k}\|^{2}) + 2\nu_{2}^{2}}{N_{k}} - \left(1 - 16\gamma^{2}L^{2}\right) \|x_{k} - y_{k}\|^{2} \\
\leq \left(1 + \frac{192\gamma^{2}\nu_{1}^{2}}{N_{k}}\right) \left(\|x_{k} - x^{*}\|^{2} + \frac{3}{4}\|x_{k} - y_{k-1}\|^{2} + 2\gamma F(x^{*})^{T}(x_{k-1} - x^{*})\right) - 2\alpha\gamma \operatorname{dist}(x_{k}, X^{*}) \\
+ \frac{192\gamma^{2}\|x^{*}\|^{2} + 32\gamma^{2}\nu_{2}^{2}}{N_{k}} - \left(1 - 16\gamma^{2}L^{2} - \frac{160\gamma^{2}\nu_{1}^{2}}{N_{k}}\right) \|x_{k} - y_{k}\|^{2} \\
= \left(1 + \frac{192\gamma^{2}\nu_{1}^{2}}{N_{k}}\right) v_{k} - \delta_{k} + \psi_{k}, \tag{14}$$

where v_k , δ_k , and ψ_k are random variables defined as

$$\begin{split} v_k &\triangleq \|x_k - x^*\|^2 + \frac{3}{4} \|x_k - y_{k-1}\|^2 + 2\gamma F(x^*)^T (x_k - x^*), \\ \delta_k &\triangleq \left(1 - 16\gamma^2 L^2 - \frac{160\gamma^2 \nu_1^2}{N_k}\right) \|x_k - y_k\|^2 + 2\alpha\gamma \mathrm{dist}\left(x_k, X^*\right), \text{ and } \psi_k \triangleq \frac{192\gamma^2 \|x^*\|^2 + 32\gamma^2 \nu_2^2}{N_k}. \end{split}$$

Since $x^* \in X^*$, $v_k \ge 0$ for every k while $\psi_k \ge 0$ for every k follows immediately. In addition, by assumption, $\gamma \le \frac{1}{8L}$ where

$$\left(16\gamma^2 L^2 + \frac{160\gamma^2 \nu_1^2}{N_k}\right) = 16\gamma^2 \tilde{L}^2 \le \frac{1}{4}, \text{ where } \tilde{L}^2 \triangleq \left(L^2 + \frac{10\nu_1^2}{N_0}\right).$$

Consequently, by the choice of γ and by noting that $\operatorname{dist}(x_k,X^*)\geq 0$ for all k, it follows that $\psi_k\geq 0$ for all k. Furthermore, $\sum_k \psi_k < \infty$ since $\sum_k \frac{1}{N_k} < \infty$. We may now invoke Lemma 2 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. when $\gamma \leq \frac{1}{8\tilde{L}}$:

$$\infty > \sum_{k} \left(\left(1 - 16\gamma^{2}L^{2} - \frac{160\gamma^{2}\nu_{1}^{2}}{N_{k}} \right) \|x_{k} - y_{k}\|^{2} + 2\alpha\gamma \operatorname{dist}\left(x_{k}, X^{*}\right) \right)$$

$$\geq \sum_{k} \left(\left(1 - \frac{1}{4} \right) \|x_{k} - y_{k}\|^{2} + 2\alpha\gamma \operatorname{dist}\left(x_{k}, X^{*}\right) \right) = \sum_{k} \left(\frac{3}{4} \|x_{k} - y_{k}\|^{2} + 2\alpha\gamma \operatorname{dist}\left(x_{k}, X^{*}\right) \right).$$

Consequently, we have that $\infty > \sum_k \left(\frac{3}{4} \|x_k - y_k\|^2 + 2\alpha\gamma \mathrm{dist}\left(y_k, X^*\right)\right)$, implying that $x_k - y_k \xrightarrow[a.s.]{k \to \infty} 0$ and $\mathrm{dist}(y_k, X^*) \xrightarrow[a.s.]{k \to \infty} 0$. Since $\{x_k\}$ and $\{y_k\}$ have the same set of limit points with probability one, we may conclude that $\mathrm{dist}(x_k, X^*) \xrightarrow[a.s.]{k \to \infty} 0$ and $\{x_k\}$ is bounded a.s.. It follows that with probability one, $\{x_k\}$ has a convergent subsequence; we denote this subsequence by \mathcal{K} and its limit point by $x_1^*(\omega)$. From hereafter, we suppress ω for ease of exposition. Since $x_k - y_k \xrightarrow[a.s.]{k \to \infty} 0$ or $x_k - x_{k-1} \xrightarrow[a.s.]{k \to \infty} 0$, we have $x_k - y_{k-1} = (x_k - x_{k-1}) - (x_{k-1} - x_{k-2}) \xrightarrow[a.s.]{k \to \infty} 0$. Since $\mathrm{dist}(x_k, X^*) \xrightarrow[a.s.]{k \to \infty} 0$, every limit point of $\{x_k\}$ lies in X^* in an a.s. sense and for any convergent

subsequence \mathcal{K} , $F(x^*)^T(x_k-x^*) \xrightarrow{k\in\mathcal{K}, k\to\infty} F(x^*)^T(x_1^*-x^*) = 0$ by Lemma 4, where $x_1^*\in X^*$ in an a.s. sense. Therefore, $\{\|x_k-y_{k-1}\|^2+F(x^*)^T(x_k-x^*)\}$ converges in an a.s. sense. Since $\{\|x_k-x^*\|^2+\|x_k-y_{k-1}\|^2+F(x^*)^T(x_k-x^*)\}$ converges a.s. (Lemma 2), it follows that $\{\|x_k-x^*\|^2\}$ converges a.s. because $\{\|x_k-y_{k-1}\|^2+F(x^*)^T(x_k-x^*)\}$ converges a.s.. Since $\{\|x_k-x^*\|^2\}$ converges a.s. for any $x^*\in X^*$, it also converges a.s. for $x^*=x_1^*$ and $\|x_k-x_1^*\| \xrightarrow{k\in\mathcal{K}, k\to\infty} 0$; i.e. some subsequence of $\{\|x_k-x_1^*\|^2\}$ converges to zero. Since $\{\|x_k-x_1^*\|\}$ is convergent a.s., we may conclude that the entire sequence $\{x_k\}$ converges to $x_1^*\in X^*$ in an a.s. sense.

Next we derive rate statements for the averaged sequence in the merely monotone regimes without imposing a weak sharpness requirement. However, we do require a compactness requirement on X, a more common restriction when conducting rate analysis. 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 (cf. [13]).

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 VI. We establish the convergence rate for $(\mathbf{v}\text{-}\mathbf{SPRG})$ by using the gap function. Importantly, we attain a rate of $\mathcal{O}(1/K)$ in terms of the expected gap and derive the oracle complexity.

Theorem 2. Consider the (**v-SPRG**) scheme and let $\{\bar{x}_K\}$ be defined as $\bar{x}_K = \sum_{k=0}^{K-1} x_k / K$, where $0 < \gamma \le \frac{1}{8\bar{L}}$ where $\tilde{L}^2 \triangleq \left(L^2 + \frac{10\nu_1^2}{N_0}\right)$ and $\{N_k\}$ is a non-decreasing sequence satisfying $\sum_{k=1}^{\infty} \frac{1}{N_k} < M$. Let Assumptions 1, 2, and 4 hold. In addition, for any $u, v \in X$, suppose that there exists a $D_X > 0$ such that $||u - v||^2 \le D_X^2$.

- (a) Then we have $\mathbb{E}[G(\bar{x}_K)] \leq \mathcal{O}(\frac{1}{K})$ for any K.
- (b) Suppose $N_k \triangleq \lfloor k^a \rfloor$ for a > 1. Then the oracle complexity to ensure that $\mathbb{E}[G(\bar{x}_K)] \leq \epsilon$ satisfies $\sum_{k=1}^K N_k \leq \mathcal{O}\left(\frac{1}{\epsilon^{a+1}}\right)$.

Proof. (a) From (13), we obtain

$$2\gamma F(y)^{T}(x_{k}-y) \leq (\|x_{k}-y\|^{2} + \frac{3}{4}\|x_{k}-y_{k-1}\|^{2} + 2\gamma F(y)^{T}(x_{k-1}-y))$$

$$-(\|x_{k+1}-y\|^{2} + \frac{3}{4}\|x_{k+1}-y_{k}\|^{2} + 2\gamma F(x^{*})^{T}(x_{k}-y)) + 8\gamma^{2}\|\bar{w}_{k}-\bar{w}_{k-1}\|^{2}$$

$$-(1-16\gamma^{2}L^{2})\|x_{k}-y_{k}\|^{2} - 2\gamma \bar{w}_{k}^{T}(y_{k}-x_{k}) - 2\gamma_{k}\bar{w}_{k}^{T}(x_{k}-y). \tag{15}$$

We now define an auxiliary sequence $\{u_k\}$ such that

$$u_{k+1} := \Pi_X(u_k - \gamma \bar{w}_k),$$

where $u_0 \in X$. We have $||u_{k+1} - y||^2 = ||\Pi_X(u_k - \gamma \bar{w}_k) - y||^2 \le ||u_k - \gamma \bar{w}_k - y||^2 = ||u_k - y||^2 - 2\gamma \bar{w}_k(y - u_k) + \gamma^2 ||\bar{w}_k||^2$. Then we may then express the last term on the right in (15) as follows.

$$2\gamma \bar{w}_k^T(y - x_k) = 2\gamma \bar{w}_k^T(y - u_k) + 2\gamma \bar{w}_k^T(u_k - x_k)$$

$$\leq \|u_k - y\|^2 - \|u_{k+1} - y\|^2 + \gamma^2 \|\bar{w}_k\|^2 + 2\gamma \bar{w}_k^T (u_k - x_k). \tag{16}$$

Summing over k and invoking (16), we obtain the following bound:

$$2\gamma \sum_{k=0}^{K-1} F(y)^T (x_k - y) \le ||x_0 - y||^2 + \frac{3}{4} ||x_0 - y_{-1}||^2 + 2\gamma F(y)^T (x_0 - y)$$

$$+ 8\gamma^2 \sum_{k=0}^{K-1} ||\bar{w}_k - \bar{w}_{k-1}||^2 - 2\gamma \sum_{k=0}^{K-1} \bar{w}_k^T (y_k - x_k) - 2\gamma \sum_{k=0}^{K-1} \bar{w}_k^T (x_k - y)$$

$$\implies \frac{2\gamma}{K} \sum_{k=0}^{K-1} F(y)^T (x_k - y) \le \frac{1}{K} (||x_0 - y||^2 + \frac{3}{4} ||x_0 - y_{-1}||^2 + 2\gamma F(y)^T (x_0 - y))$$

$$+ \frac{8\gamma^2 \sum_{k=0}^{K-1} ||\bar{w}_k - \bar{w}_{k-1}||^2}{K} + \frac{\sum_{k=0}^{K-1} 2\gamma \bar{w}_k^T (x_k - y_k)}{K} + \frac{\sum_{k=0}^{K-1} 2\gamma \bar{w}_k^T (y - x_k)}{K}$$
or $F(y)^T (\bar{x}_K - y) \le \frac{1}{2\gamma K} (||x_0 - y||^2 + \frac{3}{4} ||x_0 - y_{-1}||^2 + 2\gamma F(y)^T (x_0 - y))$

$$+ \frac{8\gamma^2 \sum_{k=0}^{K-1} ||\bar{w}_k - \bar{w}_{k-1}||^2}{2\gamma K} + \frac{\sum_{k=0}^{K-1} 2\gamma \bar{w}_k^T (y - x_k)}{2\gamma K} + \frac{\sum_{k=0}^{K-1} 2\gamma \bar{w}_k^T (x_k - y_k)}{2\gamma K}$$

$$\le \frac{1}{2\gamma K} (||x_0 - y||^2 + \frac{3}{4} ||x_0 - y_{-1}||^2 + 2\gamma F(y)^T (x_0 - y))$$

$$+ \frac{\gamma^2 \sum_{k=0}^{K-1} (8||\bar{w}_k - \bar{w}_{k-1}||^2)}{2\gamma K}$$

$$\le \frac{1}{2\gamma K} (||x_0 - y||^2 + \frac{3}{4} ||x_0 - y_{-1}||^2 + 2\gamma F(y)^T (x_0 - y))$$

$$+ \frac{\gamma^2 \sum_{k=0}^{K-1} (8||\bar{w}_k - \bar{w}_{k-1}||^2)}{2\gamma K}$$

$$\le \frac{1}{2\gamma K} (||x_0 - y||^2 + ||u_0 - y||^2 + \frac{3}{4} ||x_0 - y_{-1}||^2 + 2\gamma F(y)^T (x_0 - y))$$

$$+ \frac{\gamma^2 \sum_{k=0}^{K-1} (8||\bar{w}_k - \bar{w}_{k-1}||^2 + ||\bar{w}_k||^2)}{2\gamma K} + \frac{\sum_{k=0}^{K-1} 2\gamma \bar{w}_k^T (u_k - y_k)}{2\gamma K}$$

$$\le \frac{C_X^2}{2\gamma K} + \frac{\gamma^2 \sum_{k=0}^{K-1} (8||\bar{w}_k - \bar{w}_{k-1}||^2 + ||\bar{w}_k||^2)}{2\gamma K} + \frac{\sum_{k=0}^{K-1} 2\gamma \bar{w}_k^T (u_k - y_k)}{2\gamma K}.$$

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

$$G(\bar{x}_K) \triangleq \sup_{y \in X} F(y)^T (\bar{x}_K - y) \le \frac{2C_X^2}{2\gamma K} + \frac{\gamma^2 \sum_{k=0}^{K-1} (8\|\bar{w}_k - \bar{w}_{k-1}\|^2 + \|\bar{w}_k\|^2)}{2\gamma K} + \frac{\sum_{k=0}^{K-1} 2\gamma \bar{w}_k^T (u_k - y_k))}{2\gamma K},$$

where

$$||x_{0} - y||^{2} + ||u_{0} - y||^{2} + \frac{3}{4}||x_{0} - y_{-1}||^{2} + 2\gamma F(y)^{T}(x_{0} - y) \leq \frac{11}{4}D_{X}^{2} + 8\gamma^{2}||F(y) - F(x^{*})||^{2} + 8\gamma^{2}||F(x^{*})||^{2} + 2||x_{0} - y||^{2} \leq \frac{19}{4}D_{X}^{2} + \underbrace{8\gamma^{2}L^{2}D_{X}^{2}}_{\leq 8\gamma^{2}L^{2}D_{X}^{2}} D_{X}^{2} + 8\gamma^{2}C^{2} \leq 5D_{X}^{2} + 8\gamma^{2}C^{2} \triangleq C_{X}^{2}.$$

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

$$\mathbb{E}[G(\bar{x}_K)] \leq \frac{C_X^2}{2\gamma K} + \frac{\gamma^2 \sum_{k=0}^{K-1} 8\mathbb{E}[\|\bar{w}_k - \bar{w}_{k-1}\|^2] + \mathbb{E}[\|\bar{w}_k\|^2]}{2\gamma K} + \frac{\sum_{k=0}^{K-1} 2\gamma \mathbb{E}[\bar{w}_k^T (u_k - y_k)]}{2\gamma K}$$

$$\leq \frac{C_X^2 + \gamma^2 \sum_{k=0}^{K-1} \frac{\nu_1^2 (17\|x_k\|^2 + 16\|x_{k-1}\|^2) + 33\nu_2^2}{2\gamma K}}{2\gamma K} \leq \frac{C_X^2 + \gamma^2 \sum_{k=0}^{K-1} \frac{\nu_1^2 (66D_X^2 + 66\|x^*\|^2) + 33\nu_2^2}{N_k}}{2\gamma K}$$

$$\leq \frac{C_X^2 + \gamma^2 M(\nu_1^2 (66D_X^2 + 66\|x^*\|^2) + 33\nu_2^2)}{2\gamma K} = \frac{\widehat{C}}{K},\tag{17}$$

by defining $\widehat{C} \triangleq \left(C_X^2 + \gamma^2 M(\nu_1^2(66D_X^2 + 66\|x^*\|^2) + 33\nu_2^2)\right)/2\gamma$. It follows that $\mathbb{E}[G(\bar{x}_K)] \leq \mathcal{O}(1/K)$. (b) It follows from (a) that $K = \lfloor (\frac{\widehat{C}}{\epsilon}) \rfloor$. We have

$$\sum_{k=0}^{K-1} N_k \le \sum_{k=0}^{\lfloor (\widehat{C}/\epsilon) \rfloor - 1} (k+1)^a = \sum_{t=1}^{\lfloor (\widehat{C}/\epsilon) \rfloor} t^a \le \int_1^{(\widehat{C}/\epsilon) + 1} x^a dx \le \frac{((\widehat{C}/\epsilon) + 1)^{a+1}}{a+1} \le \left(\frac{\widetilde{C}}{\epsilon^{a+1}}\right).$$

3.2 Stochastic Subgradient Extragradient Schemes

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

Proposition 1 (a.s. convergence of (v-SSE)). Consider a sequence generated by (v-SSE). Let Assumptions 1 and 4 hold. Suppose $0 < \gamma_k = \gamma \le \frac{1}{\sqrt{2}\tilde{L}}$ where $\tilde{L}^2 \triangleq \left(L^2 + \frac{4\nu_1^2}{N_0}\right)$ and $\{N_k\}$ is a non-decreasing sequence satisfying $\sum_{k=1}^{\infty} \frac{1}{N_k} < M$. Then for any $x_0 \in X$, the sequence generated by (v-SSE) converges to a point in X^* in an a.s. sense.

Proof. By Lemma 1(ii) we have for any x^* ,

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

$$= ||x_k - x^*||^2 - ||x_k - x_{k+1}||^2 + 2\gamma_k(F(x_{k+\frac{1}{2}}) + \bar{w}_{k+\frac{1}{2}})^T(x^* - x_{k+1}). \tag{18}$$

It is clear that

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

$$= F(x_{k+\frac{1}{2}})^{T}(x_{k+1} - x_{k+\frac{1}{2}}) + F(x_{k+\frac{1}{2}})^{T}(x_{k+\frac{1}{2}} - x^{*}).$$
(19)

Substituting (19) in (18), we obtain

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

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

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

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

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

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

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

$$+ 2(x_{k+1} - x_{k+\frac{1}{2}})^T (x_k - \gamma_k F(x_{k+\frac{1}{2}}) - x_{k+\frac{1}{2}}) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T (x^* - x_{k+1}). \tag{20}$$

By definition of C_k , we have

$$(x_{k+1} - x_{k+\frac{1}{2}})^T (x_k - \gamma_k (F(x_k) + \bar{w}_k) - x_{k+\frac{1}{2}}) \le 0.$$
(21)

Substituting (21) in (20), we deduce that

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - \|x_k - x_{k+\frac{1}{2}}\|^2 - \|x_{k+\frac{1}{2}} - x_{k+1}\|^2 - 2\gamma_k F(x_{k+\frac{1}{2}})^T (x_{k+\frac{1}{2}} - x^*) \\ &+ 2\gamma_k (x_{k+1} - x_{k+\frac{1}{2}})^T (F(x_k) - F(x_{k+\frac{1}{2}})) + 2\gamma_k \bar{w}_k^T (x_{k+1} - x_{k+\frac{1}{2}}) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T (x^* - x_{k+1}) \\ &\leq \|x_k - x^*\|^2 - \|x_k - x_{k+\frac{1}{2}}\|^2 - \|x_{k+\frac{1}{2}} - x_{k+1}\|^2 - 2\gamma_k F(x_{k+\frac{1}{2}})^T (x_{k+\frac{1}{2}} - x^*) \\ &+ 2\gamma_k \|x_{k+1} - x_{k+\frac{1}{2}}\| \|F(x_k) - F(x_{k+\frac{1}{2}})\| + 2\gamma_k (\bar{w}_k - \bar{w}_{k+\frac{1}{2}})^T (x_{k+1} - x_{k+\frac{1}{2}}) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T (x^* - x_{k+\frac{1}{2}}) \\ &\leq \|x_k - x^*\|^2 - \|x_k - x_{k+\frac{1}{2}}\|^2 - \|x_{k+\frac{1}{2}} - x_{k+1}\|^2 + \frac{1}{2}\|x_{k+1} - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 L^2 \|x_k - x_{k+\frac{1}{2}}\|^2 \\ &+ 2\gamma_k^2 \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 + \frac{1}{2}\|x_{k+1} - x_{k+\frac{1}{2}}\|^2 - 2\gamma_k F(x_{k+\frac{1}{2}})^T (x_{k+\frac{1}{2}} - x^*) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T (x^* - x_{k+\frac{1}{2}}) \\ &= \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2) \|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 \\ &- 2\gamma_k F(x_{k+\frac{1}{2}})^T (x_{k+\frac{1}{2}} - x^*) + 2\gamma_k \bar{w}_{k+\frac{1}{2}}^T (x^* - x_{k+\frac{1}{2}}) \\ &= \|x_k - x^*\|^2 - (1 - 2\gamma^2 L^2) \|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma^2 \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 \\ &- 2\gamma F(x_{k+\frac{1}{2}})^T (x_{k+\frac{1}{2}} - x^*) + 2\gamma \bar{w}_{k+\frac{1}{2}}^T (x^* - x_{k+\frac{1}{2}}), \end{aligned} \tag{22}$$

by noticing that $\gamma_k = \gamma$. Define $r_{\gamma}(x) \triangleq ||x - \Pi_X(x - \gamma F(x))||$ as a residual function. We have

$$r_{\gamma}^{2}(x_{k}) = \|x_{k} - \Pi_{X}(x_{k} - \gamma F(x_{k}))\|^{2}$$

$$= \|x_{k} - x_{k + \frac{1}{2}} + \Pi_{X}(x_{k} - \gamma F(x_{k}) - \gamma \bar{w}_{k}) - \Pi_{X}(x_{k} - \gamma F(x_{k}))\|$$

$$\leq 2\|x_{k} - x_{k + \frac{1}{2}}\|^{2} + 2\gamma^{2}\|\bar{w}_{k}\|^{2}.$$

It follows that

$$-\frac{1}{2}\|x_k - x_{k+\frac{1}{2}}\|^2 \le -\frac{1}{4}r_{\gamma}^2(x_k) + \frac{1}{2}\gamma^2\|\bar{w}_k\|^2.$$
 (23)

Using (23) in (22), we obtain

$$\begin{split} &\|x_{k+1}-x^*\|^2 \leq \|x_k-x^*\|^2 - \left(\frac{1}{2}-2\gamma^2L^2\right)\|x_k-x_{k+\frac{1}{2}}\|^2 + 2\gamma^2\|\bar{w}_k-\bar{w}_{k+\frac{1}{2}}\|^2 \\ &-2\gamma F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}}-x^*) + 2\gamma \bar{w}_{k+\frac{1}{2}}^T(x^*-x_{k+\frac{1}{2}}) - \frac{1}{2}\|x_k-x_{k+\frac{1}{2}}\|^2 \\ &\leq \|x_k-x^*\|^2 - \left(\frac{1}{2}-2\gamma^2L^2\right)\|x_k-x_{k+\frac{1}{2}}\|^2 + 2\gamma^2\|\bar{w}_k-\bar{w}_{k+\frac{1}{2}}\|^2 \\ &-2\gamma F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}}-x^*) + 2\gamma \bar{w}_{k+\frac{1}{2}}^T(x^*-x_{k+\frac{1}{2}}) - \frac{1}{4}r_{\gamma}^2(x_k) + \frac{1}{2}\gamma^2\|\bar{w}_k\|^2 \\ &\leq \|x_k-x^*\|^2 - \left(\frac{1}{2}-2\gamma^2L^2\right)\|x_k-x_{k+\frac{1}{2}}\|^2 + \frac{9}{2}\gamma^2\|\bar{w}_k\|^2 + 4\gamma^2\|\bar{w}_{k+\frac{1}{2}}\|^2 \\ &-2\gamma F(x_{k+\frac{1}{2}})^T(x_{k+\frac{1}{2}}-x^*) + 2\gamma \bar{w}_{k+\frac{1}{2}}^T(x^*-x_{k+\frac{1}{2}}) - \frac{1}{4}r_{\gamma}^2(x_k). \end{split}$$

Taking expectations conditioned on \mathcal{F}_k , we obtain the following bound:

$$\mathbb{E}[\|x_{k+1} - x^*\|^2 \mid \mathcal{F}_k] \le \|x_k - x^*\|^2 - (1 - 2\gamma^2 L^2) \mathbb{E}[\|x_k - x_{k+\frac{1}{2}}\|^2 \mid \mathcal{F}_k]$$

$$+ \mathbb{E}\left[\mathbb{E}\left[4\gamma^{2} \| \bar{w}_{k+\frac{1}{2}} \|^{2} \mid \mathcal{F}_{k+\frac{1}{2}}\right] \mid \mathcal{F}_{k}\right] + \mathbb{E}\left[\frac{9}{2}\gamma^{2} \| \bar{w}_{k} \|^{2} \mid \mathcal{F}_{k}\right] \\
- \mathbb{E}\left[\mathbb{E}\left[2\gamma \bar{w}_{k+\frac{1}{2}}^{T} (x_{k+\frac{1}{2}} - x^{*}) \mid \mathcal{F}_{k+\frac{1}{2}}\right] \mid \mathcal{F}_{k}\right] - \frac{1}{4}r_{\gamma}^{2}(x_{k}) \\
\leq \|x_{k} - x^{*} \|^{2} - (1 - 2\gamma^{2}L^{2})\mathbb{E}\left[\|x_{k} - x_{k+\frac{1}{2}} \|^{2} \mid \mathcal{F}_{k}\right] \\
+ \frac{4\gamma^{2}(\nu_{1}^{2}\mathbb{E}\left[\|x_{k+\frac{1}{2}} \|^{2} \mid \mathcal{F}_{k}\right] + \nu_{2}^{2})}{N_{k}} + \frac{\frac{9}{2}\gamma^{2}(\nu_{1}^{2} \|x_{k} \|^{2} + \nu_{2}^{2})}{N_{k}} - \frac{1}{4}r_{\gamma}^{2}(x_{k}) \\
\leq \|x_{k} - x^{*} \|^{2} - (1 - 2\gamma^{2}L^{2})\mathbb{E}\left[\|x_{k} - x_{k+\frac{1}{2}} \|^{2} \mid \mathcal{F}_{k}\right] \\
+ \frac{4\gamma^{2}(2\nu_{1}^{2}\mathbb{E}\left[\|x_{k} - x_{k+\frac{1}{2}} \|^{2} \mid \mathcal{F}_{k}\right] + 2\nu_{1}^{2}\|x_{k} \|^{2} + \nu_{2}^{2})}{N_{k}} + \frac{\frac{9}{2}\gamma^{2}(\nu_{1}^{2} \|x_{k} \|^{2} + \nu_{2}^{2})}{N_{k}} - \frac{1}{4}r_{\gamma}^{2}(x_{k}) \\
\leq \|x_{k} - x^{*} \|^{2} - (1 - 2\gamma^{2}L^{2} - \frac{8\gamma^{2}\nu_{1}^{2}}{N_{k}})\mathbb{E}\left[\|x_{k} - x_{k+\frac{1}{2}} \|^{2} \mid \mathcal{F}_{k}\right] \\
+ \frac{\frac{25}{2}\gamma^{2}(2\nu_{1}^{2} \|x_{k} - x^{*}\|^{2} + 2\nu_{1}^{2} \|x^{*}\|^{2})}{N_{k}} + \frac{17\gamma^{2}\nu_{2}^{2}}{2N_{k}} - \frac{r_{\gamma}^{2}(x_{k})}{4} \\
\leq \left(1 + \frac{25\gamma^{2}\nu_{1}^{2}}{N_{k}}\right) \|x_{k} - x^{*}\|^{2} + \frac{25\gamma^{2}\nu_{1}^{2} \|x^{*}\|^{2}}{N_{k}} + \frac{17\gamma^{2}\nu_{2}^{2}}{2N_{k}} - \frac{r_{\gamma}^{2}(x_{k})}{4}, \tag{24}$$

where the penultimate inequality follows from noting that $1-2\gamma^2L^2-\frac{8\gamma^2\nu_1^2}{N_k}\leq 1-2\gamma^2(L^2+\frac{4\gamma^2\nu_1^2}{N_0}))\leq 1$, if $\gamma<\frac{1}{\sqrt{2}\tilde{L}}$ and $\tilde{L}\triangleq L^2+\frac{4\gamma^2\nu_1^2}{N_0}$. We may now apply Lemma 2 which allows us to claim that $\{\|x_k-x^*\|\}$ is convergent for any $x^*\in X^*$ and $\sum_k r_\gamma(x_k)^2<\infty$ in an a.s. sense. Therefore, in an a.s. sense, we have

$$\lim_{k \to \infty} r_{\gamma}(x_k)^2 = 0.$$

Since $\{\|x_k - x^*\|^2\}$ is a convergent sequence in an a.s. sense, $\{x_k\}$ is bounded a.s. and has a convergent subsequence. Consider any convergent subsequence of $\{x_k\}$ with index set denoted by \mathcal{K} and suppose its limit point is denoted by $\bar{x}(\omega)$. The dependence of \bar{x} on ω is suppressed for ease of exposition. We have that $0 = \lim_{k \in \mathcal{K}} r_{\gamma}(x_k) = r_{\gamma}(\bar{x})$ in an a.s. sense since $r_{\gamma}(\cdot)$ is a continuous map. It follows that \bar{x} is a solution to VI(X, F) in an a.s. sense. Since $\{\|x_k - x^*\|^2\}$ is convergent a.s. since $\bar{x} \in X^*$ in an a.s. sense. Since a subsequence of $\{\|x_k - \bar{x}\|^2\}$, denoted by \mathcal{K} , converges to zero a.s., the entire sequence $\{\|x_k - \bar{x}\|^2\}$ converges to zero. Therefore, the entire sequence $\{x_k\}$ converges a.s. to a point in X^* .

We now proceed to derive a rate statement in terms of the gap function by imposing an extra compactness requirement.

Proposition 2. Consider the (**v-SSE**) scheme and let $\{\bar{x}_K\}$ be defined as $\bar{x}_K = \frac{\sum_{k=0}^{K-1} x_{k+\frac{1}{2}}}{K}$, where $0 < \gamma \le \frac{1}{\sqrt{2}\bar{L}}$ where $\tilde{L}^2 \triangleq \left(L^2 + \frac{4\nu_1^2}{N_0}\right)$ and $\{N_k\}$ is a non-decreasing sequence such that $\sum_{k=1}^{\infty} \frac{1}{N_k} < M$. Let Assumptions 1, 2 and 4 hold. In addition, suppose there exists a $D_X > 0$ such that $\|u - v\|^2 \le D_X^2$ for any $u, v \in X$.

- (a) Then we have $\mathbb{E}[G(\bar{x}_K)] \leq \mathcal{O}(\frac{1}{K})$ for any K.
- (b) Suppose $N_k \triangleq \lfloor k^a \rfloor$, for a > 1. Then the oracle complexity to ensure that $\mathbb{E}[G(\bar{x}_K)] \leq \epsilon$ satisfies $\sum_{k=1}^K N_k \leq \mathcal{O}\left(\frac{1}{\epsilon^{a+1}}\right)$.

Proof. (a) From (22), we obtain

$$2\gamma F(y)^T(x_{k+\frac{1}{2}}-y) \leq \|x_k-y\|^2 - \|x_{k+1}-y\|^2 - (1-2\gamma^2L^2)\|x_k-x_{k+\frac{1}{2}}\|^2 + 2\gamma^2\|\bar{w}_k-\bar{w}_{k+\frac{1}{2}}\|^2$$

$$+2\gamma \bar{w}_{k+\frac{1}{2}}^{T}(y-x_{k+\frac{1}{2}}). \tag{25}$$

We now define an auxiliary sequence $\{u_k\}$ such that

$$u_{k+1} := \Pi_X(u_k - \gamma \bar{w}_{k+\frac{1}{2}}),$$

where $u_0 \in X$. With a similar analysis in (16), we may then express the last term on the right in (25) as follows.

$$2\gamma \bar{w}_{k+\frac{1}{2}}^{T}(y-x_{k+\frac{1}{2}}) = 2\gamma \bar{w}_{k+\frac{1}{2}}^{T}(y-u_{k}) + 2\gamma \bar{w}_{k+\frac{1}{2}}^{T}(u_{k}-x_{k+\frac{1}{2}})$$

$$\leq ||u_{k}-y||^{2} - ||u_{k+1}-y||^{2} + \gamma^{2}||\bar{w}_{k+\frac{1}{2}}||^{2} + 2\gamma \bar{w}_{k+\frac{1}{2}}^{T}(u_{k}-x_{k+\frac{1}{2}}). \tag{26}$$

Summing over k and invoking (26), we obtain the following bound:

$$\sum_{k=0}^{K-1} 2\gamma F(y)^T (x_{k+\frac{1}{2}} - y) \le \|x_0 - y\|^2 + 2\gamma^2 \sum_{k=0}^{K-1} \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 + 2\gamma \sum_{k=0}^{K-1} \bar{w}_{k+\frac{1}{2}}^T (y - x_{k+\frac{1}{2}})$$

$$\implies \frac{2\gamma}{K} \sum_{k=0}^{K-1} F(y)^T (x_{k+\frac{1}{2}} - y) \le \frac{1}{K} \|x_0 - y\|^2 + \frac{2\gamma^2 \sum_{k=0}^{K-1} \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2}{K} + \frac{\sum_{k=0}^{K-1} 2\gamma \bar{w}_{k+\frac{1}{2}}^T (y - x_{k+\frac{1}{2}})}{K}$$
or $F(y)^T (\bar{x}_K - y) \le \frac{1}{2\gamma K} \|x_0 - y\|^2 + \frac{2\gamma^2 \sum_{k=0}^{K-1} \|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2}{2\gamma K} + \frac{\sum_{k=0}^{K-1} 2\gamma \bar{w}_{k+\frac{1}{2}}^T (y - x_{k+\frac{1}{2}})}{2\gamma K}$

$$\le \frac{1}{2\gamma K} \|x_0 - y\|^2 + \frac{\gamma \sum_{k=0}^{K-1} 2\|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2}{2K} + \frac{\|u_0 - y\|^2 + \sum_{k=0}^{K-1} (\gamma^2 \|\bar{w}_{k+\frac{1}{2}}\|^2 + 2\gamma \bar{w}_{k+\frac{1}{2}}^T (u_k - x_{k+\frac{1}{2}}))}{2\gamma K}$$

$$\le \frac{D_X^2}{\gamma K} + \frac{\gamma \sum_{k=0}^{K-1} (2\|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 + \|\bar{w}_{k+\frac{1}{2}}\|^2)}{2K} + \frac{\sum_{k=0}^{K-1} \bar{w}_{k+\frac{1}{2}}^T (u_k - x_{k+\frac{1}{2}})}{K}.$$

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

$$G(\bar{x}_K) \triangleq \sup_{y \in X} F(y)^T(\bar{x}_K - y) \leq \frac{D_X^2}{\gamma K} + \frac{\gamma \sum_{k=0}^{K-1} (2\|\bar{w}_k - \bar{w}_{k+\frac{1}{2}}\|^2 + \|\bar{w}_{k+\frac{1}{2}}\|^2)}{2K} + \frac{\sum_{k=0}^{K-1} \bar{w}_{k+\frac{1}{2}}^T (u_k - x_{k+\frac{1}{2}})}{K}.$$

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

$$\mathbb{E}[G(\bar{x}_{K})] \leq \frac{D_{X}^{2}}{\gamma K} + \frac{\gamma \sum_{k=0}^{K-1} 2\mathbb{E}[\|\bar{w}_{k} - \bar{w}_{k+\frac{1}{2}}\|^{2}] + 2\mathbb{E}[\|\bar{w}_{k+\frac{1}{2}}\|^{2}]}{2K} + \frac{\sum_{k=0}^{K-1} \mathbb{E}[\bar{w}_{k+\frac{1}{2}}^{T}(u_{k} - x_{k+\frac{1}{2}})]}{K} \\
\leq \frac{2D_{X}^{2} + \gamma^{2} \sum_{k=0}^{K-1} \frac{\nu_{1}^{2}(4\|x_{k}\|^{2} + 6\|x_{k+\frac{1}{2}}\|^{2}) + 10\nu_{2}^{2}}{N_{k}}}{2\gamma K} \leq \frac{2D_{X}^{2} + \gamma^{2} \sum_{k=0}^{K-1} \frac{\nu_{1}^{2}(20D_{X}^{2} + 20\|x^{*}\|^{2}) + 10\nu_{2}^{2}}{N_{k}}}{2\gamma K} \\
\leq \frac{2D_{X}^{2} + \gamma^{2} M(\nu_{1}^{2}(20D_{X}^{2} + 20\|x^{*}\|^{2}) + 10\nu_{2}^{2})}{2\gamma K} = \frac{\widehat{C}}{K}, \tag{27}$$

by defining $\widehat{C} \triangleq (2D_X^2 + \gamma^2 M(\nu_1^2(20D_X^2 + 20||x^*||^2) + 10\nu_2^2)/2\gamma$. It follows that $\mathbb{E}[G(\bar{x}_K)] \leq \mathcal{O}(1/K)$. (b) We may prove this result in a fashion similar to that used in Proposition 2(b).

Remark 1. Several aspects of the prior results require further emphasis.

- (a) We observe that the rate guarantees of $\mathcal{O}(1/K)$ in terms of the gap function for variance-reduced schemes for monotone stochastic variational inequality problems matches those obtained by Iusem et al. [16] but with a lower per-iteration complexity. This rate was also achieved by more recently by Jalilzadeh and Shanbhag [18]. Notably, the latter scheme leverages an inexact proximal framework. In fact, proximal-point techniques have been useful in conducting a unified analysis for both gradient and extragradient schemes for deterministic strongly convex-concave saddle point problems as seen by the recent work by Mokhtari et al. [29] that proves a linear rate of convergence in terms of solution error.
- (b) While much of the techniques in the literature rely on uniform bounds on the conditional second moments, we allow for state-dependence akin to that adopted in [16] and do not rely on compactness for proving a.s. convergence statements. Naturally, rate statements do impose a compactness assumption. In addition, weak sharpness is only required for proving a.s. convergence for (v-SPRG) but does not find application elsewhere.
- (c) To the best of our knowledge, this is the first available rate for projected reflected gradient methods in stochastic and merely monotone regimes. The prior rate of linear convergence was provided in deterministic and strongly monotone settings [28]. We also remain unaware of rate statements for SSE schemes and this appears to be the first rate statements for such schemes.

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 4.3 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 set $X = \bigcap_{i \in I} X_i$ and random projection Π_{l_k} . The following assumption is known as linear regularity and is discussed in [39]. It indicates that this condition is a mild restriction in practice.

Assumption 5. There exists a positive scalar η such that for any $x \in \mathbb{R}^n$

$$||x - \Pi_X(x)||^2 \le \eta \max_{i \in I} ||x - \Pi_{X_i}(x)||^2,$$

where I is a finite set of indices, $I = \{1, ..., m\}$.

The following assumption requires that each constraint is sampled with at least some probability and the random samples are nearly independent, which refers to [39].

Assumption 6. The random variables $l_k, k = 0, 1, ...,$ are such that $\inf_{k \geq 0} P(l_k = X_i \mid \mathcal{F}_k) \geq \frac{\rho_i}{m}$ with probability 1 for i = 1, ..., m, where $\rho_i \in (0, 1]$ is a scalar for i = 1, ..., m.

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

Lemma 6. Let X be a closed convex subset of \mathbb{R}^n . We have

$$||y - \Pi_X(y)||^2 \le 2||x - \Pi_X(x)||^2 + 8||x - y||^2, \quad \forall x, y \in \mathbb{R}^n.$$

Proof. Since $y - \Pi_X(y) = (x - \Pi_X(x)) - (x - y) + (\Pi_X(x) - \Pi_X(y))$, we have

$$||y - \Pi_X(y)|| \le ||x - \Pi_X(x)|| + ||x - y|| + ||\Pi_X(x) - \Pi_X(y)|| \le ||x - \Pi_X(x)|| + 2||x - y||.$$

Thus,

$$||y - \Pi_X(y)||^2 \le 2||x - \Pi_X(x)||^2 + 8||x - y||^2$$

where the last inequality leverages $||a+b||^2 \le 2||a||^2 + 2||b||^2$.

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

Lemma 7. Suppose Assumptions 1-3 hold. Then, we have

$$F(x)^T(x - x^*) \ge \alpha \operatorname{dist}(\Pi_X(x), X^*) - C \operatorname{dist}(x, X), \quad \forall x \in \mathbb{R}^n.$$

Proof. We have

$$F(x)^{T}(x-x^{*}) = (F(x) - F(x^{*}))^{T}(x-x^{*}) + F(x^{*})^{T}(\Pi_{X}(x) - x^{*}) + F(x^{*})^{T}(x - \Pi_{X}(x)).$$
(28)

From the monotonicity assumption on F, we have

$$(F(x) - F(x^*))^T (x - x^*) \ge 0.$$
(29)

Since x^* is a solution, it follows that from the weak sharpness property,

$$F(x^*)^T(\Pi_X(x) - x^*) \ge \alpha \operatorname{dist}(\Pi_X(x), X^*). \tag{30}$$

Finally, by recalling that $F(x^*)^T(\Pi_X(x) - x) \leq ||F(x^*)|| ||x - \Pi_X(x)||$ and $||F(x^*)|| \leq C$ (by Assumption 2), we have that

$$F(x^*)^T(x - \Pi_X(x)) \ge -\|F(x^*)\|\|x - \Pi_X(x)\| \ge -C \operatorname{dist}(x, X).$$
(31)

By substituting (29) - (31) in (28), the result follows.

Next, we provide a simple bound on F(x).

Lemma 8. Suppose Assumptions 1-2 hold. Then for any $x \in \mathbb{R}^n$, $||F(x)||^2 \le 2L^2||x-x^*||^2 + 2C^2$.

Finally, we derive a lower bound on $\mathbb{E}[\|x_k - \Pi_{l_k}(x_k)\|^2 \mid \mathcal{F}_k]$.

Lemma 9. Suppose Assumptions 5 and 6 hold. Then for any $l_k \in I$ and any $x \in \mathbb{R}^n$,

$$\mathbb{E}[\|x - \Pi_{l_k}(x)\|^2 \mid \mathcal{F}_k] \ge \frac{\rho}{mn} \operatorname{dist}^2(x, X), \quad k \ge 0,$$

with probability 1, where $\rho \triangleq \min_{i \in I} \{\rho_i\}$.

Proof. Following from Assumption 6, we have

$$\mathbb{E}[\|x - \Pi_{l_k}(x)\|^2 \mid \mathcal{F}_k] = \sum_{i=1}^{m} P(l_k = i \mid \mathcal{F}_k) \|x - \Pi_i(x)\|^2 \ge \frac{\rho}{m} \|x - \Pi_j(x)\|^2, \quad \forall j = 1, \dots, m$$

$$\implies \mathbb{E}[\|x - \Pi_{l_k}(x)\|^2 \mid \mathcal{F}_k] \ge \frac{\rho}{m} \max_{j} \|x - \Pi_j(x)\|^2 \stackrel{\text{(Ass. 5)}}{\ge} \frac{\rho}{m\eta} \text{dist}^2(x, X).$$

4.2 SPRG with random projections

We begin with an a.s. convergence claim for (**r-SPRG**).

Theorem 3. Let Assumptions 1 – 6 hold. Suppose the steplength sequence $\{\gamma_k\}$ satisfies $\sum_{k=0}^{\infty} \gamma_k = \infty$ and $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$. 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 $k \ge 1$ and $w_k = F(y_k, \omega_k) - F(y_k)$. By Lemma 1(ii) and by noting that $x_{k+1} = \prod_{l_k} (x_k - \gamma_k F(2x_k - x_{k-1}))$ and $F(y_k, \omega_k) = F(y_k) + w_k$, we have the following inequality:

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

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

$$= ||x_k - x^*||^2 - ||x_{k+1} - x_k||^2 - 2\gamma_k F(y_k)^T (x_{k+1} - x^*) - 2\gamma_k w_k^T (x_{k+1} - x^*).$$
(32)

Recall that $||y_k - x_{k+1}||^2$ can be expressed as follows.

$$\begin{aligned} \|y_k - x_{k+1}\|^2 &= \|(2x_k - x_{k-1}) - x_{k+1}\|^2 \\ &= \|(x_k - (x_{k-1} - x_k) - x_{k+1})\|^2 \\ &= 2(x_k - x_{k-1})^T (x_k - x_{k+1}) + \|x_{k-1} - x_k\|^2 + \|x_k - x_{k+1}\|^2 \\ &= 2(x_k - x_{k-1})^T (2x_k - 2x_{k+1}) + \|x_{k-1} - x_{k+1}\|^2 \\ &= 2(x_k - x_{k-1})^T (x_k + x_{k-1} - 2x_{k+1}) + 2\|x_k - x_{k-1}\|^2 + \|x_{k-1} - x_{k+1}\|^2 \\ &= 2\|x_k - x_{k+1}\|^2 - 2\|x_{k-1} - x_{k+1}\|^2 + 2\|x_k - x_{k-1}\|^2 + \|x_{k-1} - x_{k+1}\|^2 \\ &= 2\|x_k - x_{k+1}\|^2 - \|x_{k-1} - x_{k+1}\|^2 + 2\|x_k - x_{k-1}\|^2. \end{aligned}$$

Consequently, we have that

$$\frac{1}{4}||x_k - x_{k+1}||^2 = \frac{1}{8}||y_k - x_{k+1}||^2 + \frac{1}{8}||x_{k-1} - x_{k+1}||^2 - \frac{1}{4}||x_k - x_{k-1}||^2.$$
(33)

Using (33) in (32), we obtain

$$||x_{k+1} - x^*||^2 \le ||x_k - x^*||^2 - \frac{3}{4}||x_{k+1} - x_k||^2 - \frac{1}{8}||y_k - x_{k+1}||^2 - \frac{1}{8}||x_{k-1} - x_{k+1}||^2 + \frac{1}{4}||x_k - x_{k-1}||^2 - 2\gamma_k F(y_k)^T (x_{k+1} - x^*) - 2\gamma_k w_k^T (x_{k+1} - x^*) = ||x_k - x^*||^2 - \frac{3}{4}||x_{k+1} - x_k||^2 - \frac{1}{8}||y_k - x_{k+1}||^2 - \frac{1}{8}||x_{k-1} - x_{k+1}||^2 + \frac{1}{4}||x_k - x_{k-1}||^2 - 2\gamma_k F(y_k)^T (y_k - x^*) - 2\gamma_k F(y_k)^T (x_{k+1} - y_k) - 2\gamma_k w_k^T (x_{k+1} - x^*)$$
(34)

$$\leq \|x_k - x^*\|^2 - \frac{3}{4} \|x_{k+1} - x_k\|^2 - \frac{1}{8} \|y_k - x_{k+1}\|^2 - \frac{1}{8} \|x_{k-1} - x_{k+1}\|^2 + \frac{1}{4} \|x_k - x_{k-1}\|^2 - 2\gamma_k \alpha \operatorname{dist} \left(\Pi_X(y_k), X^*\right) + 2\gamma_k C \operatorname{dist}(y_k, X) - 2\gamma_k F(y_k)^T (x_{k+1} - y_k) - 2\gamma_k w_k^T (x_{k+1} - x^*),$$
 (35)

where the last inequality follows from Lemma 7. Since

$$-2\gamma_k F(y_k)^T (x_{k+1} - y_k) \le 16\gamma_k^2 ||F(y_k)||^2 + \frac{1}{16} ||x_{k+1} - y_k||^2,$$

inequality (35) can be rewritten as follows:

$$||x_{k+1} - x^*||^2 \le ||x_k - x^*||^2 - \frac{3}{4}||x_{k+1} - x_k||^2 - \frac{1}{16}||y_k - x_{k+1}||^2 - \frac{1}{8}||x_{k-1} - x_{k+1}||^2 + \frac{1}{4}||x_k - x_{k-1}||^2 - 2\gamma_k \alpha \operatorname{dist} (\Pi_X(y_k), X^*) + 2\gamma_k C \operatorname{dist}(y_k, X) + 16\gamma_k^2 ||F(y_k)||^2 + 16\gamma_k^2 ||w_k||^2 - 2\gamma_k w_k^T(y_k - x^*)$$

$$\le ||x_k - x^*||^2 - \frac{3}{4}||x_{k+1} - x_k||^2 - \frac{1}{16}||y_k - x_{k+1}||^2 - \frac{1}{8}||x_{k-1} - x_{k+1}||^2 + \frac{1}{4}||x_k - x_{k-1}||^2 - 2\gamma_k \alpha \operatorname{dist} (\Pi_X(y_k), X^*) + 2\gamma_k C \operatorname{dist}(y_k, X) + 32\gamma_k^2 L^2 ||y_k - x^*||^2 + 32\gamma_k^2 C^2 + 16\gamma_k^2 ||w_k||^2 - 2\gamma_k w_k^T(y_k - x^*).$$

$$(36)$$

Since

$$-2\gamma_k \alpha \operatorname{dist} (\Pi_X(y_k), X^*) \le -2\gamma_k \alpha \operatorname{dist} (x_k, X^*) + 2\gamma_k \alpha \|x_k - \Pi_X(y_k)\|$$

$$\le -2\gamma_k \alpha \operatorname{dist} (x_k, X^*) + 2\gamma_k \alpha \|x_k - y_k\| + 2\gamma_k \alpha \|y_k - \Pi_X(y_k)\|$$

$$= -2\gamma_k \alpha \operatorname{dist} (x_k, X^*) + 2\gamma_k \alpha \|x_k - y_k\| + 2\gamma_k \alpha \operatorname{dist} (y_k, X),$$

we have

$$||x_{k+1} - x^*||^2 \le ||x_k - x^*||^2 - 2\gamma_k \alpha \operatorname{dist}(x_k, X^*) - \frac{3}{4}||x_{k+1} - x_k||^2 - \frac{1}{16}||y_k - x_{k+1}||^2 - \frac{1}{8}||x_{k-1} - x_{k+1}||^2 + \frac{1}{4}||x_k - x_{k-1}||^2 + 2\gamma_k \alpha ||x_k - y_k|| + 2\gamma_k (C + \alpha) \operatorname{dist}(y_k, X) + 64\gamma_k^2 L^2 ||x_k - x^*||^2 + 64\gamma_k^2 L^2 ||x_k - x_{k-1}||^2 + 32\gamma_k^2 C^2 + 16\gamma_k^2 ||w_k||^2 - 2\gamma_k w_k^T (y_k - x^*).$$
(37)

By Lemma 9,

$$\mathbb{E}[\|y_k - x_{k+1}\|^2 \mid \mathcal{F}_k] \ge \mathbb{E}[\|y_k - \Pi_{l_k} y_k\|^2 \mid \mathcal{F}_k] \ge \frac{\rho}{m\eta} \operatorname{dist}^2(y_k, X). \tag{38}$$

Taking expectations conditioned on \mathcal{F}_k and using (38) in (37), we have

$$\mathbb{E}[\|x_{k+1} - x^*\|^2 + \frac{3}{4}\|x_{k+1} - x_k\|^2 \mid \mathcal{F}_k] \leq \|x_k - x^*\|^2 - 2\gamma_k\alpha\operatorname{dist}(x_k, X^*) - \frac{1}{16}\mathbb{E}[\|y_k - x_{k+1}\|^2 \mid \mathcal{F}_k] \\ - \frac{1}{8}\mathbb{E}[\|x_{k-1} - x_{k+1}\|^2 \mid \mathcal{F}_k] + \frac{1}{4}\|x_k - x_{k-1}\|^2 + 2\gamma_k\alpha\|x_k - y_k\| + 2\gamma_k(C + \alpha)\operatorname{dist}(y_k, X) \\ + 64\gamma_k^2L^2\|x_k - x^*\|^2 + 64\gamma_k^2L^2\|x_k - x_{k-1}\|^2 + 32\gamma_k^2C^2 + 16\gamma_k^2\mathbb{E}[\|w_k\|^2 \mid \mathcal{F}_k] \\ \leq \|x_k - x^*\|^2 - 2\gamma_k\alpha\operatorname{dist}(x_k, X^*) - \frac{1}{16}\frac{\rho}{m\eta}\operatorname{dist}^2(y_k, X) + \frac{1}{4}\|x_k - x_{k-1}\|^2 + 2\gamma_k\alpha\|x_k - y_k\| \\ + 2\gamma_k(C + \alpha)\operatorname{dist}(y_k, X) + 64\gamma_k^2L^2\|x_k - x^*\|^2 + 64\gamma_k^2L^2\|x_k - x_{k-1}\|^2 + 32\gamma_k^2C^2 \\ + 16\gamma_k^2(\nu_1^2\|y_k\|^2 + \nu_2^2) \\ \leq \|x_k - x^*\|^2 - 2\gamma_k\alpha\operatorname{dist}(x_k, X^*) - \frac{1}{16}\frac{\rho}{m\eta}\operatorname{dist}^2(y_k, X) + \frac{1}{4}\|x_k - x_{k-1}\|^2 + 2\gamma_k\alpha\|x_k - y_k\| \\ + 2\gamma_k(C + \alpha)\operatorname{dist}(y_k, X) + 64\gamma_k^2L^2\|x_k - x^*\|^2 + 64\gamma_k^2L^2\|x_k - x_{k-1}\|^2 + 32\gamma_k^2C^2 \\ + 16\gamma_k^2(2\nu_1^2\|y_k - x_k\|^2 + 2\nu_1^2\|x_k\|^2 + \nu_2^2) \\ \leq \|x_k - x^*\|^2 - 2\gamma_k\alpha\operatorname{dist}(x_k, X^*) - \frac{1}{16}\frac{\rho}{m\eta}\operatorname{dist}^2(y_k, X) + \frac{1}{4}\|x_k - x_{k-1}\|^2 + 2\gamma_k\alpha\|x_k - y_k\| \\ \leq \|x_k - x^*\|^2 - 2\gamma_k\alpha\operatorname{dist}(x_k, X^*) - \frac{1}{16}\frac{\rho}{m\eta}\operatorname{dist}^2(y_k, X) + \frac{1}{4}\|x_k - x_{k-1}\|^2 + 2\gamma_k\alpha\|x_k - y_k\|$$

$$+2\gamma_{k}(C+\alpha)\operatorname{dist}(y_{k},X)+64\gamma_{k}^{2}L^{2}\|x_{k}-x^{*}\|^{2}+64\gamma_{k}^{2}L^{2}\|x_{k}-x_{k-1}\|^{2}+32\gamma_{k}^{2}C^{2}$$

$$+16\gamma_{k}^{2}(2\nu_{1}^{2}\|y_{k}-x_{k}\|^{2}+4\nu_{1}^{2}\|x_{k}-x^{*}\|^{2}+4\nu_{1}^{2}\|x^{*}\|^{2}+\nu_{2}^{2})$$

$$=\|x_{k}-x^{*}\|^{2}+\frac{3}{4}\|x_{k}-x_{k-1}\|^{2}-2\gamma_{k}\alpha\operatorname{dist}(x_{k},X^{*})-\frac{1}{2}\|x_{k}-x_{k-1}\|^{2}+2\gamma_{k}\alpha\|x_{k}-x_{k-1}\|$$

$$+2\gamma_{k}(C+\alpha)\operatorname{dist}(y_{k},X)-\frac{1}{16}\frac{\rho}{m\eta}\operatorname{dist}^{2}(y_{k},X)+64\gamma_{k}^{2}(L^{2}+\nu_{1}^{2})\|x_{k}-x^{*}\|^{2}$$

$$+32\gamma_{k}^{2}(2L^{2}+\nu_{1}^{2})\|x_{k}-x_{k-1}\|^{2}+32\gamma_{k}^{2}C^{2}+16\gamma_{k}^{2}(4\nu_{1}^{2}\|x^{*}\|^{2}+\nu_{2}^{2}). \tag{39}$$

Noting that $-\frac{1}{2}||x_k - x_{k-1}||^2 + 2\gamma_k\alpha||x_k - x_{k-1}|| = -\frac{1}{2}(||x_k - x_{k-1}|| - 2\gamma_k\alpha)^2 + 2\gamma_k^2\alpha^2$ and inserting it in (39), we have

$$\mathbb{E}[\|x_{k+1} - x^*\|^2 + \frac{3}{4}\|x_{k+1} - x_k\|^2 \mid \mathcal{F}_k] \leq \|x_k - x^*\|^2 + \frac{3}{4}\|x_k - x_{k-1}\|^2 - 2\gamma_k\alpha\operatorname{dist}(x_k, X^*) \\ - \frac{1}{2}(\|x_k - x_{k-1}\| - 2\gamma_k\alpha)^2 + 2\gamma_k^2\alpha^2 - \frac{\rho}{16m\eta}\left(\operatorname{dist}(y_k, X) - \frac{16m\eta\gamma_k(C+\alpha)}{\rho}\right)^2 + \frac{16m\eta(C+\alpha)^2}{\rho}\gamma_k^2 \\ + 64\gamma_k^2(L^2 + \nu_1^2)\|x_k - x^*\|^2 + 32\gamma_k^2(2L^2 + \nu_1^2)\|x_k - x_{k-1}\|^2 + 32\gamma_k^2C^2 + 16\gamma_k^2(4\nu_1^2\|x^*\|^2 + \nu_2^2) \\ \leq (1 + 86\gamma_k^2L^2 + 64\gamma_k^2\nu_1^2)\left(\|x_k - x^*\|^2 + \frac{3}{4}\|x_k - x_{k-1}\|^2\right) \\ - \underbrace{\left(\frac{1}{2}(\|x_k - x_{k-1}\| - 2\gamma_k\alpha)^2 + 2\gamma_k\alpha\operatorname{dist}(x_k, X^*) + \frac{\rho}{16m\eta}\left(\operatorname{dist}(y_k, X) - \frac{16m\eta\gamma_k(C+\alpha)}{\rho}\right)^2\right)}_{\beta_k} \\ + \underbrace{\left(2\gamma_k^2\alpha^2 + \frac{16m\eta(C+\alpha)^2}{\rho}\gamma_k^2 + 32\gamma_k^2C^2 + 16\gamma_k^2(4\nu_1^2\|x^*\|^2 + \nu_2^2)\right)}_{\eta_k}. \tag{40}$$

In effect, we obtain the following recursion:

$$\mathbb{E}[v_{k+1} \mid \mathcal{F}_k] \le (1 + u_k)v_k - \beta_k + \eta_k, \quad \text{a.s.}$$

where $v_k \triangleq \left(\|x_k - x^*\|^2 + \frac{3}{4}\|x_k - x_{k-1}\|^2\right)$ and $u_k = 86\gamma_k^2L^2 + 64\gamma_k^2\nu_1^2$. Since $\sum \gamma_k^2 < \infty$, it follows that u_k and η_k are summable. We may then invoke Lemma 2 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 a.s. 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 $\sum_k \frac{1}{2}\|x_k - x_{k-1} - 2\gamma_k\alpha\|^2 < \infty$ a.s. implying that $\|x_k - x_{k-1} - 2\gamma_k\alpha\| \to 0$ in a.s. sense. It follows that $\|y_k - x_k - 2\gamma_k\alpha\| \to 0$ almost surely. Since $\gamma_k \to 0$, it follows that $x_k - x_{k-1} \to 0$ in an a.s. sense. Thus $\{\|x_k - x^*\|\}$ is convergent in an a.s. sense. Consequently, $\{\|x_k - x^*\|\}$ is bounded a.s. and has a convergent subsequence a.s.. For any convergent subsequence denoted by \mathcal{K} , we have that $\{x_k\}_{k\in\mathcal{K}} \to \hat{x}(\omega)$. We proceed by contradiction and assume that $\hat{x}(\omega) \notin X^*$ with finite probability. Therefore, $\mathrm{dist}(x_k, X^*) \to h(\omega) > 0$ where $\omega \in V$ and $P(V) > 0 \Longrightarrow \sum_k \gamma_k \mathrm{dist}(x_k, X^*) = \infty$ with finite probability. But this implies that $\sum_{\omega \in V} \gamma_k \mathrm{dist}(x_k, X^*) \not< \infty$ a.s. \Longrightarrow contradiction. Thus every limit point of $\{x_k\}$ lies in X^* a.s. . Consider any such limit point $x_1^* \in X^*$. Then we have that $\{\|x_k - x_1^*\|\}$ is convergent to zero in an a.s. sense Since a subsequence of $\{\|x_k - x_1^*\|\}$ converges to zero a.s. and the entire sequence is convergent, the entire sequence converges to x_1^* a.s. .

Unlike in (**v-SPRG**), feasibility of the iterates $\{x_k\}$ cannot be maintained in (**r-SPRG**). This feasibility error arises because the random projection algorithms cannot guarantee feasibility of $\{x_k\}$. First we conduct almost-sure convergence analysis on the metric $\{\text{dist}(x_k, X)\}$ for both randomly generated algorithms and then derive the rate of convergence. To establish the rate of

convergence, we need the following lemma from [39].

Lemma 10. Suppose $\beta \in (0,1)$ and $R \geq 0$. Let $\{\delta_k\}$ and $\{\alpha_k\}$ be nonnegative sequences such that

$$\delta_{k+1} \le (1-\beta)\delta_k + R\alpha_k^2, \quad \forall k \ge 0.$$

If there exists $\bar{k} \geq 0$ such that $\alpha_{k+1}^2 \geq (1 - \frac{\beta}{2})\alpha_k^2$ for all $k \geq \bar{k}$, we have

$$\delta_k \le \frac{2R}{\beta} \alpha_k^2 + \delta_0 (1 - \beta)^k + \left(R \sum_{t=0}^{\bar{k}} \alpha_t^2 \right) (1 - \beta)^{k - \bar{k}}.$$

We also prove the following result that relates geometric rates to polynomial rates.

Lemma 11. Consider $\beta \in (0,1), d > 0, t \in \mathbb{Z}_+$, and $t \geq 1$. Then there exists a scalar $\bar{a} > 0$ such that

$$d\beta^k \le \frac{\bar{a}}{k^t}, \quad \forall k \ge 0$$

where \bar{a} is defined as

$$\bar{a} \triangleq \left(\frac{te}{\ln(1/\beta)}\right)^t$$
.

Proof. By definition of \bar{a} , we have that for all $k \geq 0$,

$$d\beta^k \le \frac{\bar{a}}{k^t} \text{ or } dk^t \beta^k \le \bar{a}.$$
 (41)

But (41) holds if

$$\bar{a} = \max_{z>0} dz^t \beta^z$$
.

If $h(z) \triangleq z^t \beta^z$, then any interior maximizer of $\max_{z \geq 0} h(z)$ satisfies

$$h'(z) = tz^{t-1}\beta^z + z^t\beta^z \ln(\beta) = 0 \implies t + z\ln(\beta) = 0 \text{ or } z^* = -\frac{t}{\ln(\beta)} = \frac{t}{\ln(1/\beta)}.$$

It is relatively simple to show that $h''(z^*) < 0$. Consequently,

$$\max_{z\geq 0} z^t \beta^z = h(z^*) = \left(\frac{t}{\ln(1/\beta)}\right)^t \beta^{-\frac{t}{\ln(\beta)}} = \left(\frac{t}{\ln(1/\beta)}\right)^t \left(\left(\frac{1}{\beta}\right)^{\frac{1}{\ln(\beta)}}\right)^t = \left(\frac{te}{\ln(1/\beta)}\right)^t.$$

Note that we utilize the relation that $(1/x)^{1/\ln(x)} = e$ in the last equality. As a result, we have $dk^t\beta^k \leq d\left(\frac{te}{\ln(1/\beta)}\right)^t = d\bar{a}^t$.

Theorem 4. Let Assumptions 1-2, 4-6 hold. Suppose $\{x_k\}$ is generated by (**r-SPRG**), where the projections are randomly generated. In addition, suppose there exists a $D_X > 0$ such that $||u-v||^2 \le D_X^2$ for any $u, v \in X$. Then the following hold.

- (a) If $\sum_{k} \gamma_{k} = \infty$ and $\sum_{k} \gamma_{k}^{2} < \infty$, then $\operatorname{dist}(x_{k}, X) \xrightarrow{k \to \infty} 0$.
- (b) Suppose $\gamma_k = 1/k^{t/2}$ where $t \geq 1$. Then $\mathbb{E}[\operatorname{dist}(x_k, X)] \leq \mathcal{O}\left(\frac{1}{k^{t/2}}\right)$ for any $k \geq \bar{k}$, where

$$\bar{k} \triangleq \left\lceil \left(\frac{1}{1 - (1 - \beta/2)^{1/t}} - 1 \right) \right\rceil.$$

(c) Suppose
$$\gamma_k = 1/k^{1/2}$$
. Suppose $\bar{x}_{K,\bar{k}} \triangleq \frac{\sum_{k=\bar{k}+\lfloor K/2\rfloor}^{\bar{k}+K} \gamma_k x_k}{\sum_{k=\bar{k}+\lfloor K/2\rfloor}^{\bar{k}+K} \gamma_k}$. Then $\mathbb{E}[\mathrm{dist}(\bar{x}_{K,\bar{k}},X)] \leq \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$.

Proof. (a) Let $z_k = x_k - \gamma_k F(2x_k - x_{k-1}, \omega_k)$. From Lemma 1, we have

$$\operatorname{dist}^{2}(x_{k+1}, X) \leq \|x_{k+1} - \Pi_{X}(z_{k})\|^{2} = \|\Pi_{l_{k}}(z_{k}) - \Pi_{X}(z_{k})\|^{2}$$
$$\leq \|z_{k} - \Pi_{X}(z_{k})\|^{2} - \|\Pi_{l_{k}}(z_{k}) - z_{k}\|^{2}. \tag{42}$$

Choose $\theta \ge \max\left\{1, \frac{3\rho}{64m\eta}\right\}$. By $||a+b||^2 \le \left(1 + \frac{4\theta m\eta}{\rho}\right) ||a||^2 + \left(1 + \frac{\rho}{4\theta m\eta}\right) ||b||^2$, we obtain

$$||z_{k} - \Pi_{X}(z_{k})||^{2} \leq ||z_{k} - \Pi_{X}(x_{k})||^{2} = ||z_{k} - x_{k} + x_{k} - \Pi_{X}(x_{k})||^{2}$$

$$\leq \left(1 + \frac{4\theta m\eta}{\rho}\right) ||z_{k} - x_{k}||^{2} + \left(1 + \frac{\rho}{4\theta m\eta}\right) ||x_{k} - \Pi_{X}(x_{k})||^{2}.$$

$$(43)$$

Combining (42) and (43), we obtain that

$$\operatorname{dist}^{2}(x_{k+1}, X) \leq \left(1 + \frac{4\theta m \eta}{\rho}\right) \|z_{k} - x_{k}\|^{2} + \left(1 + \frac{\rho}{4\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) - \|\Pi_{l_{k}}(z_{k}) - z_{k}\|^{2}. \tag{44}$$

From Lemmas 6 and 9, we have

$$\mathbb{E}[\|z_{k} - \Pi_{l_{k}}(z_{k})\|^{2} \mid \mathcal{F}_{k}] \ge \frac{\rho}{m\eta} \operatorname{dist}^{2}(z_{k}, X) \ge \frac{\rho}{\theta m\eta} \left(\frac{1}{2} \operatorname{dist}^{2}(x_{k}, X) - 4\|z_{k} - x_{k}\|^{2}\right)$$

$$\ge \frac{\rho}{2\theta m\eta} \operatorname{dist}^{2}(x_{k}, X) - \frac{4\rho}{\theta m\eta} \|z_{k} - x_{k}\|^{2} \ge \frac{\rho}{2\theta m\eta} \operatorname{dist}^{2}(x_{k}, X) - 4\|z_{k} - x_{k}\|^{2}, \tag{45}$$

where the last inequality follows from $\frac{\rho}{\theta m \eta} \leq 1$. Substituting (45) into (44), and taking conditional expectations, it follows that

$$\mathbb{E}[\operatorname{dist}^{2}(x_{k+1}, X) \mid \mathcal{F}_{k}] \leq \left(1 - \frac{\rho}{4\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) + \left(5 + \frac{4\theta m \eta}{\rho}\right) \|z_{k} - x_{k}\|^{2} \\
= \left(1 - \frac{\rho}{4\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) + \left(5 + \frac{4\theta m \eta}{\rho}\right) \|\gamma_{k} \mathbb{E}[F(2x_{k} - x_{k-1}, \omega_{k}) \mid \mathcal{F}_{k}]\|^{2} \\
\leq \left(1 - \frac{\rho}{4\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) + \left(5 + \frac{4\theta m \eta}{\rho}\right) \gamma_{k}^{2} (2\|F(2x_{k} - x_{k-1})\|^{2} + 2\mathbb{E}[\|w_{k}\|^{2} \mid \mathcal{F}_{k}]) \\
\leq \left(1 - \frac{\rho}{4\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) + \left(5 + \frac{4\theta m \eta}{\rho}\right) \gamma_{k}^{2} (4L^{2}\|2x_{k} - x_{k-1} - x^{*}\|^{2} \\
+ 4\|F(x^{*})\|^{2} + 2\nu_{1}^{2}\|y_{k}\|^{2} + 2\nu_{2}^{2}) \\
\leq \left(1 - \frac{\rho}{4\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) + \left(5 + \frac{4\theta m \eta}{\rho}\right) \gamma_{k}^{2} (4L^{2}\|2x_{k} - x_{k-1} - x^{*}\|^{2} \\
+ 4\|F(x^{*})\|^{2} + 4\nu_{1}^{2}\|2x_{k} - x_{k-1} - x^{*}\|^{2} + 4\nu_{1}^{2}\|x^{*}\|^{2} + 2\nu_{2}^{2}) \\
\leq \left(1 - \frac{\rho}{4\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) + \left(5 + \frac{4\theta m \eta}{\rho}\right) \gamma_{k}^{2} (32(L^{2} + \nu_{1}^{2})\|x_{k} - x^{*}\|^{2} \\
+ 8(L^{2} + \nu_{1}^{2})\|x_{k-1} - x^{*}\|^{2} + 4\|F(x^{*})\|^{2} + 4\nu_{1}^{2}\|x^{*}\|^{2} + 2\nu_{2}^{2}), \tag{47}$$

where $x^* \in X^*$. We now use the inequality

$$||x_k - x^*||^2 \le 2||x_k - \Pi_X(x_k)||^2 + 2||\Pi_X(x_k) - x^*||^2 \le 2\operatorname{dist}^2(x_k, X) + 2D_X^2.$$
(48)

Therefore, we have that from (47),

$$\mathbb{E}[\operatorname{dist}^{2}(x_{k+1}, X) \mid \mathcal{F}_{k}] \leq \left(1 - \frac{\rho}{4\theta m\eta}\right) \operatorname{dist}^{2}(x_{k}, X)$$

$$+ \left(5 + \frac{4\theta m\eta}{\rho}\right) \left(80(L^2 + \nu_1^2)D_X^2 + 4\nu_1^2 \|x^*\|^2 + 4C^2 + 2\nu_2^2\right)\gamma_k^2$$

$$+ 64\left(5 + \frac{4\theta m\eta}{\rho}\right)\gamma_k^2(L^2 + \nu_1^2)\mathrm{dist}^2(x_k, X) + 16\left(5 + \frac{4\theta m\eta}{\rho}\right)\gamma_k^2(L^2 + \nu_1^2)\mathrm{dist}^2(x_{k-1}, X).$$

Suppose $\gamma_0^2 \leq \frac{\rho}{8\theta m\eta}/64 \left(5 + \frac{4\theta m\eta}{\rho}\right) (L^2 + \nu_1^2)$. Since $\{\gamma_k\}$ is a diminishing sequence, it holds that $64 \left(5 + \frac{4\theta m\eta}{\rho}\right) \gamma_k^2 (L^2 + \nu_1^2) \leq \frac{\rho}{8\theta m\eta}$. Then we have

$$\mathbb{E}[\operatorname{dist}^{2}(x_{k+1}, X) \mid \mathcal{F}_{k}] \leq \left(1 - \frac{\rho}{8\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) + \frac{\rho}{32\theta m \eta} \operatorname{dist}^{2}(x_{k-1}, X) + \left(5 + \frac{4\theta m \eta}{\rho}\right) (80(L^{2} + \nu_{1}^{2})D_{X}^{2} + 4\nu_{1}^{2} \|x^{*}\|^{2} + 4C^{2} + 2\nu_{2}^{2})\gamma_{k}^{2}.$$

Since $\frac{\rho}{32\theta m\eta} \leq \frac{2}{3}$, we have $\frac{\rho}{32\theta m\eta} \leq \left(1 - \frac{\rho}{32\theta m\eta}\right) \frac{3\rho}{32\theta m\eta}$. It follows that

$$\mathbb{E}[\operatorname{dist}^{2}(x_{k+1}, X) + \frac{3\rho}{32\theta m\eta} \operatorname{dist}^{2}(x_{k}, X) \mid \mathcal{F}_{k}] \leq \left(1 - \frac{\rho}{32\theta m\eta}\right) \left(\operatorname{dist}^{2}(x_{k}, X) + \frac{3\rho}{32\theta m\eta} \operatorname{dist}^{2}(x_{k-1}, X)\right) + \left(5 + \frac{4\theta m\eta}{\rho}\right) \left(80(L^{2} + \nu_{1}^{2})D_{X}^{2} + 4\nu_{1}^{2}\|x^{*}\|^{2} + 4C^{2} + 2\nu_{2}^{2}\right)\gamma_{k}^{2}.$$

$$(49)$$

We may now invoke Lemma 3 and the summability of γ_k^2 to claim $\operatorname{dist}^2(x_k, X) \xrightarrow{k \to \infty} 0$.

(b) We begin by noting that $\gamma_{k+1}^2 \geq \left(1 - \frac{\rho}{32\theta m\eta}\right) \gamma_k^2$ when $k \geq \bar{k}$, where \bar{k} is obtained as follows when $\gamma_k = 1/k^{t/2}$.

$$\frac{1}{(k+1)^t} \ge \left(1 - \frac{\beta}{2}\right) \frac{1}{k^t} \text{ or } \frac{k}{k+1} \ge \left(1 - \frac{\beta}{2}\right)^{1/t}$$
$$\frac{1}{k+1} \le 1 - \left(1 - \frac{\beta}{2}\right)^{1/t} \text{ or } k \ge \bar{k} \triangleq \left[\left(\frac{1}{1 - (1 - \beta/2)^{1/t}} - 1\right)\right].$$

Taking unconditional expectations on (49), recalling $\gamma_{k+1}^2 \ge \left(1 - \frac{\rho}{32\theta m\eta}\right) \gamma_k^2$ when $k \ge \bar{k}$, and by leveraging Lemma 10, we have

$$\mathbb{E}[\operatorname{dist}^{2}(x_{k}, X)] \leq \underbrace{\left(\frac{320\theta m\eta}{\rho} + \frac{256\theta^{2}m^{2}\eta^{2}}{\rho^{2}}\right) \left(80(L^{2} + \nu_{1}^{2})D_{X}^{2} + 4\nu_{1}^{2}\|x^{*}\|^{2} + 4C^{2} + 2\nu_{2}^{2}\right)}_{\triangleq a} \gamma_{k}^{2} + \underbrace{\left(\operatorname{dist}^{2}(x_{0}, X) + \frac{3\rho}{32\theta m\eta}\operatorname{dist}^{2}(x_{-1}, X)\right) \left(1 - \frac{\rho}{32\theta m\eta}\right)^{k}}_{\triangleq b} + \underbrace{\left(\left(5 + \frac{4\theta m\eta}{\rho}\right) \left(80(L^{2} + \nu_{1}^{2})D_{X}^{2} + 4\nu_{1}^{2}\|x^{*}\|^{2} + 4C^{2} + 2\nu_{2}^{2}\right)\sum_{t=0}^{\bar{k}} \gamma_{t}^{2}\right) \left(\underbrace{1 - \frac{\rho}{32\theta m\eta}}_{\triangleq \beta}\right)^{k - \bar{k}}}_{\triangleq c} \leq a\gamma_{k}^{2} + b\beta^{k} + c\beta^{k - \bar{k}},$$

$$(50)$$

Consequently, for $k \geq \bar{k}$, (50) reduces to

$$\mathbb{E}[\operatorname{dist}^2(x_k, X)] \leq \frac{a}{k^t} + d\beta^k$$
, where $d \triangleq (b + c\beta^{-\bar{k}})$.

From Lemma 11, we have that for $k \geq \bar{k}$,

$$\mathbb{E}[\operatorname{dist}^{2}(x_{k}, X)] \leq \frac{a + \bar{a}}{k^{t}}, \text{ where } \bar{a} \triangleq \left(\frac{te}{\ln(1/\beta)}\right)^{t}$$
(51)

By Jensen's inequality, we have that

$$\mathbb{E}[\operatorname{dist}(x_k, X)] \le \sqrt{\mathbb{E}[\operatorname{dist}^2(x_k, X)]} \le \frac{\sqrt{a + \bar{a}}}{k^{t/2}}.$$

(c) Suppose $\bar{x}_{K,\bar{k}} \triangleq \frac{\sum_{k=\bar{k}+\lfloor K/2\rfloor}^{K+k} \gamma_k x_k}{\sum_{k=\bar{k}+\lfloor K/2\rfloor}^{K+k} \gamma_k}$ denotes the window-based weighted average from $\bar{k} + \lfloor K/2 \rfloor$ to $\bar{k} + K$. Then we have the following.

$$\mathbb{E}[\operatorname{dist}(\bar{x}_{K,\bar{k}},X)] \leq \frac{\sum_{k=\bar{k}+\lfloor K/2\rfloor}^{K+\bar{k}} \gamma_k \mathbb{E}[\operatorname{dist}(x_k,X)]}{\sum_{k=\bar{k}+\lfloor K/2\rfloor}^{K+\bar{k}} \gamma_k} \leq \frac{\sum_{k=\bar{k}+\lfloor K/2\rfloor}^{K+\bar{k}} \frac{a+\bar{a}}{k}}{\sum_{k=\bar{k}+\lceil K/2\rfloor}^{K+\bar{k}} \frac{1}{\sqrt{k}}} \leq \frac{(a+\bar{a})(\ln(2)+1)}{\frac{\sqrt{K}}{2\sqrt{k}+2}} \leq \mathcal{O}\left(\frac{1}{\sqrt{K}}\right). \quad (52)$$

We now provide a rate of convergence for the iterates in terms of the gap function expressed at a projection of the averaged sequence. Note the difference between this result and that in the previous Section where $\bar{x}_{K,\bar{k}}$ is feasible; here, the lack of feasibility requires utilizing the average of the projection $\Pi_X(\bar{y}_k)$ instead of the standard weighted average. Recall that the previous result derives a rate statement for the infeasibility. In addition, we also derive an and oracle complexity statement for ensuring that the condition

$$\mathbb{E}[G(\bar{y}_{K,\bar{k}}))] \leq \mathcal{O}\left(\frac{1}{\sqrt{K}}\right), \text{ where } \bar{y}_{K,\bar{k}} = \frac{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \gamma_k \Pi_X(y_k)}{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \gamma_k}.$$

Proposition 3. Let Assumptions 1-2 and 4-6 hold. Let $\gamma_k = \frac{1}{\sqrt{k}}$. In addition, for any $u, v \in X$, suppose that there exists a $D_X > 0$ such that $||u-v||^2 \le D_X^2$. Then the following holds for any sequence generated by (**r-SPRG**) in an expected value sense, where $\bar{y}_{K,\bar{k}} = \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_k \Pi_X(y_k)}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_k}$.

- (a) $\mathbb{E}[G(\bar{y}_{K,\bar{k}})] \leq \mathcal{O}\left(\frac{1}{\sqrt{K}}\right);$
- (b) The oracle complexity to compute an $\bar{y}_{K,\bar{k}}$ such that $\mathbb{E}[G(\bar{y}_{K,\bar{k}})] \leq \epsilon$ is bounded by $\mathcal{O}\left(\frac{1}{\epsilon^2}\right)$.

Proof. (a) Invoking the analysis of Lemma 7, without invoking (30), we obtain

$$F(x)^{T}(x-x^{*}) \ge F(x^{*})^{T}(\Pi_{X}(x)-x^{*}) - C\operatorname{dist}(x,X), \quad \forall x \in \mathbb{R}^{n}.$$
(53)

Using this property in (34) and rewriting it with a similar manner as (36), we have

$$||x_{k+1} - x^*||^2 \le ||x_k - x^*||^2 - \frac{3}{4}||x_{k+1} - x_k||^2 - \frac{1}{16}||y_k - x_{k+1}||^2 - \frac{1}{8}||x_{k-1} - x_{k+1}||^2 + \frac{1}{4}||x_k - x_{k-1}||^2 - 2\gamma_k F(x^*)^T (\Pi_X(y_k) - x^*) + 2\gamma_k C \operatorname{dist}(y_k, X) + 32\gamma_k^2 L^2 ||y_k - x^*||^2$$

$$+32\gamma_k^2C^2+16\gamma_k^2||w_k||^2-2\gamma_kw_k^T(y_k-x^*).$$

It follows that

$$\begin{aligned} &\|x_{k+1} - x^*\|^2 + \frac{3}{4} \|x_{k+1} - x_k\|^2 \le \|x_k - x^*\|^2 - 2\gamma_k F(x^*)^T (\Pi_X(y_k) - x^*) \\ &- \frac{1}{16} \|y_k - x_{k+1}\|^2 - \frac{1}{8} \|x_{k-1} - x_{k+1}\|^2 + \frac{1}{4} \|x_k - x_{k-1}\|^2 + 2\gamma_k C \mathrm{dist}(y_k, X) \\ &+ 64\gamma_k^2 L^2 \|x_k - x^*\|^2 + 64\gamma_k^2 L^2 \|x_k - x_{k-1}\|^2 + 32\gamma_k^2 C^2 + 16\gamma_k^2 \|w_k\|^2 - 2\gamma_k w_k^T (y_k - x^*) \\ &\le \|x_k - x^*\|^2 + \frac{3}{4} \|x_k - x_{k-1}\|^2 - 2\gamma_k F(x^*)^T (\Pi_X(y_k) - x^*) - \frac{1}{2} \|x_k - x_{k-1}\|^2 \\ &+ 2\gamma_k C \mathrm{dist}(y_k, X) - \frac{1}{16} \|y_k - x_{k+1}\|^2 + 64\gamma_k^2 L^2 \|x_k - x^*\|^2 + 64\gamma_k^2 L^2 \|x_k - x_{k-1}\|^2 \\ &+ 32\gamma_k^2 C^2 + 16\gamma_k^2 \|w_k\|^2 - 2\gamma_k w_k^T (y_k - x^*). \end{aligned}$$

We have the following inequality by replacing x^* with y:

$$2\gamma_{k}F(y)^{T}(\Pi_{X}(y_{k})-y) \leq \|x_{k}-y\|^{2} + \frac{3}{4}\|x_{k}-x_{k-1}\|^{2} - (\|x_{k+1}-y\|^{2} + \frac{3}{4}\|x_{k+1}-x_{k}\|^{2})$$

$$+ 2\gamma_{k}C\operatorname{dist}(y_{k},X) - \frac{1}{16}\|y_{k}-x_{k+1}\|^{2} + 64\gamma_{k}^{2}L^{2}\|x_{k}-y\|^{2} + 64\gamma_{k}^{2}L^{2}\|x_{k}-x_{k-1}\|^{2}$$

$$+ 32\gamma_{k}^{2}C^{2} + 16\gamma_{k}^{2}\|w_{k}\|^{2} - 2\gamma_{k}w_{k}^{T}(y_{k}-y).$$

$$(54)$$

We now define an auxiliary sequence $\{u_k\}$ such that

$$u_{k+1} := \Pi_X(u_k - \gamma_k w_k),$$

where $u_0 \in X$. We may then express the last term on the right in (54) as follows.

$$2\gamma_k w_k^T(y - y_k) = 2\gamma_k w_k^T(y - u_k) + 2\gamma_k w_k^T(u_k - y_k)$$

$$\leq ||u_k - y||^2 - ||u_{k+1} - y||^2 + \gamma_k^2 ||w_k||^2 + 2\gamma_k w_k^T(u_k - y_k).$$
(55)

Summing over k and invoking (55), we obtain the following bound:

$$\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} 2\gamma_k F(y)^T (\Pi_X(y_k) - y) \le ||x_0 - y||^2 + \frac{3}{4} ||x_0 - x_{-1}||^2
+ \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} (2\gamma_k C \operatorname{dist}(y_k, X) - \frac{1}{16} ||y_k - x_{k+1}||^2)
+ \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} (64L^2 ||x_k - y||^2 + 64L^2 ||x_k - x_{k+1}||^2 + 32C^2) \gamma_k^2 + \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} 17\gamma_k^2 ||w_k||^2
+ ||u_0 - y||^2 + \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} 2\gamma_k w_k^T (u_k - y_k).$$
(56)

Dividing both sides of (56) by $\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \gamma_k$, we have

$$2F(y)^{T}(\bar{y}_{K,\bar{k}} - y) \leq \frac{\|x_{0} - y\|^{2} + \frac{3}{4}\|x_{0} - x_{-1}\|^{2} + \|u_{0} - y\|^{2}}{\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \left(2\gamma_{k}C\operatorname{dist}(y_{k}, X) - \frac{1}{16}\|y_{k} - x_{k+1}\|^{2}\right)}{\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \gamma_{k}}$$

$$+\frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}(64L^2||x_k-y||^2+64L^2||x_k-x_{k-1}||^2+32C^2)\gamma_k^2}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_k}+\frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}17\gamma_k^2||w_k||^2+\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}2\gamma_kw_k^T(u_k-y_k)}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_k}\\ \leq \frac{B_1}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_k}+\frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\left(2\gamma_kC\mathrm{dist}(y_k,X)-\frac{1}{16}||y_k-x_{k+1}||^2\right)}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_k}\\ +\frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}(64L^2||x_k-y||^2+64L^2||x_k-x_{k-1}||^2+32C^2)\gamma_k^2}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}}+\frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}2\gamma_kw_k^T(u_k-y_k)}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_k}\\ \leq \frac{B_1}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_k}+\frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\left(2\gamma_kC\mathrm{dist}(y_k,X)-\frac{1}{16}||y_k-x_{k+1}||^2\right)}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_k}\\ +\frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}(128L^2(\mathrm{dist}^2(x_k,X)+D_X^2)+64L^2||x_k-x_{k-1}||^2+32C^2)\gamma_k^2}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}}17\gamma_k^2||w_k||^2+\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}2\gamma_kw_k^T(u_k-y_k)}\\ \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_k},\\ +\frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}17\gamma_k^2||w_k||^2+\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}2\gamma_kw_k^T(u_k-y_k)}{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_k}},$$

where $||x_0 - y||^2 + \frac{3}{4}||x_0 - x_{-1}||^2 + ||u_0 - y||^2 \le 2||x_0 - x^*||^2 + 2||x^* - y||^2 + \frac{3}{4}||x_0 - x_{-1}||^2 + 2||u_0 - x^*||^2 + 2||x^* - y||^2 \le 2||x_0 - x^*||^2 + \frac{3}{4}||x_0 - x_{-1}||^2 + 2||u_0 - x^*||^2 + 2D_X^2 \triangleq B_1$. By taking supremum over $y \in X$, we obtain the following inequality:

$$G(\bar{y}_{K,\bar{k}}) \stackrel{\triangle}{=} \sup_{y \in X} F(y)^{T} (\bar{y}_{K,\bar{k}} - y) \leq \frac{B_{1}}{2\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \left(2\gamma_{k}C\operatorname{dist}(y_{k}, X) - \frac{1}{16}\|y_{k} - x_{k+1}\|^{2}\right)}{2\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} (128L^{2}(\operatorname{dist}^{2}(x_{k}, X) + D_{X}^{2}) + 64L^{2}\|x_{k} - x_{k-1}\|^{2} + 32C^{2})\gamma_{k}^{2}}{2\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} 17\gamma_{k}^{2}\|w_{k}\|^{2} + \sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} 2\gamma_{k}w_{k}^{T}(u_{k} - y_{k})}{2\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \gamma_{k}}.$$

$$(57)$$

Taking unconditional expectation and using (38) in (57), we have

$$\begin{split} \mathbb{E}[G(\bar{y}_{K,\bar{k}})] &\leq \frac{B_1}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} + \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \left(2\gamma_k C \mathbb{E}[\operatorname{dist}(y_k,X)] - \frac{\rho}{16m\eta} \mathbb{E}[\operatorname{dist}^2(y_k,X)]\right)}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k } \\ &+ \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} (128L^2(\mathbb{E}[\operatorname{dist}^2(x_k,X)] + D_X^2) + 64L^2 \mathbb{E}[\|x_k - x_{k-1}\|^2] + 32C^2)\gamma_k^2}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k } \\ &+ \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} (17\gamma_k^2(32\nu_1^2 \mathbb{E}[\operatorname{dist}^2(x_k,X)] + 8\nu_1^2 \mathbb{E}[\operatorname{dist}^2(x_{k-1},X)] + 40\nu_1^2 D_X^2 + 2\nu_1^2 \|x^*\|^2 + \nu_2^2))}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k } \\ &\leq \frac{B_1}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} + \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \left(\frac{16m\eta C^2}{\rho} \gamma_k^2\right)}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k } \\ &+ \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} (128L^2(\mathbb{E}[\operatorname{dist}^2(x_k,X)] + D_X^2) + 64L^2 \mathbb{E}[\|x_k - x_{k-1}\|^2] + 32C^2)\gamma_k^2}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k } \\ &+ \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} (17\gamma_k^2(32\nu_1^2 \mathbb{E}[\operatorname{dist}^2(x_k,X)] + 8\nu_1^2 \mathbb{E}[\operatorname{dist}^2(x_{k-1},X)] + 40\nu_1^2 D_X^2 + 2\nu_1^2 \|x^*\|^2 + \nu_2^2))}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k } \\ &\stackrel{\text{From } (51)}{\leq \frac{B_1}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k} + \frac{16m\eta C^2}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k + \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} (512L^2((a+\bar{a})/\bar{k}) + 320D_X^2 + 32C^2)\gamma_k^2}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k + \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_k}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k} \\ &= \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_k}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k} + \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_k}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k} + \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_k}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k} \\ &= \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_k}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k} + \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_k}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k} \\ &= \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_k}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k} \gamma_k}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}} \gamma_k} \\ &= \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{$$

$$+\frac{\left(17(40\nu_{1}^{2}(\bar{b}/\bar{k})+40\nu_{1}^{2}D_{X}^{2}+2\nu_{1}^{2}\|x^{*}\|^{2}+\nu_{2}^{2})\right)\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_{k}^{2}}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}}\gamma_{k}}.$$
(58)

We now leverage the following lower bound on the denominator for $K \geq 1$:

$$\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} k^{-\frac{1}{2}} \ge \int_{K/2+\bar{k}}^{K+\bar{k}} (x+1)^{-\frac{1}{2}} dx = 2\sqrt{K+\bar{k}+1} - 2\sqrt{K/2+\bar{k}+1} \ge \frac{\sqrt{K}}{(2\sqrt{\bar{k}+2})}.$$
 (59)

Similarly an upper bound may be constructed:

$$\sum_{k=|K/2|+\bar{k}}^{K+\bar{k}} \gamma_k^2 = \sum_{k=|K/2|+\bar{k}}^{K+\bar{k}} k^{-1} \le \int_{K/2+\bar{k}}^{K+\bar{k}} x^{-1} dx + \frac{1}{\lfloor K/2\rfloor+\bar{k}} \le \log 2 + 1.$$
 (60)

By substituting (59) and (60) in (58), we obtain that the following holds:

$$\mathbb{E}[G(\bar{y}_{K,\bar{k}})] \le \mathcal{O}\left(\frac{1}{\sqrt{K}}\right).$$

(b) From (a), we know that $K_{\epsilon} = \mathcal{O}(1/\epsilon^2)$ to ensure that $\mathbb{E}[G(\bar{y}_{K,\bar{k}})] \leq \mathcal{O}\left(\frac{1}{\sqrt{K}}\right) \leq \epsilon$. It follows that

$$\sum_{k=1}^{K_{\epsilon}} 1 = K_{\epsilon} = \mathcal{O}\left(\frac{1}{\epsilon^2}\right).$$

4.3 SSE with random projections

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

Proposition 4. Let Assumptions 1 – 6 hold. Suppose the steplength sequence $\{\gamma_k\}$ satisfies $\sum_{k=0}^{\infty} \gamma_k = \infty$, $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$, and $\gamma_k \leq \frac{1}{2\sqrt{L^2+2\nu_1^2}}$. Then any sequence generated by (**r-SSE**) converges to a solution $x^* \in X$ in an a.s. sense.

Proof. By Lemma 1(ii), we have

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

$$= ||x_k - x^*||^2 - ||x_k - x_{k+1}||^2 + 2\gamma_k(F(x_{k+\frac{1}{2}}) + w_{k+\frac{1}{2}})^T(x^* - x_{k+1}). \tag{61}$$

It is clear that

$$F(x_{k+\frac{1}{2}})^{T}(x_{k+1} - x^{*}) = F(x_{k+\frac{1}{2}})^{T}(x_{k+1} - x_{k+\frac{1}{2}}) + F(x_{k+\frac{1}{2}})^{T}(x_{k+\frac{1}{2}} - x^{*}).$$
 (62)

Using (62) in (61), we obtain

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

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

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

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

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

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

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

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

$$(63)$$

Employing a similar approach as in Proposition 1, we obtain that

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2) \|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|w_k - w_{k+\frac{1}{2}}\|^2 \\ &+ 2\gamma_k w_{k+\frac{1}{2}}^T (x^* - x_{k+\frac{1}{2}}) - 2\gamma_k F(x_{k+\frac{1}{2}})^T (x_{k+\frac{1}{2}} - x^*) \\ &\leq \|x_k - x^*\|^2 - (1 - 2\gamma_k^2 L^2) \|x_k - x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2 \|w_k - w_{k+\frac{1}{2}}\|^2 \\ &+ 2\gamma_k w_{k+\frac{1}{2}}^T (x^* - x_{k+\frac{1}{2}}) - 2\gamma_k \alpha \text{dist} \left(\Pi_X(x_{k+\frac{1}{2}}), X^*\right) + 2\gamma_k C \text{dist}(x_{k+\frac{1}{2}}, X). \end{aligned}$$
(64)

Invoking weak sharpness property, we have

$$-2\gamma_k \alpha \operatorname{dist}\left(\Pi_X(x_{k+\frac{1}{2}}), X^*\right) \le -2\gamma_k \alpha \operatorname{dist}\left(x_k, X^*\right) + 2\gamma_k \alpha \|x_k - x_{k+\frac{1}{2}}\| + 2\gamma_k \alpha \operatorname{dist}\left(x_{k+\frac{1}{2}}, X\right)$$
(65)

and

$$2\gamma_{k}(C+\alpha)\operatorname{dist}(x_{k+\frac{1}{2}},X) \leq 2\gamma_{k}(C+\alpha)\operatorname{dist}(x_{k},X) + 2\gamma_{k}(C+\alpha)\|x_{k} - x_{k+\frac{1}{2}}\|$$

$$\leq 2\gamma_{k}(C+\alpha)\operatorname{dist}(x_{k},X) + 4\gamma_{k}^{2}(C+\alpha)^{2} + \frac{1}{4}\|x_{k} - x_{k+\frac{1}{2}}\|^{2}, \tag{66}$$

Using (65) and (66) in (64), we obtain

$$\begin{split} &\|x_{k+1}-x^*\|^2 \leq \|x_k-x^*\|^2 - (1-2\gamma_k^2L^2)\|x_k-x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2\|w_k-w_{k+\frac{1}{2}}\|^2 \\ &+ 2\gamma_k w_{k+\frac{1}{2}}^T (x^*-x_{k+\frac{1}{2}}) - 2\gamma_k \alpha \mathrm{dist} \left(\Pi_X(x_{k+\frac{1}{2}}), X^*\right) + 2\gamma_k C \mathrm{dist}(x_{k+\frac{1}{2}}, X) \\ &\leq \|x_k-x^*\|^2 - (1-2\gamma_k^2L^2)\|x_k-x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2\|w_k-w_{k+\frac{1}{2}}\|^2 \\ &+ 2\gamma_k w_{k+\frac{1}{2}}^T (x^*-x_{k+\frac{1}{2}}) - 2\gamma_k \alpha \mathrm{dist} \left(x_k, X^*\right) + 2\gamma_k \alpha \|x_k-x_{k+\frac{1}{2}}\| + 2\gamma_k (C+\alpha) \mathrm{dist}(x_{k+\frac{1}{2}}, X) \\ &\leq \|x_k-x^*\|^2 - (1-2\gamma_k^2L^2)\|x_k-x_{k+\frac{1}{2}}\|^2 + 2\gamma_k^2\|w_k-w_{k+\frac{1}{2}}\|^2 + 2\gamma_k w_{k+\frac{1}{2}}^T (x^*-x_{k+\frac{1}{2}}) \\ &- 2\gamma_k \alpha \mathrm{dist} \left(x_k, X^*\right) + 2\gamma_k \alpha \|x_k-x_{k+\frac{1}{2}}\| + 2\gamma_k (C+\alpha) \mathrm{dist}(x_k, X) + 4\gamma_k^2 (C+\alpha)^2 + \frac{1}{4}\|x_k-x_{k+\frac{1}{2}}\|^2 \\ &\leq \|x_k-x^*\|^2 - 2\gamma_k \alpha \mathrm{dist} \left(x_k, X^*\right) - \left(\frac{5}{8} - 2\gamma_k^2L^2\right)\|x_k-x_{k+\frac{1}{2}}\|^2 - \frac{1}{8}(\|x_k-x_{k+\frac{1}{2}}\| - 8\gamma_k \alpha)^2 \\ &+ 8\gamma_k^2 \alpha^2 + 4\gamma_k^2 (C+\alpha)^2 + 2\gamma_k (C+\alpha) \mathrm{dist}(x_k, X) + 2\gamma_k^2 \|w_{k+\frac{1}{2}} - w_k\|^2 - 2\gamma_k w_{k+\frac{1}{2}}^T \left(x_{k+\frac{1}{2}} - x^*\right). \end{split}$$

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

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

$$+8\gamma_{k}^{2}\alpha^{2}+4\gamma_{k}^{2}(C+\alpha)^{2}+2\gamma_{k}(C+\alpha)\operatorname{dist}(x_{k},X)+2\gamma_{k}^{2}(\nu_{1}^{2}\mathbb{E}[\|x_{k+\frac{1}{2}}\|^{2}\mid\mathcal{F}_{k}]+\nu_{1}^{2}\|x_{k}\|^{2}+2\nu_{2}^{2})$$

$$\leq\|x_{k}-x^{*}\|^{2}-2\gamma_{k}\alpha\operatorname{dist}(x_{k},X^{*})-\left(\frac{5}{8}-2\gamma_{k}^{2}L^{2}\right)\mathbb{E}[\|x_{k}-x_{k+\frac{1}{2}}\|^{2}\mid\mathcal{F}_{k}]$$

$$+8\gamma_{k}^{2}\alpha^{2}+4\gamma_{k}^{2}(C+\alpha)^{2}+2\gamma_{k}(C+\alpha)\operatorname{dist}(x_{k},X)$$

$$+2\gamma_{k}^{2}(2\nu_{1}^{2}\mathbb{E}[\|x_{k}-x_{k+\frac{1}{2}}\|^{2}\mid\mathcal{F}_{k}]+3\nu_{1}^{2}\|x_{k}\|^{2}+2\nu_{2}^{2})$$

$$\leq\|x_{k}-x^{*}\|^{2}-2\gamma_{k}\alpha\operatorname{dist}(x_{k},X^{*})-\left(\frac{5}{8}-2\gamma_{k}^{2}L^{2}\right)\mathbb{E}[\|x_{k}-x_{k+\frac{1}{2}}\|^{2}\mid\mathcal{F}_{k}]$$

$$+8\gamma_{k}^{2}\alpha^{2}+4\gamma_{k}^{2}(C+\alpha)^{2}+2\gamma_{k}(C+\alpha)\operatorname{dist}(x_{k},X)$$

$$+2\gamma_{k}^{2}(2\nu_{1}^{2}\mathbb{E}[\|x_{k}-x_{k+\frac{1}{2}}\|^{2}\mid\mathcal{F}_{k}]+6\nu_{1}^{2}\|x_{k}-x^{*}\|^{2}+6\nu_{1}^{2}\|x^{*}\|^{2}+2\nu_{2}^{2})$$

$$=(1+12\gamma_{k}^{2}\nu_{1}^{2})\|x_{k}-x^{*}\|^{2}-2\gamma_{k}\alpha\operatorname{dist}(x_{k},X^{*})-\left(\frac{5}{8}-2\gamma_{k}^{2}(L^{2}+2\nu_{1}^{2})\right)\mathbb{E}[\|x_{k}-x_{k+\frac{1}{2}}\|^{2}\mid\mathcal{F}_{k}]$$

$$+8\gamma_{k}^{2}\alpha^{2}+4\gamma_{k}^{2}(C+\alpha)^{2}+2\gamma_{k}(C+\alpha)\operatorname{dist}(x_{k},X)+2\gamma_{k}^{2}(6\nu_{1}^{2}\|x^{*}\|^{2}+2\nu_{2}^{2}). \tag{67}$$

According to Lemma 9, we have

$$\mathbb{E}[\|x_{k} - x_{k + \frac{1}{2}}\|^{2} \mid \mathcal{F}_{k}] = \mathbb{E}[\|x_{k} - \Pi_{l_{k}}(x_{k} - \gamma_{k}F(x_{k}, \omega_{k}))\|^{2} \mid \mathcal{F}_{k}]$$

$$\geq \mathbb{E}[\|x_{k} - \Pi_{l_{k}}(x_{k})\|^{2} \mid \mathcal{F}_{k}] \geq \frac{\rho}{mn} \operatorname{dist}^{2}(x_{k}, X).$$
(68)

where the last inequality follows from Lemma 6. Multiplying (68) by $\frac{1}{8}$ and using it in (67), we have

$$\mathbb{E}[\|x_{k+1} - x^*\|^2 \mid \mathcal{F}_k] \leq (1 + 12\gamma_k^2 \nu_1^2) \|x_k - x^*\|^2 - 2\gamma_k \alpha \operatorname{dist}(x_k, X^*) \\
- \left(\frac{1}{2} - 2\gamma_k^2 (L^2 + 2\nu_1^2)\right) \mathbb{E}[\|x_k - x_{k+\frac{1}{2}}\|^2 \mid \mathcal{F}_k] + 8\gamma_k^2 \alpha^2 + 4\gamma_k^2 (C + \alpha)^2 \\
+ 2\gamma_k (C + \alpha) \operatorname{dist}(x_k, X) - \frac{\rho}{8m\eta} \operatorname{dist}^2(x_k, X) + 2\gamma_k^2 (6\nu_1^2 \|x^*\|^2 + 2\nu_2^2) \\
= (1 + 12\gamma_k^2 \nu_1^2) \|x_k - x^*\|^2 - 2\gamma_k \alpha \operatorname{dist}(x_k, X^*) - \left(\frac{1}{2} - 2\gamma_k^2 (L^2 + 2\nu_1^2)\right) \mathbb{E}[\|x_k - x_{k+\frac{1}{2}}\|^2 \mid \mathcal{F}_k] \\
+ 8\gamma_k^2 \alpha^2 + 4\gamma_k^2 (C + \alpha)^2 - \frac{\rho}{8m\eta} \left(\operatorname{dist}(x_k, X) - \frac{8m\eta\gamma_k (C + \alpha)}{\rho}\right)^2 \\
+ \frac{8m\eta(C + \alpha)^2}{\rho} \gamma_k^2 + 2\gamma_k^2 (6\nu_1^2 \|x^*\|^2 + 2\nu_2^2) \\
\leq (1 + 12\gamma_k^2 \nu_1^2) \|x_k - x^*\|^2 - 2\gamma_k \alpha \operatorname{dist}(x_k, X^*) + 8\gamma_k^2 \alpha^2 + 4\gamma_k^2 (C + \alpha)^2 \\
+ \frac{8m\eta(C + \alpha)^2}{\rho} \gamma_k^2 + 2\gamma_k^2 (6\nu_1^2 \|x^*\|^2 + 2\nu_2^2)$$
(69)

Now we may invoke Lemma 2. It follows that $\{\|x_k - x^*\|^2\}$ is convergent in an a.s. sense and $\sum_k 2\gamma_k \alpha \operatorname{dist}(x_k, X^*) < \infty$ a.s. . We first show that $\operatorname{dist}(x_k, X^*) \xrightarrow{k \to \infty} 0$ a.s. . We proceed by contradiction and assume that with finite probability, $\operatorname{dist}(x_k, X^*) \to h(\omega) > 0$ for $\omega \in V$ where $\mathbb{P}(V) > 0$. Since $\sum_k \gamma_k = \infty$, it follows that $\sum_k \gamma_k \operatorname{dist}(x_k, X^*) = \infty$ with finite probability. But this contradicts $\sum 2\gamma_k \alpha \operatorname{dist}(x_k, X^*) < \infty$ a.s., implying that $\operatorname{dist}(x_k, X^*) \to 0$ in an a.s. sense. In a similar fashion as in Proposition 3, we may show that the entire sequence of $\{x_k\}$ is convergent to a random point in X^* .

We continue with an analysis of the infeasibility sequence.

Proposition 5. Let Assumptions 1-2, 4-6 hold. Suppose $\{x_k\}$ is generated by $(\mathbf{r}\text{-}\mathbf{SSE})$, where the projections are randomly generated. In addition, suppose there exists a $D_X > 0$ such that $||u-v||^2 \le D_X^2$ for any $u, v \in X$. Then the following hold.

(a) If
$$\sum_{k} \gamma_{k} = \infty$$
 and $\sum_{k} \gamma_{k}^{2} < \infty$, then $\operatorname{dist}(x_{K}, X) \xrightarrow{k \to \infty} 0$.

(b) Suppose
$$\gamma_k = 1/k^{t/2}$$
 where $t \ge 1$. Then $\mathbb{E}[\operatorname{dist}(x_k, X)] \le \mathcal{O}\left(\frac{1}{k^{t/2}}\right)$ for any $k \ge \bar{k}$, where

$$\bar{k} \triangleq \left\lceil \left(\frac{1}{1 - (1 - \beta/2)^{1/t}} - 1 \right) \right\rceil.$$

(c) Suppose
$$\gamma_k = 1/k^{1/2}$$
 and $\bar{x}_{K,\bar{k}} \triangleq \frac{\sum_{k=\bar{k}+\lfloor K/2\rfloor}^{\bar{k}+K} \gamma_k x_k}{\sum_{k=\bar{k}+\lfloor K/2\rfloor}^{K+\bar{k}} \gamma_k}$. Then $\mathbb{E}[\operatorname{dist}(\bar{x}_{K,\bar{k}},X)] \leq \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$.

Proof. (a) Let
$$z_k = x_k - \gamma_k F(x_{k+\frac{1}{2}}, \omega_{k+\frac{1}{2}})$$
. Choose $\theta \ge \max\left\{1, \frac{\rho}{m\eta}\right\}$. We have

$$\operatorname{dist}^{2}(x_{k+1}, X) \leq \|x_{k+1} - \Pi_{X}(x_{k+\frac{1}{2}})\|^{2} = \|\Pi_{T_{k}}(z_{k}) - x_{k+\frac{1}{2}} + x_{k+\frac{1}{2}} - \Pi_{X}(x_{k+\frac{1}{2}})\|^{2} \\
\leq \left(1 + \frac{16\theta m\eta}{\rho}\right) \|\Pi_{C_{k}}(z_{k}) - x_{k+\frac{1}{2}}\|^{2} + \left(1 + \frac{\rho}{16\theta m\eta}\right) \|x_{k+\frac{1}{2}} - \Pi_{X}(x_{k+\frac{1}{2}})\|^{2} \\
= \left(1 + \frac{16\theta m\eta}{\rho}\right) \|\Pi_{C_{k}}(z_{k}) - \Pi_{l_{k}}(x_{k})\|^{2} + \left(1 + \frac{\rho}{16\theta m\eta}\right) \|x_{k+\frac{1}{2}} - \Pi_{X}(x_{k+\frac{1}{2}})\|^{2} \\
= \left(1 + \frac{16\theta m\eta}{\rho}\right) \|\Pi_{C_{k}}(z_{k}) - \Pi_{C_{k}}(x_{k})\|^{2} + \left(1 + \frac{\rho}{16\theta m\eta}\right) \|x_{k+\frac{1}{2}} - \Pi_{X}(x_{k+\frac{1}{2}})\|^{2} \\
\leq \left(1 + \frac{16\theta m\eta}{\rho}\right) \|z_{k} - x_{k}\|^{2} + \left(1 + \frac{\rho}{16\theta m\eta}\right) \|x_{k+\frac{1}{2}} - \Pi_{X}(x_{k+\frac{1}{2}})\|^{2}, \tag{70}$$

where we leverage $||a+b||^2 \le \left(1 + \frac{16\theta m\eta}{\rho}\right) ||a||^2 + \left(1 + \frac{\rho}{16\theta m\eta}\right) ||b||^2$. From (70), we can deduce

$$\begin{split} \mathbb{E}[\mathrm{dist}^2(x_{k+1}, X) \mid \mathcal{F}_k] &\leq \left(1 + \frac{16\theta m\eta}{\rho}\right) \|\gamma_k \mathbb{E}[F(x_{k+\frac{1}{2}}, \omega_{k+\frac{1}{2}}) \mid \mathcal{F}_k]\|^2 + \left(1 + \frac{\rho}{16\theta m\eta}\right) \mathbb{E}[\mathrm{dist}^2(x_{k+\frac{1}{2}}, X) \mid \mathcal{F}_k] \\ &\leq \left(1 + \frac{16\theta m\eta}{\rho}\right) \gamma_k^2 (4(L^2 + \nu_1^2) \|x_{k+\frac{1}{2}} - x^*\|^2 + 4 \|F(x^*)\|^2 + 4\nu_1^2 \|x^*\|^2 + 2\nu_2^2) \\ &\quad + \left(1 + \frac{\rho}{16\theta m\eta}\right) \mathbb{E}[\mathrm{dist}^2(x_{k+\frac{1}{2}}, X) \mid \mathcal{F}_k] \\ &\leq \left(1 + \frac{16\theta m\eta}{\rho}\right) \gamma_k^2 (8(L^2 + \nu_1^2) D_X^2 + 4C^2 + 4\nu_1^2 \|x^*\|^2 + 2\nu_2^2) \\ &\quad + \left(1 + \frac{\rho}{16\theta m\eta} + 8\left(1 + \frac{16\theta m\eta}{\rho}\right) (L^2 + \nu_1^2) \gamma_k^2\right) \mathbb{E}[\mathrm{dist}^2(x_{k+\frac{1}{2}}, X) \mid \mathcal{F}_k]. \end{split}$$

Suppose $8\left(5 + \frac{16\theta m\eta}{\rho}\right)(L^2 + \nu_1^2)\gamma_k^2 \le \frac{\rho}{16\theta m\eta}$. Then we have

$$\mathbb{E}[\operatorname{dist}^{2}(x_{k+1}, X) \mid \mathcal{F}_{k}] \leq \left(1 + \frac{16\theta m\eta}{\rho}\right) \gamma_{k}^{2} (8(L^{2} + \nu_{1}^{2}) D_{X}^{2} + 4C^{2} + 4\nu_{1}^{2} ||x^{*}||^{2} + 2\nu_{2}^{2}) + \left(1 + \frac{\rho}{8\theta m\eta}\right) \mathbb{E}[\operatorname{dist}^{2}(x_{k+\frac{1}{2}}, X) \mid \mathcal{F}_{k}].$$

$$(71)$$

We can bound the second term using a similar way with (46) as follows:

$$\mathbb{E}[\operatorname{dist}^{2}(x_{k+\frac{1}{2}}, X) \mid \mathcal{F}_{k}] \leq \left(1 - \frac{\rho}{4\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) + \left(5 + \frac{4\theta m \eta}{\rho}\right) \|\gamma_{k} \mathbb{E}[F(x_{k}, \omega_{k}) \mid \mathcal{F}_{k}]\|^{2}$$

$$\leq \left(1 - \frac{\rho}{4\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) + \left(5 + \frac{4\theta m \eta}{\rho}\right) \gamma_{k}^{2} (4(L^{2} + \nu_{1}^{2}) \|x_{k} - x^{*}\|^{2}$$

$$+ 4\|F(x^{*})\|^{2} + 4\nu_{1}^{2} \|x^{*}\|^{2} + 2\nu_{2}^{2}). \tag{72}$$

Using (48) in (72), it follows that

$$\mathbb{E}[\operatorname{dist}^{2}(x_{k+\frac{1}{2}}, X) \mid \mathcal{F}_{k}] \leq \left(1 - \frac{\rho}{4\theta m \eta} + \underbrace{8\left(5 + \frac{4\theta m \eta}{\rho}\right)(L^{2} + \nu_{1}^{2})\gamma_{k}^{2}}\right) \operatorname{dist}^{2}(x_{k}, X)$$

$$+ \left(5 + \frac{4\theta m \eta}{\rho}\right) \gamma_{k}^{2}(8(L^{2} + \nu_{1}^{2})D_{X}^{2} + 4C^{2} + 4\nu_{1}^{2}\|x^{*}\|^{2} + 2\nu_{2}^{2})$$

$$\leq \left(1 - \frac{\rho}{8\theta m \eta}\right) \operatorname{dist}^{2}(x_{k}, X) + \left(5 + \frac{4\theta m \eta}{\rho}\right) \gamma_{k}^{2}(8(L^{2} + \nu_{1}^{2})D_{X}^{2} + 4C^{2} + 4\nu_{1}^{2}\|x^{*}\|^{2} + 2\nu_{2}^{2}).$$

$$(73)$$

Using (73) in (71), we obtain

$$\mathbb{E}[\operatorname{dist}^{2}(x_{k+1}, X) \mid \mathcal{F}_{k}] \leq \left(1 - \frac{\rho^{2}}{64\theta^{2}m^{2}\eta^{2}}\right) \operatorname{dist}^{2}(x_{k}, X) + \hat{D}\gamma_{k}^{2},\tag{74}$$

where $\hat{D} \triangleq \left(1 + \frac{16\theta m\eta}{\rho} + \left(1 + \frac{\rho}{8\theta m\eta}\right)\left(5 + \frac{4\theta m\eta}{\rho}\right)\right) \left(8(L^2 + \nu_1^2)D_X^2 + 4C^2 + 4\nu_1^2\|x^*\|^2 + 2\nu_2^2\right)$. We may now invoke Lemma 3 and the summability of γ_k^2 to claim $\operatorname{dist}^2(x_k, X) \xrightarrow{k \to \infty} 0$.

(b) Let $\beta \triangleq 1 - \frac{\rho^2}{64\theta^2 m^2 \eta^2}$. The conclusion holds by using a similar fashion with Proposition 4(b).

(c) We can derive the result using
$$(52)$$
.

We conclude this section with a rate of convergence of the gap function in terms of the projection of the averaged sequence for (**r-SSE**) and the associated oracle complexity bound.

Proposition 6. Let Assumptions 1-2, 4-6 hold. Let $\gamma_k = \frac{\gamma_0}{\sqrt{k}}$ and assume $\gamma_0 \leq \frac{1}{2\sqrt{L^2+5\nu_1^2}}$. In addition, for any $u, v \in X$, suppose that there exists a $D_X > 0$ such that $||u-v||^2 \leq D_X^2$. Then the following holds for any sequence generated by (**r-SSE**) in an expected value sense, where $\bar{y}_{K,\bar{k}} \triangleq \frac{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \gamma_k \Pi_X(x_{k+\frac{1}{2}})}{\sum_{k=1}^{K+\bar{k}} \gamma_k \Gamma_X(x_{k+\frac{1}{2}})}$.

- (a) $\mathbb{E}[G(\bar{y}_{K,\bar{k}})] \leq \mathcal{O}\left(\frac{1}{\sqrt{K}}\right);$
- (b) The oracle complexity to compute an $\bar{y}_{K,\bar{k}}$ such that $\mathbb{E}[\operatorname{dist}(\bar{y}_{K,\bar{k}},X^*)] \leq \epsilon$ is bounded by $\mathcal{O}\left(\frac{1}{\epsilon^2}\right)$.

Proof. (a) Using (53) in (63), we obtain the following inequality which is similar with (64)

$$||x_{k+1} - x^*||^2 \le ||x_k - x^*||^2 - (1 - 2\gamma_k^2 L^2) ||x_k - x_{k+\frac{1}{2}}||^2 + 2\gamma_k^2 ||w_k - w_{k+\frac{1}{2}}||^2 + 2\gamma_k w_{k+\frac{1}{2}}^T (x^* - x_{k+\frac{1}{2}}) - 2\gamma_k F(x^*)^T (\Pi_X(x_{k+\frac{1}{2}}) - x^*) + 2\gamma_k C \operatorname{dist}(x_{k+\frac{1}{2}}, X).$$

Similarly with (54), we have

$$\begin{split} &2\gamma_{k}F(y)^{T}(\Pi_{X}(x_{k+\frac{1}{2}})-y)\leq\|x_{k}-y\|^{2}-\|x_{k+1}-y\|^{2}-(1-2\gamma_{k}^{2}L^{2})\|x_{k}-x_{k+\frac{1}{2}}\|^{2}\\ &+2\gamma_{k}^{2}\|w_{k}-w_{k+\frac{1}{2}}\|^{2}+2\gamma_{k}w_{k+\frac{1}{2}}^{T}(y-x_{k+\frac{1}{2}})+2\gamma_{k}C\mathrm{dist}(x_{k},X)+4\gamma_{k}^{2}C^{2}+\frac{1}{4}\|x_{k}-x_{k+\frac{1}{2}}\|^{2}\\ &\leq\|x_{k}-y\|^{2}-\|x_{k+1}-y\|^{2}-\frac{1}{4}\|x_{k}-x_{k+\frac{1}{2}}\|^{2}+2\gamma_{k}^{2}\|w_{k}-w_{k+\frac{1}{2}}\|^{2}\\ &+2\gamma_{k}w_{k+\frac{1}{2}}^{T}(y-x_{k+\frac{1}{2}})+2\gamma_{k}C\mathrm{dist}(x_{k},X)+4\gamma_{k}^{2}C^{2}-(\frac{1}{2}-2\gamma_{k}^{2}L^{2})\|x_{k}-x_{k+\frac{1}{2}}\|^{2}, \end{split}$$

We now define an auxiliary sequence $\{u_k\}$ such that

$$u_{k+1} := \Pi_X(u_k - \gamma_k w_k),$$

where $u_0 \in X$. We may then express $2\gamma_k w_{k+\frac{1}{2}}^T (y-x_{k+\frac{1}{2}})$ as follows.

$$2\gamma_{k}w_{k+\frac{1}{2}}^{T}(y-x_{k+\frac{1}{2}}) = 2\gamma_{k}w_{k+\frac{1}{2}}^{T}(y-u_{k}) + 2\gamma_{k}w_{k+\frac{1}{2}}^{T}(u_{k}-x_{k+\frac{1}{2}})$$

$$\leq ||u_{k}-y||^{2} - ||u_{k+1}-y||^{2} + \gamma_{k}^{2}||w_{k+\frac{1}{2}}||^{2} + 2\gamma_{k}w_{k+\frac{1}{2}}^{T}(u_{k}-x_{k+\frac{1}{2}}).$$
 (75)

Summing over k and invoking (75), we obtain the following bound:

$$\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} 2\gamma_k F(y)^T (\Pi_X(x_{k+\frac{1}{2}}) - y) \leq \|x_0 - y\|^2 + \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} (2\gamma_k C \operatorname{dist}(x_k, X) - \frac{1}{4} \|x_k - x_{k+\frac{1}{2}}\|^2)$$

$$+ \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} 4C^2 \gamma_k^2 + \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} (4\gamma_k^2 \|w_k\|^2 + 5\gamma_k^2 \|w_{k+\frac{1}{2}}\|^2) + \|u_0 - y\|^2$$

$$+ \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} 2\gamma_k w_{k+\frac{1}{2}}^T (u_k - x_{k+\frac{1}{2}}) - \sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} (\frac{1}{2} - 2\gamma_k^2 L^2) \|x_k - x_{k+\frac{1}{2}}\|^2.$$

Dividing both sides by $\sum_{k=|K/2|+\bar{k}}^{K+\bar{k}} \gamma_k$, we have

$$\begin{split} &2F(y)^T (\bar{y}_{K,\bar{k}} - y) \leq \frac{B_2}{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \gamma_k} + \frac{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \left(2\gamma_k C \mathrm{dist}(x_k, X) - \frac{1}{4} \|x_k - x_{k+\frac{1}{2}} \|^2 \right)}{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \gamma_k} \\ &+ \frac{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} 4C^2 \gamma_k^2}{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \gamma_k} + \frac{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} (4\gamma_k^2 \|w_k\|^2 + 5\gamma_k^2 \|w_{k+\frac{1}{2}} \|^2) + \sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} 2\gamma_k w_{k+\frac{1}{2}}^T (u_k - x_{k+\frac{1}{2}})}{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \gamma_k} \\ &- \frac{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \left(\frac{1}{2} - 2\gamma_k^2 L^2\right) \|x_k - x_{k+\frac{1}{2}} \|^2}{\sum_{k=\lfloor K/2 \rfloor + \bar{k}}^{K+\bar{k}} \gamma_k}, \end{split}$$

where $||x_0 - y||^2 + ||u_0 - y||^2 \le 2||x_0 - x^*||^2 + 2||x^* - y||^2 + 2||u_0 - x^*||^2 + 2||x^* - y||^2 \le 2||x_0 - x^*||^2 + 2||u_0 - x^*||^2 + 2D_X^2 \triangleq B_2$. By taking supremum over $y \in X$, we obtain the following inequality:

$$G(\bar{y}_{K,\bar{k}}) \triangleq \sup_{y \in X} F(y)^{T} (\bar{y}_{K,\bar{k}} - y) \leq \frac{B_{2}}{2\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \left(2\gamma_{k}C \operatorname{dist}(x_{k}, X) - \frac{1}{4} \|x_{k} - x_{k+\frac{1}{2}}\|^{2}\right)}{2\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} 4C^{2}\gamma_{k}^{2}} + \frac{\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \left(4\gamma_{k}^{2} \|w_{k}\|^{2} + 5\gamma_{k}^{2} \|w_{k+\frac{1}{2}}\|^{2}\right) + \sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} 2\gamma_{k} w_{k+\frac{1}{2}}^{T} (u_{k} - x_{k+\frac{1}{2}})}{2\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \gamma_{k}} - \frac{\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \left(\frac{1}{2} - 2\gamma_{k}^{2} L^{2}\right) \|x_{k} - x_{k+\frac{1}{2}}\|^{2}}{2\sum_{k=\lfloor K/2\rfloor + \bar{k}}^{K+\bar{k}} \gamma_{k}}.$$

$$(76)$$

Taking unconditional expectation and using (68) in (76), we have

$$\mathbb{E}[G(\bar{y}_{K,\bar{k}})] \leq \frac{B_{2}}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \left(2\gamma_{k}C\mathbb{E}[\operatorname{dist}(x_{k},X)] - \frac{\rho}{4m\eta}\mathbb{E}[\operatorname{dist}^{2}(x_{k},X)]\right)}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} 4C^{2}\gamma_{k}^{2}}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \left(\gamma_{k}^{2}(56\nu_{1}^{2}\mathbb{E}[\operatorname{dist}^{2}(x_{k},X)] + 56\nu_{1}^{2}D_{X}^{2} + 28\nu_{1}^{2}||x^{*}||^{2} + 9\nu_{2}^{2})\right)}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_{k}} - \frac{\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \left(\frac{1}{2} - 2\gamma_{k}^{2}(L^{2} + 5\nu_{1}^{2})\right)||x_{k} - x_{k+\frac{1}{2}}||^{2}}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_{k}}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} 4C^{2}\gamma_{k}^{2}}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_{k}} + \frac{(56\nu_{1}^{2}(\bar{b}/\bar{k}) + 56\nu_{1}^{2}D_{X}^{2} + 28\nu_{1}^{2}||x^{*}||^{2} + 9\nu_{2}^{2})\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_{k}}^{K+\bar{k}}}{2\sum_{k=\lfloor K/2\rfloor+\bar{k}}^{K+\bar{k}} \gamma_{k}}. \tag{77}$$

By substituting (59) and (60) in (77), we obtain that the following holds:

$$\mathbb{E}[G(\bar{y}_{K,\bar{k}})] \le \mathcal{O}\left(\frac{1}{\sqrt{K}}\right).$$

(b) The result follows using the same avenue as Proposition 3(b).

Remark 2. Proving a.s. convergence of our randomized projection schemes relies on imposing weak-sharpness assumptions as seen in [17]. However, gap statements do not impose such a requirement. Since variance-reduction techniques cannot directly overcome the impact of the infeasibility in the iterates, the resulting rate diminishes to $\mathcal{O}(1/\sqrt{K})$, similar to that seen in the classical rate statements for standard stochastic projection schemes for monotone stochastic variational inequality problems [21, 47].

5 Numerical Results

In this section, we apply the schemes on a stochastic Nash-Cournot equilibrium problem (Section 5.1) and the computation of the invariant distribution of a Markov chain (Section 5.2).

5.1 A Stochastic Nash-Cournot Equilibrium Problem

In this section, we present and compare the computational results of applying the proposed schemes on a stochastic Nash-Cournot equilibrium problem. This game is assumed that \mathcal{I} firms compete over a network of \mathcal{J} nodes. Level of production and sales of firm $i \in \mathcal{I}$ at node $j \in \mathcal{J}$ are denoted by q_{ij} and s_{ij} , respectively. Furthermore, we assume the cost of production at node j is $C_{ij}(q_{ij})$ and the price at node j is denoted by $Q_j(\bar{s}_j,\xi) = a_j(\xi) - b_j\bar{s}_j$, where \bar{s}_j is the aggregate sales at node j defined as $\bar{s}_j \triangleq \sum_{i=1}^{\mathcal{I}} s_{ij}$. For simplicity, we assume the transportation costs are zero. Thus, each firm i will solve a profit maximization problem given by the following:

$$\max \mathbb{E}\left[\sum_{j\in\mathcal{J}}(Q_j(\bar{s}_j,\xi)s_{ij}-C_{ij}(q_{ij}))\right]$$

subject to
$$\sum_{j \in \mathcal{J}} q_{ij} = \sum_{j \in \mathcal{J}} s_{ij}, \quad 0 \le q_{ij} \le \operatorname{cap}_{ij}, \quad s_{ij} \ge 0, \quad \forall j \in \mathcal{J}$$

This is an instance of a generalized Nash equilibrium problem (GNEP) with *shared constraints* and a variational equilibrium (VE) (cf. [12]) of this (GNEP) given by a solution to $VI(\mathcal{X}, F)$, where

$$x \triangleq \begin{pmatrix} s \\ q \end{pmatrix}, s \triangleq \begin{pmatrix} s_{\bullet,1} \\ \vdots \\ s_{\bullet,\mathcal{J}} \end{pmatrix}, q \triangleq \begin{pmatrix} q_{\bullet,1} \\ \vdots \\ q_{\bullet,\mathcal{J}} \end{pmatrix}, \text{ and } z_{\bullet,j} \triangleq \begin{pmatrix} z_{1j} \\ \vdots \\ x_{\mathcal{I}j} \end{pmatrix}, \mathbf{a} \triangleq \begin{pmatrix} a_1 \mathbf{1} \\ \vdots \\ a_{\mathcal{J}} \mathbf{1} \end{pmatrix},$$

$$\mathbf{B} \triangleq \begin{pmatrix} b_1 \mathbf{D} \\ \vdots \\ b_{\mathcal{J}} \mathbf{D} \end{pmatrix}, \mathbf{D} \triangleq (I + \mathbf{1} \mathbf{1}^T), c \triangleq \begin{pmatrix} c_{\bullet,1} \\ \vdots \\ c_{\bullet,\mathcal{J}} \end{pmatrix}, c_{\bullet,j} \triangleq \begin{pmatrix} c_{1j} \\ \vdots \\ c_{\mathcal{I}j} \end{pmatrix}, F(x) \triangleq \begin{pmatrix} Bs - \mathbf{a} \\ c \end{pmatrix},$$
and
$$\mathcal{X} \triangleq \left\{ x = (s,q) \mid \sum_{j \in \mathcal{J}} q_{ij} = \sum_{j \in \mathcal{J}} s_{ij}, i = 1, \cdots, \mathcal{I}, \ 0 \leq q_{ij} \leq \operatorname{cap}_{ij}, \quad s_{ij} \geq 0, \ i \in \mathcal{I}, j \in \mathcal{J} \right\}.$$

Before proceeding, we prove that $VI(\mathcal{X}, F)$ satisfies the required assumptions where $\mathcal{X} \subseteq \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^n$:

(i) F(x) is Lipschitz on \mathbb{R}^n by noting that for any $x_1, x_2 \in \mathbb{R}^n$, the following holds.

$$||F(x_1) - F(x_2)|| = ||\mathbf{B}(s_1 - s_2)|| \le ||\mathbf{B}|| ||s_1 - s_2|| \le ||\mathbf{B}|| ||x_1 - x_2||.$$

(ii) F(x) is monotone on \mathbb{R}^n by noting that for any $x_1, x_2 \in \mathbb{R}^n$, the following holds.

$$(F(x_1) - F(x_2))^T(x_1 - x_2) = (\mathbf{B}(s_1 - s_2))^T(s_1 - s_2) > 0,$$

since $\mathbf{B} \succ 0$, a consequence of $b_j > 0$ for all j and $(I + \mathbf{1}\mathbf{1}^T \succ 0$. Note that F is merely monotone by noting that if $x_1 = (\mathbf{0}, q_1)$ and $x_2 = (\mathbf{0}, q_2)$, we have

$$(F(x_1) - F(x_2))^T (x_1 - x_2) = 0.$$

(iii) \mathcal{X} is a compact set by recalling that $0 \leq q_{ij} \leq \operatorname{cap}_{ij}$ for every i, j and by observing that for any i, j.

$$s_{ij} \le \sum_{j \in \mathcal{J}} s_{ij} = \sum_{j \in \mathcal{J}} q_{ij} \le \sum_{j \in \mathcal{J}} \operatorname{cap}_{ij}.$$

Furthermore, $||F(x)|| \le (||A|| + ||c||)||s|| \le (||A|| + ||c||)||$ cap||, for all x and cap denotes the vector of capacities over all nodes and firms. One assumption that is generally more challenging to verify is the weak-sharpness requirement. There may be approaches for claiming that such a condition holds by leveraging weak sharpness (cf. [13, Ch. 3]) and this remains a goal of future work.

Problem parameters. We assume that there are $\mathcal{I}=5$ firms and $\mathcal{J}=4$ nodes whille the capacity $\operatorname{cap}_{ij}=300$, for all i,j. We assume that $c_{ij}=1.5$ and d_{ij} is a positive constant for all i,j. Furthermore, for all $j, b_j=0.05$ and $a_j(\xi)\sim U[49.5,50.5]$ where U[a,b] denotes the uniform distribution on the interval [a,b].

Algorithm parameters. We choose $\gamma = 0.1$ which satisfies the requirements of (**v-SPRG**) and (**v-SSE**) by noting that L = 0.3 and $\nu_1 = 0$. In addition, we choose $\gamma_k = \frac{\gamma}{\sqrt{k}}$ for (**r-SPRG**) and (**r-SSE**). Finally, in variance-reduced settings, we choose $N_k = |k^{1.1}|$ for $k \ge 0$.

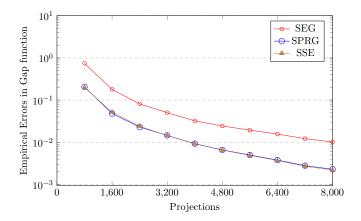


Figure 3: Convergence based on projections under mere monotonicity

Recall that SEG requires two projections onto the set while the two proposed schemes just require one. We compare their performance under the same number of projections in Fig. 3. Next we change the size and parameters of the original game to ascertain parametric sensitivity. In Table 3, we consider a set of 16 problems where the settings, the empirical errors, and elapsed time are shown in Table 3. Table 3 shows the performance after 4000 iterations and we observe that while SEG has almost the same empirical error as the others but requires significantly more computational effort. To examine the impact 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 4 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 introduced.

Key findings. The key findings are that (**v-SPRG**) and (**v-SSE**) produce comparable empirical errors to (**v-SEG**) but do so in approximately 65% of the time utilized by (**v-SEG**). Moreover, the presence of variance reduction allows for significant improvement in empirical error in comparision with the single-sample counterparts (See Table 4).

5.2 Markov Invariant Distribution Approximation

We now test the performance of the random projection schemes on an example from [39] 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 π . 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 Σ . Then we use a projection approach to approximate $\pi = P^T \pi$ as $\Sigma x = \Pi_X(P^T \Sigma x)$, where $X \triangleq \{x \mid \Sigma x \geq 0, e^T \Sigma x = 1\}$. It has been proved [4, 39] that this projected equation is equivalent to the variational inequality problem VI(X, Sx). where $S = \Sigma^T (I - P^T)\Sigma$. Before proceeding, we verify that this problem satisfies the required assumptions.

Table 3: I	Errors and	elansed	time com	parison o	f the	three	schemes	with	different	parameters
Table 9. I	arrors and	Clapsca	UIIIC COIII	parison o	1 0110	UIIICC	SOLICILIOS	WILLIAM	CHILCH CHI	parametri

	(v-SEG)	Time	(v-SSE)	Time	(v-SPRG)	Time
$\mathcal{I} = 5, \mathcal{J} = 4, c_{ij} = 2, b_j = 0.05$	9.1e-3	2.4e3s	9.1e-3	1.6e3s	9.2e-3	1.5e3s
$\mathcal{I} = 6, \mathcal{J} = 4, c_{ij} = 2, b_j = 0.05$	1.0e-2	2.4e3s	1.1e-2	1.6e3s	1.1e-2	1.5e3s
$\mathcal{I} = 5, \mathcal{J} = 5, c_{ij} = 2, b_{j} = 0.05$	1.2e-2	2.5e3s	1.2e-2	1.8e3s	1.2e-2	1.5e3s
$\mathcal{I} = 6, \mathcal{J} = 5, c_{ij} = 2, b_{j} = 0.05$	1.2e-2	2.5e3s	1.1e-2	1.9e3s	1.3e-2	1.5e3s
$\mathcal{I} = 5, \mathcal{J} = 4, c_{ij} = 1, b_j = 0.05$	9.1e-3	2.3e3s	9.2e-3	1.7e3s	9.3e-3	1.4e3s
$\mathcal{I} = 6, \mathcal{J} = 4, c_{ij} = 1, b_j = 0.05$	1.1e-2	2.3e3s	1.1e-2	1.8e3s	1.1e-2	1.4e3s
$\mathcal{I} = 5, \mathcal{J} = 5, c_{ij} = 1, b_j = 0.05$	1.2e-2	2.4e3s	1.3e-2	1.8e3s	1.3e-2	1.5e3s
$\mathcal{I} = 6, \mathcal{J} = 5, c_{ij} = 1, b_j = 0.05$	1.2e-2	2.4e3s	1.3e-2	1.9e3s	1.3e-2	1.5e3s
$\mathcal{I} = 5, \mathcal{J} = 4, c_{ij} = 2, b_{j} = 0.1$	1.1e-2	2.4e3s	1.1e-2	1.6e3s	1.2e-2	1.4e3s
$\mathcal{I} = 6, \mathcal{J} = 4, c_{ij} = 2, b_{j} = 0.1$	1.1e-2	2.4e3s	1.0e-2	1.6e3s	1.1e-2	1.5e3s
$\mathcal{I} = 5, \mathcal{J} = 5, c_{ij} = 2, b_{j} = 0.1$	1.2e-2	2.4e3s	1.1e-2	1.7e3s	1.2e-2	1.4e3s
$\mathcal{I} = 6, \mathcal{J} = 5, c_{ij} = 2, b_{j} = 0.1$	1.1e-2	2.5e3s	1.2e-2	1.8e3s	1.3e-2	1.4e3s
$\mathcal{I} = 5, \mathcal{J} = 4, c_{ij} = 1, b_{j} = 0.1$	1.0e-2	2.4e3s	1.0e-2	1.7e3s	1.1e-2	1.3e3s
$\mathcal{I} = 6, \mathcal{J} = 4, c_{ij} = 1, b_{j} = 0.1$	1.1e-2	2.4e3s	1.1e-2	1.6e3s	1.1e-2	1.3e3s
$\mathcal{I} = 5, \mathcal{J} = 5, c_{ij} = 1, b_{j} = 0.1$	1.2e-2	2.4e3s	1.2e-2	1.8e3s	1.1e-2	1.4e3s
$\mathcal{I} = 6, \mathcal{J} = 5, c_{ij} = 1, b_j = 0.1$	1.1e-2	2.4e3s	1.1e-2	1.7e3s	1.2e-3	1.0e3s

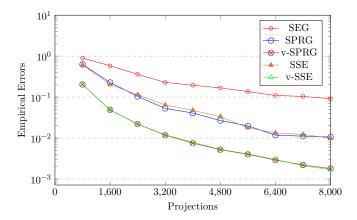


Figure 4: Performance comparison between variance reduced schemes and original ones

- (i) The mapping Sx is a monotone map on \mathbb{R}^n , since $(S+S^T)/2$ is a positive semidefinite matrix, a result that follows from P being a transition matrix. In addition, Sx is a Lipschitz continuous map with constant ||S||.
- (ii) The set X is clearly compact and $||F(x)|| \le ||S|| ||x||$, since ||x|| is bounded on X.

Problem parameters. The transition matrix P is randomly generated. We choose the columns of Σ based on sinusoidal functions of various frequencies (see [39] for details).

Algorithm parameters. We choose $\gamma_0 = 0.1$ which satisfies the requirements of (**v-SPRG**) and (**v-SSE**) by noting that L = 1.0036 (in our setting) and $\nu_1 = 0$. Finally, for (**r-SPRG**) and (**r-SSE**), we choose $\gamma_k = \frac{\gamma}{\sqrt{k}}$, for $k \ge 0$. We choose $N_k = \lfloor k^{1.1} \rfloor$ in variance-reduced settings.

Figure 5 illustrates the empirical behavior of all of the schemes considered. We record the elapsed time and empirical errors of each scheme for 10 different transition matrices, as shown in Table 5 while the comparison between the original variance-reduced schemes and their random projection variants is shown in Table 6.

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**)

Table 4: 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	2.4e3s	1.1e-1	1.7e3s	7.5e-3	1.9e3s	1.1e-1	1.5e3s	7.4e-3	1.6e3s
24	1.3e-1	2.4e3s	1.4e-1	1.8e3s	7.7e-3	2.0e3s	1.3e-1	1.5e3s	7.7e-3	1.7e3s
28	1.8e-1	2.7e3s	1.7e-1	1.9e3s	7.9e-3	2.1e3s	1.9e-1	1.6e3s	8.0e-3	1.7e3s
32	2.0e-1	2.8e3s	1.9e-1	1.9e3s	8.3e-3	2.2e3s	2.0e-1	1.7e3s	8.2e-3	1.8e3s
36	2.5e-1	3.1e3s	2.5e-1	2.2e3s	8.7e-3	2.4e3s	2.4e-1	2.0e3s	8.8e-3	2.1e3s
40	3.4e-1	3.2e3s	3.5e-1	2.3e3s	9.0e-3	2.5e3s	3.5e-1	2.1e3s	9.1e-3	2.2e3s

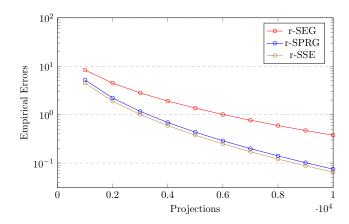


Figure 5: Empirical behavior for random projection variants

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 5: Comparison of schemes with differing transition matrices

Matrix	(r-SEG)	Time	(r-SSE)	Time	(r-SPRG)	Time
No.1	7.7e-2	1.4e3s	6.5e-2	1.4e3s	7.5e-2	0.7e3s
No.2	4.0e-2	1.3e3s	3.9e-2	1.4e3s	4.0e-2	0.7e3s
No.3	1.8e-2	1.3e3s	1.7e-2	1.4e3s	1.8e-2	0.7e3s
No.4	5.2e-2	1.4e3s	4.9e-2	1.4e3s	5.1e-2	0.7e3s
No.5	4.7e-2	1.3e3s	4.4e-2	1.4e3s	4.6e-2	0.7e3s
No.6	5.9e-2	1.3e3s	5.5e-2	1.4e3s	5.8e-2	0.7e3s
No.7	2.7e-2	1.4e3s	2.6e-2	1.4e3s	2.7e-2	0.7e3s
No.8	5.8e-2	1.3e3s	5.3e-2	1.4e3s	5.7e-2	0.7e3s
No.9	2.6e-2	1.4e3s	2.3e-2	1.4e3s	2.5e-2	0.7e3s
No.10	3.3e-2	1.4e3s	3.1e-2	1.4e3s	3.2e-2	0.7e3s

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 halfpace, the second of which is computable in closed form (a subgradient extragradient scheme). In both instances, under a variance-reduced regime, we derive a.s. convergence statements without imposing a compactness

Table 6: Comparision between original schemes and random projection variants

	(v-SEG)	(r-SEG)	(v-SSE)	(r-SSE)	(v-SRPG)	(r-SRPG)
Error	4.3e-3	7.7e-2	3.7e-3	6.5e-2	4.2e-3	7.5e-2
Time	2.8e4s	1.4e3s	1.6e4s	1.4e3s	1.5e4s	0.7e3s

requirement and while allowing for state-dependent noise. Notably, the sequences achieve a non-asymptotic rate of $\mathcal{O}(1/K)$ in terms of the expected gap function of an averaged sequence, matching its deterministic counterpart. Furthermore, when this set is given by the intersection of a large number of convex sets, we develop a random projection variant for each scheme. Again, a.s. convergence guarantees are developed. Since the sequence of iterates is no longer feasible, we proceed to develop rate guarantees for both the expected infeasibility of iterates as well as the expected gap function of a projected averaged sequence of iterates. 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] Y. CENSOR, A. GIBALI, AND S. 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, 165 (2017), pp. 113–149.
- [9] 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.
- [10] 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.
- [11] A. Defazio, F. Bach, and S. Lacoste-Julien, Saga: A fast incremental gradient method with support for non-strongly convex composite objectives, in Advances in Neural Information Processing Systems, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, eds., vol. 27, Curran Associates, Inc., 2014, pp. 1646–1654.
- [12] F. FACCHINEI AND C. KANZOW, Generalized nash equilibrium problems, 4or, 5 (2007), pp. 173-210.
- [13] F. FACCHINEI AND J. S. PANG, Finite-dimensional variational inequalities and complementarity problems, Springer Science & Business Media, 2007.
- [14] F. Yousefian, A. Nedić, and U. V. Shanbhag, On stochastic mirror-prox algorithms for stochastic cartesian variational inequalities: Randomized block coordinate and optimal averaging schemes, Set-Valued and Variational Analysis, (2018), pp. 789–819.

- [15] Y. G. HSIEH, F. IUTZELER, J. MALICK, AND P. MERTIKOPOULOS, On the convergence of single-call stochastic extra-gradient methods, arXiv:1908.08465, (2019).
- [16] A. 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.
- [17] A. N. IUSEM, A. JOFRÉ, AND P. THOMPSON, Incremental constraint projection methods for monotone stochastic variational inequalities, Math. Oper. Res., 44 (2019), pp. 236–263.
- [18] A. Jalilzadeh and U. V. Shanbhag, A proximal-point algorithm with variable sample-sizes (PPAWSS) for monotone stochastic variational inequality problems, in 2019 Winter Simulation Conference, WSC 2019, National Harbor, MD, USA, December 8-11, 2019, IEEE, 2019, pp. 3551–3562.
- [19] H. Jiang and H. Xu, Stochastic approximation approaches to the stochastic variational inequality problem, IEEE Transactions on Automatic Control, 53 (2008), pp. 1462–1475.
- [20] R. JOHNSON AND T. ZHANG, Accelerating stochastic gradient descent using predictive variance reduction, in Advances in Neural Information Processing Systems, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, eds., vol. 26, Curran Associates, Inc., 2013, pp. 315–323.
- [21] A. Juditsky, A. Nemirovski, and C. Tauvel, Solving variational inequalities with stochastic mirror-prox algorithm, Stochastic Systems, 1 (2011), pp. 17–58.
- [22] A. Kannan and U. V. Shanbhag, Distributed computation of equilibria in monotone Nash games via iterative regularization techniques, SIAM J. Optim., 22 (2012), pp. 1177–1205.
- [23] A. KANNAN AND U. V. SHANBHAG, Optimal stochastic extragradient schemes for pseudomonotone stochastic variational inequality problems and their variants, Comp. Opt. and Appl., 74 (2019), pp. 779-820.
- [24] 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.
- [25] I. Konnov, Equilibrium Models and Variational Inequalities, ISSN, Elsevier Science, 2007.
- [26] G. Korpelevich, The extragradient method for finding saddle points and other problems, Matecon, 12 (1976), pp. 747–756.
- [27] J. KOSHAL, A. NEDIC, AND U. V. SHANBHAG, Regularized iterative stochastic approximation methods for stochastic variational inequality problems, Automatic Control, IEEE Transactions on, 58 (2013), pp. 594–609.
- [28] Y. Malitsky, Projected reflected gradient methods for monotone variational inequalities, SIAM Journal on Optimization, 25 (2015), pp. 502–520.
- [29] A. MOKHTARI, A. OZDAGLAR, AND S. PATTATHIL, A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach, arXiv:1908.08465, (2019).
- [30] A. Nedić, Random projection algorithms for convex set intersection problems, in Decision and Control (CDC), 2010 49th IEEE Conference on, IEEE, 2010, pp. 7655–7660.
- [31] A. Nedić, Random algorithms for convex minimization problems, Mathematical programming, 129 (2011), pp. 225–253.
- [32] A. Nemirovski, Prox-method with rate of convergence $\mathcal{O}(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, SIAM Journal on Optimization, 15 (2004), pp. 229–251.
- [33] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, Robust stochastic approximation approach to stochastic programming, SIAM Journal on Optimization, 19 (2009), pp. 1574–1609.
- [34] B. T. Polyak, Introduction to optimization, Optimization Software New York, 1987.
- [35] U. RAVAT AND U. V. SHANBHAG, On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games, SIAM Journal on Optimization, 21 (2011), pp. 1168-1199.
- [36] H. Robbins and S. Monro, A stochastic approximation method, Ann. Math. Statistics, 22 (1951), pp. 400-407.
- [37] R. T. ROCKAFELLAR AND R. J.-B. WETS, Variational analysis, vol. 317, Springer Science & Business Media, 2009.
- [38] A. Shapiro, D. Dentcheva, and A. Ruszcsynski, Lectures on stochastic programming: modeling and theory, vol. 16, SIAM, 2014.

- [39] M. Wang and D. P. Bertsekas, Incremental constraint projection methods for variational inequalities, Mathematical Programming, 150 (2015), pp. 321–363.
- [40] M. WANG AND D. P. BERTSEKAS, Stochastic first-order methods with random constraint projection, SIAM Journal on Optimization, 26 (2016), pp. 681–717.
- [41] M. WANG, Y. CHEN, J. LIU, AND Y. GU, Random multi-constraint projection: Stochastic gradient methods for convex optimization with many constraints, arXiv preprint arXiv:1511.03760, (2015).
- [42] H. Xu, Sample average approximation methods for a class of stochastic variational inequality problems, Asia-Pacific Journal of Operational Research, 27 (2010), pp. 103–119.
- [43] H. Yin, U. V. Shanbhag, and P. G. Mehta, Nash equilibrium problems with scaled congestion costs and shared constraints, IEEE Trans. Automat. Control, 56 (2011), pp. 1702–1708.
- [44] F. Yousefian, A. Nedić, and U. V. Shanbhag, A regularized smoothing stochastic approximation (RSSA) algorithm for stochastic variational inequality problems, in Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World, IEEE Press, 2013, pp. 933–944.
- [45] F. Yousefian, A. Nedić, and U. V. Shanbhag, Optimal robust smoothing extragradient algorithms for stochastic variational inequality problems, in Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on, IEEE, 2014, pp. 5831–5836.
- [46] F. Yousefian, A. Nedić, and U. V. Shanbhag, Self-tuned stochastic approximation schemes for non-Lipschitzian stochastic multi-user optimization and Nash games, IEEE Transactions on Automatic Control, 61 (2016), pp. 1753–1766.
- [47] F. Yousefian, A. Nedić, and U. V. Shanbhag, On smoothing, regularization, and averaging in stochastic approximation methods for stochastic variational inequality problems, Math. Program., 165 (2017), pp. 391–431.