

A REGULARIZED ADAPTIVE STEPLENGTH STOCHASTIC APPROXIMATION SCHEME FOR MONOTONE STOCHASTIC VARIATIONAL INEQUALITIES

Farzad Yousefian
Angelia Nedić
Uday V. Shanbhag

UIUC, 104 S. Mathews,
Urbana, IL 61801, USA

ABSTRACT

We consider the solution of monotone stochastic variational inequalities and present an adaptive steplength stochastic approximation framework with possibly multivalued mappings. Traditional implementations of SA have been characterized by two challenges. First, convergence of standard SA schemes requires a strongly or strictly monotone single-valued mapping, a requirement that is rarely met. Second, while convergence requires that the steplength sequences need to satisfy $\sum_k \gamma_k = \infty$ and $\sum_k \gamma_k^2 < \infty$, little guidance is provided on a choice of sequences. In fact, standard choices such as $\gamma_k = 1/k$ may often perform poorly in practice. Motivated by the minimization of a suitable error bound, a recursive rule for prescribing steplengths is proposed for strongly monotone problems. By introducing a regularization sequence, extensions to merely monotone regimes are proposed. Finally, an iterative smoothing extension is suggested for accommodating multivalued mappings. Preliminary numerical results suggest that the schemes prove effective.

1 INTRODUCTION

In this paper, we consider the solution of a monotone stochastic variational inequality, denoted by $\text{VI}(X, F)$, where $X \subseteq \mathbb{R}^n$ is a closed and convex set, the i th component of $F : X \rightarrow \mathbb{R}^n$ is defined as $F_i(x) \triangleq \mathbb{E}[f_i(x; \xi)]$ and $f : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^n$. Note that ξ is a random variable with $\xi : \Omega \rightarrow \mathbb{R}^d$ and the associated probability space is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$. A solution of $\text{VI}(X, F)$ is denoted by x^* where $x^* \in \text{SOL}(X, F)$, the solution set of $\text{VI}(X, F)$, if $(x - x^*)^T F(x^*) \geq 0$ for all $x \in X$. It may be recalled that the mapping $F(x)$ is monotone over X if $(F(x) - F(y))^T (x - y) \geq 0$ for all $x, y \in X$.

Variational inequalities (cf. Facchinei and Pang 2003) are a broad class of objects that allows for capturing the set of solutions to convex optimization problems as well as noncooperative Nash games over continuous strategy sets. Stochastic variational inequalities are a relatively less studied class of problems and our interest lies in extensions in which the mapping F comprises of expectations in its components, as examined by Jiang and Xu (2008) and Ravat and Shanbhag (2011). Given that the components of $F(x)$ contains expectations, standard approaches for solving variational inequalities (cf. Facchinei and Pang 2003) cannot be leveraged since this necessitates having analytical expressions for the expectations. Instead, we turn to a simulation-based approach for obtaining solutions. Jiang and Xu (2008) appear to be amongst the first to have analyzed such schemes in the context of strongly monotone stochastic variational inequalities. Extensions to merely monotone regimes have been recently provided by Koshal et al. (2010) where Tikhonov regularization and proximal-point schemes are overlaid on standard SA techniques.

One of the motivations for the present work lies in the lack of guidance in choosing steplength sequences; notably, certain choices may lead to significant degradation in performance. Adaptive steplength stochastic approximation procedures have been studied extensively since the earliest work by Robbins and Monro (1951). Subsequently, Kesten (1958) suggested a sequence that adapts to the observed data. Subsequently, under suitable conditions, Sacks (1958) proved that a choice of the form a/k is optimal from the standpoint

of minimizing the asymptotic variance. Yet, the challenge lies in estimating the “optimal” a . Subsequently, Venter (1967) in what is possibly amongst the first *adaptive* steplength SA schemes, considered sequences of the form a_k/k where a_k is updated by leveraging past information. Averaging techniques were applied by Polyak and Juditsky (1992). More recently, Broadie et al. (2011) consider an adaptive Kiefer-Wolfowitz (KW) SA algorithm and derive upper bounds on its mean-squared error.

Yousefian, Nedić, and Shanbhag (2011) presented two adaptive steplength SA schemes for strongly convex programs, both of which produce a sequence of iterates guaranteed to converge almost surely to the true solution. Of these, the first relies on minimizing the upper bound on the error at every step; in fact, such a minimization leads to a recursive rule for updating steplengths and the scheme is referred to as a recursive SA (or RSA) scheme. The second scheme introduces regular reductions in the steplength when a suitable error criterion is satisfied and the associated scheme is referred to as a cascading SA scheme (or CSA) scheme. A random local smoothing approach is employed for accommodating the approximate solution of nonsmooth stochastic convex programs. In this paper, we revisit our RSA scheme with the intent of introducing three key generalizations: (i) First, we extend the regime of applicability of the RSA scheme to strongly monotone stochastic VIs; (ii) Second, we overlay a regularization scheme that facilitates addressing merely monotone stochastic VIs; and (iii) Third, we introduce a smoothing parameter that allows for solving problems with strongly monotone but multivalued maps. The remainder of the paper is organized as follows. In Section 2, we introduce a regularized recursive steplength stochastic approximation scheme for monotone stochastic VIs. We show that this algorithm produces iterates that converge in expectation to the solution. In Section 3, we introduce an iterative smoothing generalization of the recursive scheme. Finally, in Section 4, we present some preliminary numerical results and provide some concluding remarks in Section 5.

2 A REGULARIZED RECURSIVE STEPLENGTH STOCHASTIC APPROXIMATION SCHEME

In Section 2.1, we introduce a regularized SA scheme, akin to that developed by Koshal, Nedić, and Shanbhag (2010). By introducing a recursive rule for updating the steplengths, we present the regularized recursive SA scheme or RRSA scheme in Section 2.2.

2.1 A Regularized Stochastic Approximation scheme

Consider a regularized stochastic approximation scheme of the form:

$$\begin{aligned} x_{k+1} &= \Pi_X(x_k - \gamma_k(F(x_k) + \eta_k x_k + w_k)) \quad \text{for all } k \geq 0, \\ w_k &= f(x_k, \xi_k) - F(x_k), \end{aligned} \tag{1}$$

where $\{\gamma_k\}$ is the stepsize sequence, $\{\eta_k\}$ is a nonnegative sequence, and $x_0 \in X$ is a random initial vector that is independent of the random variable ξ and such that $E[\|x_0\|^2] < \infty$. We make the following two assumptions through our analysis, of which the first pertains to problem parameters and errors and the second to the algorithm parameters (steplength and regularization) sequences.

Assumption 1

- (a) The sets $X \subset \mathbb{R}^n$ are closed and convex; (b) $F(x)$ is Lipschitz with constant L over the set X .
- (c) The stochastic errors w_k satisfy $\sum_{k=0}^{\infty} \gamma_k^2 E[\|w_k\|^2 | \mathcal{F}_k] < \infty$ almost surely.

Assumption 2 Let the following hold:

- (a) $0 < \gamma_k < \frac{\eta_k}{2(\eta_k + L)^2}$ for all $k \geq 0$; (b) $\lim_{k \rightarrow \infty} \eta_k = 0$;
- (c) $\sum_{k=0}^{\infty} \gamma_k \eta_k = \infty$; (d) $\sum_{k=0}^{\infty} \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^2} (1 + \frac{1}{\gamma_k \eta_k}) < \infty$;
- (e) $\lim_{k \rightarrow \infty} \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^3 \gamma_k} (1 + \frac{1}{\gamma_k \eta_k}) = 0$; (f) $\lim_{k \rightarrow \infty} \frac{\gamma_k}{\eta_k} E[\|w_k\|^2 | \mathcal{F}_k] = 0$.

We also make use Lemma 11, Pg. 50 from the monograph by Polyak (1987).

Lemma 1 Let $\{v_k\}$ be a sequence of nonnegative random variables, where $E[v_0] < \infty$, and let $\{\alpha_k\}$ and $\{\beta_k\}$ be deterministic scalar sequences such that:

$$\begin{aligned} E[v_{k+1}|v_0, \dots, v_k] &\leq (1 - \alpha_k)v_k + \beta_k \quad \text{a.s for all } k \geq 0, \\ 0 \leq \alpha_k \leq 1, \quad \beta_k &\geq 0, \quad \sum_{k=0}^{\infty} \alpha_k = \infty, \quad \sum_{k=0}^{\infty} \beta_k < \infty, \quad \lim_{k \rightarrow \infty} \frac{\beta_k}{\alpha_k} = 0. \end{aligned}$$

Then, $v_k \rightarrow 0$ almost surely.

We let \mathcal{F}_k denote the history of the method up to time k , i.e., $\mathcal{F}_k = \{x_0, \xi_0, \xi_1, \dots, \xi_{k-1}\}$ for $k \geq 1$ and $\mathcal{F}_0 = \{x_0\}$. The following Lemma is a combined result of Lemma 3 and Proposition 1 from Koshal et al. (2010).

Lemma 2 Assume that $\text{SOL}(X, F)$ is nonempty and $X \in \mathbb{R}^n$ be closed and convex. Furthermore, let the map $F : X \rightarrow \mathbb{R}^n$ be continuous and monotone over X . Consider the Tikhonov sequence $\{y_k\}$ for $\text{VI}(X, F)$, i.e., $\{y_k\}$ is the sequence of exact solution to $\text{VI}(X, F + \eta_k \mathbf{I})$, $k \geq 0$, with $\eta_k > 0$ for all k . Then

(a) For all $k \geq 1$ we have

$$\|y_k - y_{k-1}\| \leq M_y \frac{|\eta_{k-1} - \eta_k|}{\eta_k}, \quad (2)$$

where M_y is a norm bound on the Tikhonov sequence, i.e., $\|y_k\| \leq M_y$ for all $k \geq 0$.

(b) For all $k \geq 1$ we have

$$\begin{aligned} E[\|x_{k+1} - y_k\|^2 | \mathcal{F}_k] &\leq q_k(1 + \gamma_k \eta_k) \|x_k - y_{k-1}\|^2 \\ &+ q_k M_y^2 \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^2} \left(1 + \frac{1}{\gamma_k \eta_k}\right) + \gamma_k^2 E[\|w_k\|^2 | \mathcal{F}_k], \end{aligned} \quad (3)$$

where $q_k = 1 - 2\gamma_k \eta_k + \gamma_k^2 (\eta_k + L)^2$.

In the remainder of this subsection, we prove convergence of Tikhonov algorithm under weaker assumptions than shown by Koshal et al. (2010). More specifically, we weaken the assumptions $\lim_{k \rightarrow \infty} \frac{\gamma_k}{\eta_k} (\eta_k + L)^2 = 0$ and $\lim_{k \rightarrow \infty} \gamma_k \eta_k = 0$ by part (a) of Assumption 2.

Proposition 1 Let Assumptions 1 and 2 hold. Also, assume that $\text{SOL}(X, F)$ is nonempty. Then, the sequence $\{x_k\}$ generated by iterative Tikhonov scheme (2) converges to the least-norm solution x^* of $\text{VI}(X, F)$ almost surely and for $k \geq 1$ we have

$$E[\|x_{k+1} - y_k\|^2 | \mathcal{F}_k] \leq \left(1 - \frac{\eta_k}{2} \gamma_k\right) \|x_k - y_{k-1}\|^2 + q_k M_y^2 \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^2} \left(1 + \frac{1}{\gamma_k \eta_k}\right) + \gamma_k^2 E[\|w_k\|^2 | \mathcal{F}_k], \quad (4)$$

where $q_k = 1 - 2\gamma_k \eta_k + \gamma_k^2 (\eta_k + L)^2$.

Proof. From Lemma 2, we have

$$E[\|x_{k+1} - y_k\|^2 | \mathcal{F}_k] \leq q_k(1 + \gamma_k \eta_k) \|x_k - y_{k-1}\|^2 + q_k M_y^2 \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^2} \left(1 + \frac{1}{\gamma_k \eta_k}\right) + \gamma_k^2 E[\|w_k\|^2 | \mathcal{F}_k].$$

Next, we estimate the coefficient $q_k(1 + \eta_k \gamma_k)$. From part (a) of Assumption 2 for $k \geq 0$ we have

$$\frac{(\eta_k + L)^2}{\eta_k} \gamma_k < \frac{1}{2} < 2 \Rightarrow -2\eta_k \gamma_k + \gamma_k^2 (\eta_k + L)^2 < 0 \Rightarrow q_k < 1.$$

Therefore, using this and the definition of q_k we obtain

$$q_k(1 + \eta_k \gamma_k) = q_k + \eta_k \gamma_k q_k < q_k + \eta_k \gamma_k = 1 - \eta_k \gamma_k + \gamma_k^2 (\eta_k + L)^2.$$

Obviously, when $0 < \gamma_k < \frac{\eta_k}{2(\eta_k + L)^2}$, we have

$$1 - \eta_k \gamma_k + \gamma_k^2 (\eta_k + L)^2 \leq 1 - u_k,$$

where $u_k \triangleq \frac{\eta_k}{2} \gamma_k$. Therefore, we obtain

$$\mathbb{E}[\|x_{k+1} - y_k\|^2 \mid \mathcal{F}_k] \leq (1 - \frac{\eta_k}{2} \gamma_k) \|x_k - y_{k-1}\|^2 + q_k M_y \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^2} (1 + \frac{1}{\gamma_k \eta_k}) + \gamma_k^2 \mathbb{E}[\|w_k\|^2 \mid \mathcal{F}_k]. \quad (5)$$

To claim the convergence of the sequence $\|x_{k+1} - y_k\|$, we show that Lemma 1 applies to relation (5). Let us define v_k as

$$v_k \triangleq q_k M_y^2 \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^2} (1 + \frac{1}{\gamma_k \eta_k}) + \gamma_k^2 \mathbb{E}[\|w_k\|^2 \mid \mathcal{F}_k].$$

Therefore we have

$$\mathbb{E}[\|x_{k+1} - y_k\|^2 \mid \mathcal{F}_k] \leq (1 - u_k) \|x_k - y_{k-1}\|^2 + v_k, \quad \text{for all } k > K,$$

and it suffices to examine a shifted sequence. We need to verify that u_k and v_k satisfy the conditions of Lemma 1 for $k \geq 0$. Note that from part (a) of Assumption 2 for $k \geq 0$, $u_k < 1$ since for any $L > 0$, we have $0 < \gamma_k < \frac{\eta_k}{2(\eta_k + L)^2} < \frac{2}{\eta_k}$. Assumption 2 (c) implies that $\sum_{k=K}^{\infty} u_k = \infty$. Next, we consider the ratio $\frac{v_k}{u_k}$ and observe that $\frac{v_k}{u_k} \geq 0$ for $k \geq 0$. We show that its upper bound converges to zero. Since $0 < q_k < 1$, the ratio $\frac{v_k}{u_k}$ can be bounded as follows

$$\begin{aligned} \frac{v_k}{u_k} &\leq \frac{2}{\eta_k \gamma_k} M_y^2 \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^2} (1 + \frac{1}{\gamma_k \eta_k}) + \frac{2\gamma_k^2}{\eta_k \gamma_k} \mathbb{E}[\|w_k\|^2 \mid \mathcal{F}_k] \\ &= M_y^2 \frac{2(\eta_{k-1} - \eta_k)^2}{\eta_k^3 \gamma_k} (1 + \frac{1}{\gamma_k \eta_k}) + \frac{2\gamma_k}{\eta_k} \mathbb{E}[\|w_k\|^2 \mid \mathcal{F}_k]. \end{aligned} \quad (6)$$

By Assumption 2 (e) and (f), the above two terms converge to zero, implying that $\lim_{k \rightarrow \infty} \frac{v_k}{u_k} = 0$. Finally, by Assumption 1 (c) and Assumption 2 (d), $\sum_{k=K}^{\infty} v_k < \infty$. It follows from Lemma 1 that $\|x_k - y_{k-1}\|$ converges to zero almost surely. This and the fact that Tikhonov sequence converges to the least-norm solution x^* of $VI(X, F)$ imply that $x_k \rightarrow x^*$ almost surely. \square

Remark on choices of γ_k and η_k : While Assumption 2 appears rather difficult to satisfy, Koshal et al. (2010), show that a stronger form of Assumption 2 is satisfied by $\gamma_k = k^{-a}$ and $\eta_k = k^{-b}$ with $a + b < 1$ and $a > b$. We conclude with a corollary that relies on the following assumption on the errors.

Assumption 3 The errors w_k are such that for some $v > 0$ $\mathbb{E}[\|w_k\|^2 \mid \mathcal{F}_k] \leq v^2$ a.s. for all $k \geq 0$.

The following corollary gives a parametric upper bound for the error; this result is applied to obtain the recursive scheme in next part.

Corollary 1 Let Assumptions 1, 2, and 3 hold. Also, assume that $SOL(X, F)$ is nonempty. Then, the sequence $\{x_k\}$ generated by iterative Tikhonov scheme (2) converges to the least-norm solution x^* of $VI(X, F)$ almost surely and for $k \geq 1$ we have

$$\mathbb{E}[\|x_{k+1} - y_k\|^2] \leq (1 - \frac{\eta_k}{2} \gamma_k) \mathbb{E}[\|x_k - y_{k-1}\|^2] + q_k M_y^2 \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^2} (1 + \frac{1}{\gamma_k \eta_k}) + \gamma_k^2 v^2, \forall k \geq 1, \quad (7)$$

where $q_k = 1 - 2\gamma_k \eta_k + \gamma_k^2 (\eta_k + L_k)^2$.

Proof. The convergence result follows from Proposition 1 while (7) follows from (4) by taking expectations and invoking Assumption 3. \square

2.2 A Regularized Recursive Steplength SA Scheme

A challenge associated with the implementation of diminishing steplength schemes lies in determining an appropriate sequence $\{\gamma_k\}$. The key result of this section is the motivation and introduction of a scheme that *adaptively* updates the steplength across successive iterations; such a rule is derived from the minimization of a suitably defined error function at each step. Let us view the quantity $\mathbb{E}[\|x_{k+1} - y_k\|^2]$ as an error, denoted by e_{k+1} , and arising from the use of the stepsize values $\gamma_0, \gamma_1, \dots, \gamma_k$. Thus, in the worst case, the error satisfies the following recursive relation for any $k \geq 1$:

$$e_{k+1}(\gamma_1, \dots, \gamma_k) = \left(1 - \frac{\eta_k}{2}\gamma_k\right)e_k(\gamma_1, \dots, \gamma_{k-1}) + q_k M_y^2 \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^2} \left(1 + \frac{1}{\gamma_k \eta_k}\right) + v^2 \gamma_k^2,$$

where $q_k = 1 - 2\gamma_k \eta_k + \gamma_k^2 (\eta_k + L_k)^2$, e_1 is a positive scalar, η_k is a positive regularization parameter and v^2 is the upper bound for the second moments of the error norms $\|w_k\|$. Then, it seems natural to investigate if the stepsizes $\gamma_1, \gamma_1 \dots, \gamma_k$ can be selected so as to minimize the error e_{k+1} . It turns out that this can indeed be achieved at each iteration. Let us now define the following parameters:

$$\begin{aligned} M_k &\triangleq M_y^2 \frac{(\eta_{k-1} - \eta_k)^2}{\eta_k^2}, & a_k &\triangleq \frac{\eta_k}{2}, & b_k &\triangleq \frac{M_k}{\eta_k}, & c_k &\triangleq -M_k, \\ d_k &\triangleq M_k \left(\frac{(\eta_k + L)^2}{\eta_k} - 2\eta_k \right), & f_k &\triangleq M_k (\eta_k + L)^2 + v^2. \end{aligned} \quad (8)$$

Then, (2.2) may be rewritten as follows for $k \geq 1$:

$$e_k(\gamma_1, \dots, \gamma_{k-1}) = (1 - a_{k-1}\gamma_{k-1})e_{k-1}(\gamma_1, \dots, \gamma_{k-2}) + \frac{b_{k-1}}{\gamma_{k-1}} + c_{k-1} + d_{k-1}\gamma_{k-1} + f_{k-1}\gamma_{k-1}^2. \quad (9)$$

The presentation of our recursive steplength scheme is significantly simplified by defining a function $D_k : \mathbb{R}^+ \rightarrow \mathbb{R}$ such that

$$D_k(t) \triangleq \frac{1}{a_k} \left(-\frac{b_k}{t^2} + d_k + 2f_k t \right). \quad (10)$$

Regularized recursive steplength SA (RRSA) scheme: For $k \geq 1$, given an $e_1 > 0$, the RRSA scheme generates a sequence $\{x_k\}$ where x_{k+1} is updated as per (1), $\{\eta_k\}$ satisfies Assumption 2 and γ_k satisfies

$$D_1(\gamma_1) - e_1 = 0, \quad (11)$$

$$D_{k+1}(\gamma_{k+1}) = (1 - a_k \gamma_k) D_k(\gamma_k) + \frac{b_k}{\gamma_k} + c_k + d_k \gamma_k + f_k \gamma_k^2. \quad (12)$$

For $k \geq 1$, the above equations lead to a polynomial function of the third degree with respect to γ_k . We assume that at each iterate, this equation has a positive real root denoted by γ_k . Furthermore, we often refer to $e_k(\gamma_1, \dots, \gamma_{k-1})$ by e_k whenever this is unambiguous. In our main result of this subsection, we show that the stepsizes $\gamma_i, i = 1, \dots, k-1$, minimize the errors e_k over the range that $0 < \gamma_k < \frac{\eta_k}{2(\eta_k + L)^2}$ for $k \geq 1$, where L is the Lipschitz constant associated with the mapping $F(x)$.

Proposition 2 Let $e_k(\gamma_1, \dots, \gamma_{k-1})$ be defined as in (2.2), where $e_1 > 0$ is such that (11) has a real positive root as γ_1^* , and we assume that for each $k \geq 1$, equation (12) has a real positive root as γ_k^* . Let the sequence $\{\gamma_k^*\}$ be given by (11)–(12). Suppose that the sequence $\{\eta_k\}_{k=0}^\infty$ is nonincreasing and fixed. Then, the following hold:

- (a) The error e_k satisfies $e_k(\gamma_1^*, \dots, \gamma_{k-1}^*) = \frac{2}{\eta_k} \left(-\frac{b_k}{\gamma_k^{*2}} + d_k + 2f_k \gamma_k^* \right)$ for all $k \geq 1$, where b_k, d_k , and f_k are defined in (8).

(b) For each $k \geq 1$, the vector $(\gamma_1^*, \gamma_1^*, \dots, \gamma_k^*)$ minimizes $e_{k+1}(\gamma_1, \dots, \gamma_k)$ over the set

$$\mathbb{G}_k \triangleq \left\{ \alpha \in \mathbb{R}^k : 0 < \alpha_j < \frac{\eta_j}{2(\eta_j + L)^2} \text{ for } j = 1, \dots, k \right\}.$$

Additionally, for any $k \geq 2$ and any $(\gamma_1, \dots, \gamma_{k-1}) \in \mathbb{G}_{k-1}$, we have

$$e_k(\gamma_1, \dots, \gamma_{k-1}) - e_k(\gamma_1^*, \dots, \gamma_{k-1}^*) \geq \left(\frac{b_k}{\gamma_k^{*2}} + f_k \right) (\gamma_k - \gamma_k^*)^2 \geq 0.$$

Proof. (a) We use induction on k to prove our result. Note that the result holds trivially for $k = 1$ from (11) and definition of function D_1 . Next, assume that we have $e_k(\gamma_1^*, \dots, \gamma_{k-1}^*) = \frac{1}{a_k}(-\frac{b_k}{\gamma_k^*} + d_k + 2f_k\gamma_k^*)$ for some k , and consider the case for $k + 1$. By the definition of the error e_k in (2.2), we have

$$\begin{aligned} e_{k+1}(\gamma_1^*, \dots, \gamma_k^*) &= (1 - a_k\gamma_k^*)e_k(\gamma_1^*, \dots, \gamma_{k-1}^*) + \frac{b_k}{\gamma_k^*} + c_k + d_k\gamma_k^* + f_k\gamma_k^{*2} \\ &= (1 - a_k\gamma_k^*)D_k(\gamma_k^*) + \frac{b_k}{\gamma_k^*} + c_k + d_k\gamma_k^* + f_k\gamma_k^{*2} \\ &= D_{k+1}(\gamma_{k+1}^*) = \frac{1}{a_{k+1}}\left(-\frac{b_{k+1}}{\gamma_{k+1}^*} + d_{k+1} + 2f_{k+1}\gamma_{k+1}^*\right), \end{aligned}$$

where the second equality follows by the inductive hypothesis and definition of D_k in (10), the third inequality follows by (12), and the last inequality follows by definition of D_{k+1} in (10).

(b) We now show that $(\gamma_1^*, \gamma_2^*, \dots, \gamma_{k-1}^*)$ minimizes the error e_k for all $k \geq 2$. We again use mathematical induction on k . By the definition of the error e_2 , we have

$$e_2(\gamma_1) - e_2(\gamma_1^*) = (1 - a_1\gamma_1)e_1 + \frac{b_1}{\gamma_1} + c_1 + d_1\gamma_1 + f_1\gamma_1^2 - (1 - a_1\gamma_1^*)e_1 - \frac{b_1}{\gamma_1^*} - c_1 - d_1\gamma_1^* - f_1\gamma_1^{*2}.$$

Using (11) and definition of D_1 in (10) we have

$$\begin{aligned} e_2(\gamma_1) - e_2(\gamma_1^*) &= a_1(\gamma_1^* - \gamma_1)\left(\frac{1}{a_1}\right)\left(-\frac{b_1}{\gamma_1^*} + d_1 + 2f_1\gamma_1^*\right) + b_1\left(\frac{1}{\gamma_1} - \frac{1}{\gamma_1^*}\right) + d_1(\gamma_1 - \gamma_1^*) + f_1(\gamma_1^2 - \gamma_1^{*2}) \\ &= -\frac{b_1}{\gamma_1^*} + d_1\gamma_1^* + 2f_1\gamma_1^{*2} + \frac{b_1}{\gamma_1^2}\gamma_1 - d_1\gamma_1 - 2f_1\gamma_1^*\gamma_1 + b_1\left(\frac{1}{\gamma_1} - \frac{1}{\gamma_1^*}\right) + d_1(\gamma_1 - \gamma_1^*) + f_1(\gamma_1^2 - \gamma_1^{*2}) \\ &= b_1\left(1 - \frac{\gamma_1}{\gamma_1^*}\right)\left(\frac{1}{\gamma_1} - \frac{1}{\gamma_1^*}\right) + f_1(\gamma_1^2 + \gamma_1^{*2} - 2\gamma_1^*\gamma_1) = \frac{b_1(\gamma_1^* - \gamma_1)^2}{\gamma_1^2\gamma_1} + f_1(\gamma_1 - \gamma_1^*)^2 \geq 0, \end{aligned}$$

where the last inequality follows by positiveness of b_1 and f_1 . Now suppose that $e_k(\gamma_1, \dots, \gamma_{k-1}) \geq e_k(\gamma_1^*, \dots, \gamma_{k-1}^*)$ holds for some k and any $(\gamma_1, \dots, \gamma_{k-1}) \in \mathbb{G}_k$. We want to show that $e_{k+1}(\gamma_1, \dots, \gamma_k) \geq e_{k+1}(\gamma_1^*, \dots, \gamma_k^*)$ holds as well for all $(\gamma_1, \dots, \gamma_k) \in \mathbb{G}_{k+1}$. To simplify the notation we use e_{k+1}^* to denote the error e_{k+1} evaluated at $(\gamma_1^*, \gamma_1^*, \dots, \gamma_k^*)$, and e_{k+1} when evaluating at an arbitrary vector $(\gamma_1, \gamma_1, \dots, \gamma_k) \in \mathbb{G}_{k+1}$. Using (2.2) we have

$$\begin{aligned} e_{k+1} - e_{k+1}^* &= (1 - a_k\gamma_k)e_k + \frac{b_k}{\gamma_k} + c_k + d_k\gamma_k + f_k\gamma_k^2 - (1 - a_k\gamma_k^*)e_k^* - \frac{b_k}{\gamma_k^*} - c_k - d_k\gamma_k^* - f_k\gamma_k^{*2} \\ &\geq (1 - a_k\gamma_k)e_k^* + \frac{b_k}{\gamma_k} + c_k + d_k\gamma_k + f_k\gamma_k^2 - (1 - a_k\gamma_k^*)e_k^* - \frac{b_k}{\gamma_k^*} - c_k - d_k\gamma_k^* - f_k\gamma_k^{*2} \end{aligned}$$

where the last inequality follows by the induction hypothesis and the observation that $(1 - a_k \gamma_k) > 0$ from the feasibility of the sequence $\{\gamma_j\}_{j=1}^k$ with respect to \mathbb{G}_k . Using (12) and the definition of D_k in (10),

$$\begin{aligned} e_{k+1} - e_{k+1}^* &= a_k(\gamma_k^* - \gamma_k) \left(\frac{1}{a_k} \right) \left(-\frac{b_k}{\gamma_k^*} + d_k + 2f_k \gamma_k^* \right) + b_k \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_k^*} \right) + d_k(\gamma_k - \gamma_k^*) + f_k(\gamma_k^2 - \gamma_k^{*2}) \\ &= -\frac{b_k}{\gamma_k^*} + d_k \gamma_k^* + 2f_k \gamma_k^{*2} + \frac{b_k}{\gamma_k^{*2}} \gamma_k - d_k \gamma_k - 2f_k \gamma_k^* \gamma_k + b_k \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_k^*} \right) + d_k(\gamma_k - \gamma_k^*) + f_k(\gamma_k^2 - \gamma_k^{*2}) \\ &= b_k \left(1 - \frac{\gamma_k}{\gamma_k^*} \right) \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_k^*} \right) + f_k(\gamma_k^2 + \gamma_k^{*2} - 2\gamma_k^* \gamma_k) = \frac{b_k(\gamma_k^* - \gamma_k)^2}{\gamma_k^{*2} \gamma_k} + f_k(\gamma_k - \gamma_k^*)^2 \geq 0, \end{aligned}$$

where the last inequality follows by positiveness of b_k and f_k . Hence, we have $e_k(\gamma_1, \dots, \gamma_{k-1}) - e_k(\gamma_1^*, \dots, \gamma_{k-1}^*) \geq \nu^2(\gamma_k - \gamma_k^*)^2$ for all $k \geq 2$ and all $(\gamma_1, \dots, \gamma_{k-1}) \in \mathbb{G}_k$. Therefore, for all $k \geq 2$, the vector $(\gamma_1, \dots, \gamma_{k-1}) \in \mathbb{G}_k$ is a minimizer of the error e_k . \square

2.3 Convergence of RRSA Scheme

In this part, we show that the RRSA scheme converges in expectation to the solution for a fixed choice of regularization sequence by employing the following result proved by Koshal et al. (2010).

Proposition 3 Let Assumptions 1(a)-(b), and 3 hold and suppose that $\text{SOL}(X, F)$ is nonempty. Consider the choice $\eta_k = k^{-b}$ and $\gamma_k = k^{-a}$ for all k , where $a, b \in (0, 1)$, $a + b < 1$, and $a > b$. Then, this choice satisfies Assumptions 1(c), and 2 and the sequence $\{x_k\}$ generated by iterative Tikhonov scheme (2) converges to the least-norm solution x^* of $\text{VI}(X, F)$ almost surely.

Proposition 4 Let Assumptions 1(a)-(b), and 3 hold and suppose that $\text{SOL}(X, F)$ is nonempty. Consider the choice $\eta_k = k^{-b}$ for all k , where $b \in (0, 1/2)$ and suppose that $\{\gamma_k^*\}_{k=1}^\infty$ is given by the adaptive scheme (11)-(12). Then, the sequence $\{x_k\}$ generated by iterative Tikhonov scheme (2) converges in expectation to the least-norm solution x^* of $\text{VI}(X, F)$.

Proof. Let $\tilde{\gamma}_k = k^{-a}$ for all k , where $a + b < 1$ and $b < a$. Now suppose that $\{\bar{x}_k\}$ is generated by the iterative Tikhonov scheme (2) with $\{\tilde{\gamma}_k\}$ and $\{\eta_k\}$. Then from Lemma 3, Assumptions 1(c), and 2 are satisfied by this choice of $\{\eta_k\}$ and $\{\tilde{\gamma}_k\}$ and therefore by Proposition 1, $\mathbb{E}[\|\bar{x}_{k+1} - y_k\|^2]$ goes to zero where $\{y_k\}$ is the sequence of exact solution to $\text{VI}(X, F + \eta_k \mathbf{I})$ for $k \geq 0$. This implies that the upper bound sequence defined by (2.2) goes to zero. Now, from Proposition 2(b), we know that for the sequence $\{x_k\}$ generated by the iterative Tikhonov scheme (2) with $\{\gamma_k^*\}$ and $\{\eta_k\}$ the following inequality holds for $k \geq 0$:

$$e_k(\gamma_1, \dots, \gamma_{k-1}) - e_k(\gamma_1^*, \dots, \gamma_{k-1}^*) \geq 0. \tag{13}$$

Since $\mathbb{E}[\|\bar{x}_{k+1} - y_k\|^2] - \mathbb{E}[\|x_{k+1} - y_k\|^2] \geq 0$ and by noting that $\mathbb{E}[\|\bar{x}_{k+1} - y_k\|^2] \rightarrow 0$ when k goes to infinity, we conclude that the sequence $\{x_k\}$ generated by the RRSA scheme converges in expectation to the least-norm solution x^* of $\text{VI}(X, F)$. \square

Remark on almost-sure convergence: It is worth noting that almost-sure convergence of the estimators may be proved by showing that the sequence $\{\gamma_k\}$ satisfies the requirements of Lemma 1. This requires a deeper analysis of the roots and will not be pursued further, given the size restrictions of this paper.

3 AN ITERATIVE SMOOTHING RECURSIVE SA SCHEME

Yousefian, Nedić, and Shanbhag (2011) considered stochastic optimization problems with nonsmooth integrands and inspired by Lakshmanan and Farias (2008), employed a *random local smoothing* of the objective in constructing a stochastic approximation scheme. In fact, a more extensive study of the literature

revealed that such smoothing techniques had been employed extensively in the past and were referred to as Steklov-Sobolev smoothing techniques. Succinctly, such smoothing approaches led to an approximation that was shown to admit Lipschitzian properties. In fact, the growth rate (with problem size) of the Lipschitz constant associated with the gradient of the objective may be quantified under different types of smoothing distributions, as derived by Yousefian, Nedić, and Shanbhag (2011) and Lakshmanan and Farias (2008).

In this section, we make two extensions to our earlier work. First, we extend the approach to address stochastic variational inequalities which may arise from either stochastic optimization problems or Nash games. Second, while our earlier approach obtained an approximate solution since a small, but fixed, smoothing parameter was employed, we consider how this smoothing parameter may be reduced after every iteration, leading to an iterative smoothing scheme, with the intent of obtaining the *true* solution. For purposes of simplicity, we assume that the original mapping is strongly monotone, implying that there is no need for employing a regularization term.

3.1 Nonsmooth Stochastic Optimization Problems and Nash games

We motivate our approach by considering a nonsmooth stochastic Nash game. Consider an N -player Nash game in which the i th player solves the following problem, given $x_{-i} \triangleq (x_j)_{j \neq i \in \{1, \dots, N\}}$:

$$\text{Ag}(x_{-i}) \quad \min_{x_i \in X_i} \mathbb{E}[f_i(x_i; x_{-i}, \xi)], \tag{14}$$

where for $i = 1, \dots, N$, $X_i \subseteq \mathbb{R}^{n_i}$ is a closed and convex set, $\sum_{i=1}^N n_i = n$ and $f_i : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}$ is a continuous convex function for all $x_{-i} \in \prod_{j \neq i} X_j$. Recall that a Nash equilibrium is given by a tuple $\{x_i^*\}_{i=1}^N$ such that x_i^* solves $\text{Ag}(x_{-i}^*)$ for $i = 1, \dots, N$. Then the resulting Nash equilibrium is given by a solution to a multi-valued variational inequality $\text{VI}(X, \partial F)$ where $X \triangleq \prod_{i=1}^N X_i$ and $\partial F(x) \triangleq \partial_{x_i} \mathbb{E}[f_i(x; \xi)]$ for $i = 1, \dots, N$. Note that if $f_i(x; \xi)$ is differentiable in x_i , given x_{-i} , $\text{VI}(X, \partial F)$ reduces to $\text{VI}(X, F)$ where $F(x) = (\nabla_{x_i} \mathbb{E}[f_i(x; \xi)])_{i=1}^N$ is a single-valued mapping. Clearly, if $N = 1$, then this game reduces to a stochastic optimization problem.

3.2 An Iterative Smoothing Extension of the RSA Scheme

Consider a *smoothed* approximation of the stochastic Nash game in which the i th player solves the following *smoothed* problem:

$$\min_{x_i \in X_i} \mathbb{E}[\mathbb{E}[f_i(x_i + z_i; x_{-i}, \xi) \mid \xi]], \tag{15}$$

where the inner expectation is with respect to $z_i \in \mathbb{R}^{n_i}$, a random vector with a compact support. Then $z \triangleq (z_i)_{i=1}^N \in \mathbb{R}^n$ is an n -dimensional random vector with a probability distribution over the n -dimensional ball centered at the origin and with radius ε . If \hat{f}_i is defined as $\hat{f}_i \triangleq \mathbb{E}\{f_i(x_i + z_i; x_{-i}, \xi) \mid \xi\}$, then $\hat{F}(x) \triangleq (\nabla_{x_i} \hat{f}_i)_{i=1}^N$. For the mapping \hat{F} to be well defined, we need to enlarge the underlying set X so that the mapping $F(x+z)$ is defined for every $x \in X$. In particular, for a set $X \subseteq \mathbb{R}^n$ and $\varepsilon > 0$, we let X_ε be the set defined by $X_\varepsilon = \{y \mid y = x + z, x \in X, z \in \mathbb{R}^n, \|z\| \leq \varepsilon\}$. We employ a uniform distribution for purposes of smoothing but a normal distribution may also work, as considered by Lakshmanan and Farias (2008). However, distributions with finite support seem more appropriate for capturing local behavior of a function, as well as to deal with the problems where the function itself has a restricted domain. Our choice lends itself readily for computation of the resulting Lipschitz constant and for assessment of the growth of the Lipschitz constant with the size of the problem. Suppose $z \in \mathbb{R}^n$ is a random vector with uniform distribution over the n -dimensional ball centered at the origin and with a radius ε , i.e., z has the following probability density function:

$$p_u(z; \varepsilon) = \begin{cases} \frac{1}{c_n \varepsilon^n} & \text{for } \|z\| \leq \varepsilon, \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

where $c_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)}$ and Γ is the gamma function given by

$$\Gamma\left(\frac{n}{2} + 1\right) = \begin{cases} \left(\frac{n}{2}\right)! & \text{if } n \text{ is even,} \\ \sqrt{\pi} \frac{n!!}{2^{(n+1)/2}} & \text{if } n \text{ is odd,} \end{cases}$$

where $!!$ is the double factorial symbol. We consider a stochastic approximation scheme in which the parameter specifying the size of the support of the smoothing distribution is reduced after every iteration:

$$\begin{aligned} x_{k+1} &= \Pi_X(x_k - \gamma_k(\hat{F}(x_k) + w_k)) \quad \text{for all } k \geq 0, \\ w_k &= s_k - \hat{F}(x_k), \quad \text{where } s_k \in \prod_{i=1}^N \partial_{x_i} f_i(x_i + z_i^k; x_{-i}, \xi), \end{aligned}$$

where z_k is drawn from a uniform distribution $p_u(z; \varepsilon_k)$ and $\varepsilon_k \rightarrow 0$. In effect, the step computation is modified by a randomness drawn from a uniform distribution with steadily decreasing support. While this seems intuitive from the standpoint of recovering convergence, the rate at which ε_k is reduced remains crucial from the standpoint of recovering convergence. Yousefian, Nedić, and Shanbhag (2011) derive a growth property for the Lipschitz constant associated with the smoothed gradient of the optimization problem; in particular, we show that the Lipschitz constant is of the form

$$\|\hat{F}(x) - \hat{F}(y)\| \leq \left(\kappa \frac{n!!}{(n-1)!!} \frac{C}{\varepsilon} \right) \|x - y\| \quad \text{for all } x, y \in X,$$

where $\kappa = \frac{2}{\pi}$ if n is even and $\kappa = 1$ otherwise. Therefore, as $\varepsilon \rightarrow 0$, the Lipschitz constant grows to infinity. Yet, if this growth rate is sufficiently small, one may recover almost-sure convergence. The crux of the proof lies in showing that the recursive steplength rule is feasible with respect to the Lipschitz constants which are growing in each step. More specifically, \mathbb{G}_k requires that $0 < \gamma_j < \frac{\eta}{2(\eta + L(\varepsilon_j))^2}$ for $j = 1, \dots, k$. Under this modified specification of \mathbb{G}_k , the optimality of the steplength choices would need to be established. Almost-sure convergence follows if this steplength sequence satisfies $\sum_k \gamma_k^2 < \infty$ and $\sum_k \gamma_k = \infty$; verifying these requirements is a focus of ongoing research.

4 NUMERICAL RESULTS

In this section, we present some preliminary numerical results detailing the performance of the proposed schemes. In Section 4.1, we consider a stochastic network utility problem and compare the performance of the proposed RRSA scheme with an ITR scheme and a standard SA implementation with stepsizes given by $\gamma_k = 1/k$. A sensitivity analysis is also conducted for different values of problem parameters in an effort to gauge the stability of the performance of each scheme. Nonsmooth stochastic variational problems are examined in Section 4.2 where an iterative smoothing counterpart of the RSA scheme is examined. Note that the computational results were developed on Matlab 7.0 on a Linux OS. Furthermore, the true solutions were computed by solving a sample-average approximation (SAA) problem.

4.1 A Stochastic Network Utility Problem

We consider a spatial network with n users competing over L_1 links. Suppose the i th user's utility function is denoted by $\xi_i \log(1 + x_i)$. Additionally, a congestion cost is imposed on every user of the form $c(x) = \|Ax\|^2$ where A is defined as the adjacency matrix with binary elements that specifies the set of links traversed by the traffic generated by a particular user. We assume that the user traffic rates are restricted by capacity constraints $\sum_{i=1}^n A_{li}x_i \leq C_l$ for every link l where C_l is the aggregate traffic through link l . The resulting equilibrium conditions are given by a stochastic monotone variational inequality.

Employing the same starting point, we generate 100 replications for each scheme and compare the average error norm with respect to the reference solution as computed from the solution of an SAA problem. Figure 1 shows the trajectory of the mean error for the ITR, RRSA, and HSA schemes. Note that the x -axis denotes the iteration number while the y -axis denotes the logarithm of the averaged error over a 100 replications. In other words, for a fixed problem and a fixed scheme, we run the simulation 100 times and averaged the terms $\|x_k - x^*\|$ where x^* is the solution of $VI(X, F)$ from solving an SAA problem and $1 \leq k \leq N$. Here, we assume that $C_1 = (0.10, 0.15, 0.20, 0.10, 0.15, 0.20, 0.20, 0.15, 0.25) = 0.1C_2 = 0.01C_3$ and x is constrained to be nonnegative. We also assume that ξ_i is a uniform random variable between zero and β for any i where β is a positive parameter. Table 1 displays the 90% confidence intervals for nine parameter settings, categorized into three groups. In the first group, the sensitivity of the results to changing N is examined while the second and third groups investigate the impact of changing β and C .

Insights: We observe that the adaptive scheme performs favorably in comparison with the other two schemes from several standpoints. From Figure 1, we observe that the RRSA scheme tends to perform well across three different problem settings while the performance of a standard SA implementation is characterized by tremendous variability. In fact, as seen in Figures 1a and 1b, the HSA scheme performs poorly. It should be emphasized that the standard HSA scheme is not guaranteed to converge for merely monotone variational problems but is often a de-facto choice in simulation-based optimization. Furthermore, when HSA does outperform RRSA (as in Figure 1c), it does not contradict the optimality of adaptive stepsizes since the HSA scheme is not part of the feasible set of sequences that RRSA is optimized over. Table 1 shows the sensitivity of the three schemes to changes in parameters. One can immediately see that the RRSA scheme produces iterates with lower average error and is relatively robust to parametric changes. The HSA scheme, on the contrary, is extremely sensitive to such modifications, particularly when they arise in the form of a change in β .

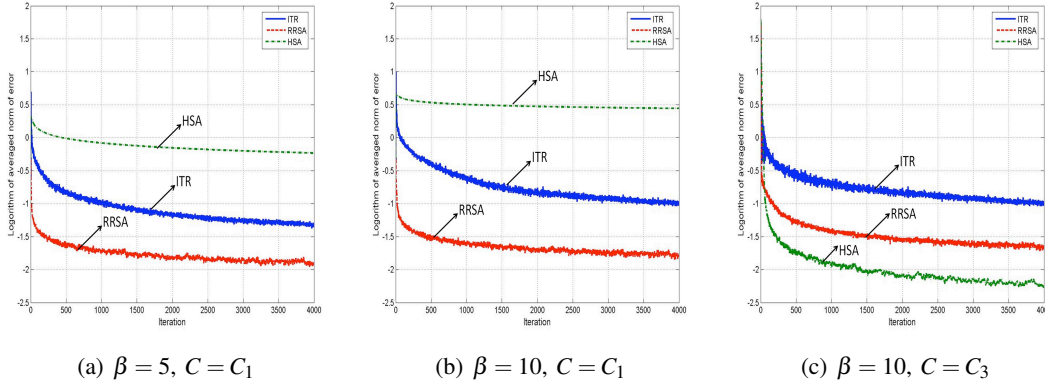


Figure 1: Trajectories of average error for the stochastic network utility problem.

Table 1: Parametric Sensitivity of Confidence Intervals.

-	$P(i)$	N	β	C	ITR - 90% CI	RRSA - 90% CI	HSA - 90% CI
N	1	4000	5	C_1	$[4.34e-2, 5.05e-2]$	$[1.10e-2, 1.32e-2]$	$[5.01e-1, 6.69e-1]$
	2	2000	5	C_1	$[5.93e-2, 6.99e-2]$	$[1.45e-2, 1.71e-2]$	$[6.09e-1, 7.93e-1]$
	3	1000	5	C_1	$[9.58e-2, 1.11e-1]$	$[1.64e-2, 1.93e-2]$	$[7.31e-1, 9.30e-1]$
β	4	4000	10	C_1	$[9.79e-2, 1.16e-1]$	$[1.44e-2, 1.71e-2]$	$[2.52e-0, 3.04e-0]$
	5	4000	10	C_2	$[9.79e-2, 1.16e-1]$	$[1.53e-2, 1.82e-2]$	$[5.10e-3, 6.30e-3]$
	6	4000	10	C_3	$[9.79e-2, 1.16e-1]$	$[1.88e-2, 2.23e-2]$	$[5.10e-3, 6.30e-3]$
C	7	4000	20	C_1	$[2.88e-1, 3.23e-1]$	$[1.87e-2, 2.22e-2]$	$[7.30e-0, 8.50e-0]$
	8	4000	50	C_1	$[1.83e-1, 1.91e-1]$	$[2.75e-2, 3.25e-2]$	$[2.25e+1, 2.57e+1]$
	9	4000	100	C_1	$[4.62e-0, 4.75e-0]$	$[3.85e-2, 4.48e-2]$	$[4.80e+1, 5.45e+1]$

4.2 A Nonsmooth Stochastic Optimization Problem

Next, we examine the following nonsmooth optimization problem:

$$\min_{x \in X} \left\{ f(x) = \mathbb{E} \left[\phi \left(\sum_{i=1}^n \left(\frac{i}{n} + \xi_i \right) x_i \right) \right] + \frac{\eta}{2} \|x\|^2 \right\}, \quad (17)$$

where $X = \{x \in \mathbb{R}^n | x \geq 0, \sum_{i=1}^n x_i = 1\}$ and ξ_i are independent and identically distributed random variables with mean zero and variance one for $i = 1, \dots, n$. The function $\phi(\cdot)$ is a piecewise linear convex function given by $\phi(t) = \max_{1 \leq i \leq m} \{v_i + s_i t\}$, where v_i and s_i are constants between zero and one, and $f(x, \xi) = \phi(\sum_{i=1}^n (\frac{i}{n} + \xi_i)x_i)$. By applying our smoothing approach, the modified problem is given by

$$\min_{x \in X} \left\{ \hat{f}(x) \triangleq \mathbb{E} \left[\phi \left(\sum_{i=1}^n \left(\frac{i}{n} + \xi_i \right) (x_i + z_i) \right) \right] + \frac{\eta}{2} \|x + z\|^2 \right\}, \quad (18)$$

where $z \in \mathbb{R}^n$ is the uniform distribution on a ball with radius ε with independent elements z_i , $1 \leq i \leq n$. While space limitations preclude a detailed numerical study, we provide schematics of the performance of our iterative smoothing extension of the RSA scheme in Figure 2 for two problem instances in which the smoothing parameter is reduced at $1/k$ and $1/\sqrt{k}$, of which the former displays better properties.

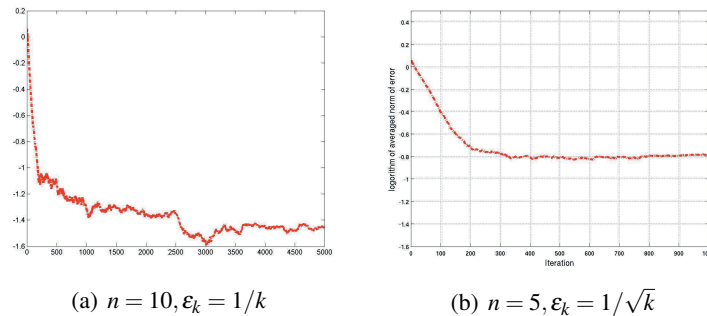


Figure 2: The stochastic utility problem: random smoothing with adaptive stepsizes.

5 CONCLUDING REMARKS

Traditionally, a key challenge in the implementation of SA schemes lies in the choice of the steplength sequences. In earlier work in the context of stochastic convex optimization, we developed a simple recursive procedure that was sensitive to problem parameters but relied on strong convexity. We extend this regularized recursive SA scheme to the regime of merely monotone stochastic variational inequalities by overlaying a regularization parameter that is reduced after every gradient step. Convergence in mean of the associated sequence is presented. The second part of the paper introduces an iterative smoothing generalization that allows for multivalued mappings. Preliminary results on a class of stochastic convex and nonsmooth problems suggest that adaptive schemes perform well in comparison with standard implementations.

REFERENCES

Broadie, M., D. Cicek, and A. Zeevi. 2011. “General Bounds and Finite-Time Improvement for the Kiefer-Wolfowitz Stochastic Approximation Algorithm”. *Operations Research* (To appear).
 Facchinei, F., and J.-S. Pang. 2003. *Finite-dimensional variational inequalities and complementarity problems. Vols. I,II*. Springer Series in Operations Research. New York: Springer-Verlag.

- Jiang, H., and H. Xu. 2008. “Stochastic approximation approaches to the stochastic variational inequality problem”. *IEEE Transactions on Automatic Control* 53 (6): 1462–1475.
- Kesten, H. 1958. “Accelerated stochastic approximation”. *Ann. Math. Statist* 29:41–59.
- Koshal, J., A. Nedić, and U. V. Shanbhag. 2010. “Single Timescale Regularized Stochastic Approximation Schemes for Monotone Nash games under Uncertainty”. *Proceedings of the IEEE Conference on Decision and Control (CDC)*:231–236.
- Lakshmanan, H., and D. Farias. 2008. “Decentralized Recourse Allocation In Dynamic Networks of Agents”. *SIAM Journal on Optimization* 19 (2): 911–940.
- Polyak, B. 1987. *Introduction to optimization*. New York: Optimization Software, Inc.
- Polyak, B., and A. Juditsky. 1992. “Acceleration of stochastic approximation by averaging”. *SIAM J. Control Optim.* 30 (4): 838–855.
- Ravat, U., and U. V. Shanbhag. 2011. “On the Characterization of Solution Sets in Smooth and Nonsmooth Convex Stochastic Nash Games”. *SIAM Journal of Optimization* (To appear).
- Robbins, H., and S. Monro. 1951. “A stochastic approximation method”. *Ann. Math. Statistics* 22:400–407.
- Sacks, J. 1958. “Asymptotic distribution of stochastic approximation procedures”. *Ann. Math. Statist.* 29:373–405.
- Venter, J. H. 1967. “An extension of the Robbins-Monro procedure”. *Ann. Math. Statist.* 38:181–190.
- Yousefian, F., A. Nedić, and U. Shanbhag. 2011. “On Stochastic Gradient and Subgradient Methods with Adaptive Steplength Sequences”. *Automatica* (To appear).

AUTHOR BIOGRAPHIES

FARZAD YOUSEFIAN received his B.S. degree in 2006, and M.S. degree in 2008, from Sharif University of Technology, both in Industrial Engineering. He has been in the Ph.D. program at the Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign since 2008. His current research interests lie in the development of algorithms in the regime of variational inequalities and optimization in uncertain and nonsmooth settings.

ANGELIA NEDIĆ received her B.S. degree from the University of Montenegro (1987) and M.S. degree from the University of Belgrade (1990), both in Mathematics. She received her Ph.D. degrees from Moscow State University (1994) in Mathematics and Mathematical Physics, and from Massachusetts Institute of Technology in Electrical Engineering and Computer Science (2002). She has been at the BAE Systems Advanced Information Technology from 2002-2006. Since 2006 she is an Assistant professor at the Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign. Her general interest is in optimization including fundamental theory, models, algorithms, and applications. Her current research interest is focused on large-scale convex optimization, distributed multi-agent optimization, and duality theory with applications in decentralized optimization. She received an NSF Faculty Early Career Development (CAREER) Award in 2008 in Operations Research.

UDAY V. SHANBHAG received his Ph.D. degree in operations research from the Department of Management Science and Engineering, Stanford University, Stanford, CA, in 2006. He also holds a Bachelors degree from the Indian Institute of Technology (IIT), Bombay (1993) and masters degrees from the Massachusetts Institute of Technology (MIT), Cambridge (1998). He is currently an Assistant Professor in industrial and enterprise systems engineering at the University of Illinois at Urbana-Champaign and his interests lies in the development of analytical and algorithmic tools in the context of optimization and variational problems, in regimes complicated by uncertainty, dynamics and nonsmoothness. Dr. Shanbhag received the triennial A.W. Tucker Prize for his dissertation from the mathematical programming society (MPS) in 2006 and the Computational Optimization and Applications (COAP) Best Paper Award in 2007 (with Walter Murray).