Proceedings of the 2011 Winter Simulation Conference
S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, eds.

A regularized adaptive steplength stochastic approximation scheme for monotone stochastic variational inequalities

Farzad Yousefian

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

Angelia Nedić

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

Uday V. Shanbhag

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

ABSTRACT

We consider the solution of monotone stochastic variational inequalities via stochastic approximation (SA) techniques. 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 when considering constrained problems with possibly nonsmooth objectives. Second, while convergence of the scheme requires that the steplength sequences need to satisfy $\sum_k \gamma_k = \infty$ and $\sum_k \gamma_k^2 < \infty$, little guidance is provided for a specific choice. In fact, standard choices such as $\gamma_k = 1/k$ can be seen to perform poorly on certain problems. Motivated by these two shortcomings, we develop an adaptive steplength stochastic approximation framework in which the stepsize is updated recursively; more specifically, at each iteration, the steplength is updated via a recursive rule contingent on the stepsize at the previous iteration and some problem parameters. By overlaying a regularization parameter, such a scheme allows for solving stochastic monotone variational problems. Next, we consider a class of stochastic multivalued variational problems arising from either stochastic optimization problems or stochastic Nash games. In this setting, we present an iterative smoothing generalization of this recursive stochastic approximation rule that is based on introducing a random local perturbations in the player objectives. Such an algorithm necessitates taking projection steps with respect to a perturbed problem and requires that the perturbation or smoothing parameter is reduced after every step. Such a smoothing approach allows for deriving a Lipschitzian property on the associated mapping which may then be leveraged in developing convergence statements. Preliminary numerical results suggest that both the regularized and smoothing generalizations of the recursive scheme perform well and are relatively robust to parametric variations.

1 Introduction

In this paper, we consider the solution of a monotone stochastic variational inequality, denoted by $\mathrm{VI}(X,F)$ where $X \subseteq \mathbb{R}^n$ is a closed and convex set and the mapping $F : X \to \mathbb{R}^n$ is given by $F(x) \triangleq \mathsf{E}[f(x,\xi)]$. Note that ξ is a random variable with $\xi : \Omega \to \mathbb{R}^d$ and the associated probability space is denoted by $(\Omega, \mathscr{F}, \mathbb{P})$. A solution of $\mathrm{VI}(X,F)$ is denoted by x^* where $x^* \in \mathrm{SOL}(X,F)$, the solution set of $\mathrm{VI}(X,F)$, 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 (Facchinei and Pang 2003) are a particularly broad class of objects that allow 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 (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 (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. (Koshal, Nedić, and Shanbhag 2010) where Tikhonov regularization and proximal-point schemes are overlaid on standard SA techniques.

One of the motivations for this 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 in 1950 (Robbins and Monro 1951). In 1957, Kesten (Kesten 1958) suggested an sequence that adapts to the observed data. Subsequently, under suitable conditions, Sacks (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, Ventner (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. More recently, Broadie et al. in (Broadi, Cicek, and Zeevi) consider an adaptive Kiefer-Wolfowitz (KW) stochastic approximation algorithm and derive general upper bounds on its mean-squared error.

In recent work (Yousefian, Nedic, and Shanbhag 2011), we present two adaptive steplength SA schemes for strongly convex programs, both of which produce a sequence of iterates that is 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 suitably error criterion is satisfied and the resulting 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 merely monotone stochastic VIs; (ii) Second, we overlay a regularization scheme that facilitates addressing merely monotone stochastic VIs; and (iii) Third, we iteratively update the smoothing parameter to allow for solving multivalued generalizations. 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 a smoothing-based generalization of the recursive scheme for stochastic Nash games where we assume that the perturbation parameter is a decreasing sequence. Finally, in Section 4, we present some preliminary numerical results.

2 A regularized recursive steplength stochastic approximation scheme

In section 2.1, we introduce a regularized variant of stochastic approximation scheme, akin to that in work by Koshal et al. (Koshal, Nedić, and Shanbhag 2010). In such a regime, a range of steplength and regularization sequences are provided to ensure convergence of the overall scheme in (Koshal, Nedić, and Shanbhag 2010). In Section 2.2, we modify the proposed scheme in Section 2.1 by providing a framework for recursively updating the steplength .

2.1 A regularized stochastic approximation scheme

Consider a regularized stochastic approximation scheme of the form:

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

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 $\mathsf{E}\left[\|x_0\|^2\right] < \infty$. We make the following two assumptions through our analysis.

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 $\sum_{k=0}^{\infty} \gamma_k^2 \mathsf{E}\left[\|w_k\|^2 \mid \mathscr{F}_k\right] < \infty$ 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) $\lim_{k\to\infty} \eta_k = 0$;

(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) $\lim_{k\to\infty} \frac{(\eta_{k-1}-\eta_k)^2}{\eta_k^3\gamma_k}(1+\frac{1}{\gamma_k\eta_k}) = 0$; (f) $\lim_{k\to\infty} \frac{\gamma_k}{\eta_k} \mathsf{E}\left[\|w_k\|^2 \mid \mathscr{F}_k\right] = 0$.

We also make use of the following result, which can be found in (Polyak 1987) (see Lemma 11 in page 50).

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

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

Then, $v_k \to 0$ almost surely.

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

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

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

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

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

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

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 in (Koshal, Nedić, and Shanbhag 2010). We replace assumptions $\lim_{k\to\infty}\frac{\gamma_k}{\eta_k}(\eta_k+L)^2=0$ and $\lim_{k\to\infty}\gamma_k\eta_k=0$ by part (a) of Assumption 2 which is a weaker condition.

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

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

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

Proof. From Lemma 2 we have

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

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

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

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

$$\mathsf{E}\left[\|x_{k+1}-y_k\|^2\mid\mathscr{F}_k\right]\leq(1-\frac{\eta_k}{2}\gamma_k)\|x_k-y_{k-1}\|^2+q_kM_y\frac{(\eta_{k-1}-\eta_k)^2}{\eta_k^2}(1+\frac{1}{\gamma_k\eta_k})+\gamma_k^2\mathsf{E}\left[\|w_k\|^2\mid\mathscr{F}_k\right]. \quad (6)$$

To claim the convergence of the sequence $\|x_{k+1}-y_k\|$, we show that Lemma 1 applies to relation (6). Let us define

$$v_k\triangleq q_kM_y^2\frac{(\eta_{k-1}-\eta_k)^2}{\eta_k^2}(1+\frac{1}{\gamma_k\eta_k})+\gamma_k^2\mathsf{E}\left[\|w_k\|^2\mid\mathscr{F}_k\right].$$

Therefore we have

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

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 $k>K$. Note that from part (a) of Assumption 2 for $k>K$, $u_k<1$ since for any $L>0$

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

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

$$\begin{aligned}\frac{v_k}{u_k} &\leq \frac{2}{\eta_k\gamma_k} M_y^2 \frac{(\eta_{k-1}-\eta_k)^2}{\eta_k^2}(1+\frac{1}{\gamma_k\eta_k}) + \frac{2\gamma_k^2}{\eta_k\gamma_k} \mathsf{E}\left[\|w_k\|^2 \mid \mathscr{F}_k\right] \\ &= M_y^2 \frac{2(\eta_{k-1}-\eta_k)^2}{\eta_k^3\gamma_k}(1+\frac{1}{\gamma_k\eta_k}) + \frac{2\gamma_k}{\eta_k} \mathsf{E}\left[\|w_k\|^2 \mid \mathscr{F}_k\right]. \end{aligned} \tag{7}$$

By Assumption 2 (e) and (f), the above two terms converge to zero, implying that $\lim_{k\to\infty} \frac{v_k}{u_k} = 0$. 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 $\mathrm{VI}(X,F)$ imply that $x_k \to x^*$ almost surely. □

Remark on choices of γ_k and η_k: While Assumption 2 appears rather difficult to satisfy, in (Koshal, Nedić, and Shanbhag 2010), we 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 this section with a corollary of the above result that relies on the following assumption on the errors.

Assumption 3 The errors w_k are such that for some $\nu > 0$ $\mathsf{E}\left[\|w_k\|^2 | \mathscr{F}_k\right] \leq \nu^2 \quad a.s.$ for all $k \geq 0$.

The following corollary gives a parametric upper bound for the error which 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

$$\mathsf{E}\left[\|x_{k+1}-y_k\|^2\right] \leq (1-\frac{\eta_k}{2}\gamma_k)\mathsf{E}\left[\|x_k - y_{k-1}\|^2\right] + q_k M_y^2 \frac{(\eta_{k-1}-\eta_k)^2}{\eta_k^2}(1+\frac{1}{\gamma_k\eta_k}) + \gamma_k^2\nu^2, \forall k \geq 1, \tag{8}$$

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 (8) follows from (4) by taking expectations and invoking Assumption 3. □

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 $\mathsf{E}\left[\|x_{k+1}-y_k\|^2\right]$ as an error, denoted by e_{k+1}, and arising from the use of the stepsize values $\gamma_0, \gamma_1, \ldots, \gamma_k$. Thus, in the worst case, the error satisfies the following recursive relation for any $k \geq 1$:

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

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 ν^2 is the upper bound for the second moments of the error norms $\|w_k\|$. Then, it seems natural to investigate if the stepsizes $\gamma_1, \gamma_1 \ldots, \gamma_k$ 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:

$$M_k \triangleq M_y^2\frac{(\eta_{k-1}-\eta_k)^2}{\eta_k^2}, \qquad a_k \triangleq \frac{\eta_k}{2}, \qquad b_k \triangleq \frac{M_k}{\eta_k}, \qquad c_k \triangleq -M_k,$$
$$d_k \triangleq M_k\left(\frac{(\eta_k+L)^2}{\eta_k}-2\eta_k\right), \qquad f_k \triangleq M_k(\eta_k+L)^2+\nu^2. \tag{9}$$

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

$$e_k(\gamma_1,\ldots,\gamma_{k-1}) = (1-a_{k-1}\gamma_{k-1})e_{k-1}(\gamma_1,\ldots,\gamma_{k-2}) + \frac{b_{k-1}}{\gamma_{k-1}} + c_{k-1} + d_{k-1}\gamma_{k-1} + f_{k-1}\gamma_{k-1}^2. \tag{10}$$

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

$$D_k(t) \triangleq \frac{1}{a_k}\left(-\frac{b_k}{t^2} + d_k + 2f_k t\right). \tag{11}$$

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

$$D_1(\gamma_1) - e_1 = 0, \tag{12}$$
$$D_{k+1}(\gamma_{k+1}) = (1-a_k\gamma_k)D_k(\gamma_k) + \frac{b_k}{\gamma_k^*} + c_k + d_k\gamma_k^* + f_k{\gamma_k}^2 \tag{13}$$

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 $e_k(\gamma_1,\ldots,\gamma_{k-1})$ by e_k whenever this is unambiguous. In our main result of this subsection, we show that the stepsizes γ_i, $i = 1,\ldots,k-1$, minimize the errors e_k over the range that $0 < \gamma_k < \frac{\eta_k}{(\eta_k+L)^2}$ for $k \geq 1$, where L is the Lipschitz constant associated with the mapping $F(x)$.

Proposition 2 Let $e_k(\gamma_1,\ldots,\gamma_{k-1})$ be defined as in (2.2), where $e_1 > 0$ is such that (12) has a real positive root as γ_1^*, and we assume that for each $k \geq 1$, equation (13) has a real positive root as γ_k^*. Let the sequence $\{\gamma_k^*\}$ be given by (12)–(13). Suppose that the sequence $\{\eta_k\}_{k=0}^\infty$ is nonincreasing and fixed. Then, the following hold:

(a) The error e_k satisfies $e_k(\gamma_1^*,\ldots,\gamma_{k-1}^*) = \frac{2}{\eta_k}(-\frac{b_k}{{\gamma_k^*}^2} + d_k + 2f_k\gamma_k^*)$ for all $k \geq 1$, where b_k, d_k, and f_k are defined in (9).

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

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

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

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

Proof. (a) We use induction on k to prove our result. Note that the result holds trivially for $k = 1$ from (12) and definition of function D_1. Next, assume that we have $e_k(\gamma_1^*,\ldots,\gamma_{k-1}^*) = \frac{1}{a_k}(-\frac{b_k}{\gamma_k^*} + d_k + 2f_k\gamma_k^*)$ for some k, and consider the case for $k+1$. By the definition of the error e_k in (2.2), we have

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

where the second equality follows by the inductive hypothesis and definition of D_k in (11), the third inequality follows by (13), and the last inequality follows by definition of D_{k+1} in (11).
(b) We now show that $(\gamma_1^*,\gamma_1^*,\ldots,\gamma_{k-1}^*)$ 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 (12) and definition of D_1 in (11) we have

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

where the last inequality follows by positiveness of b_1 and f_1. Now suppose that $e_k(\gamma_1,\ldots,\gamma_{k-1})\geq e_k(\gamma_1^*,\ldots,\gamma_{k-1}^*)$ holds for some k and any $(\gamma_1,\ldots,\gamma_{k-1})\in\mathbb{G}_k$. We want to show that $e_{k+1}(\gamma_1,\ldots,\gamma_k)\geq e_{k+1}(\gamma_1^*,\ldots,\gamma_k^*)$ holds as well for all $(\gamma_1,\ldots,\gamma_k)\in\mathbb{G}_{k+1}$. To simplify the notation we use e_{k+1}^* to denote the error e_{k+1} evaluated at $(\gamma_1^*,\gamma_1^*,\ldots,\gamma_k^*)$, and e_{k+1} when evaluating at an arbitrary vector $(\gamma_1,\gamma_1,\ldots,\gamma_k)\in\mathbb{G}_{k+1}$. 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 (13) and the definition of D_k in (11),

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

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

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 in (Koshal, Nedić, and Shanbhag 2010).

Proposition 3 Let Assumptions 1(a)-(b), and 3 hold and suppose that $\mathrm{SOL}(X,F)$ 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 $\mathrm{VI}(X,F)$ almost surely.

Proposition 4 Let Assumptions 1(a)-(b), and 3 hold and suppose that $\mathrm{SOL}(X,F)$ 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 (12)-(13). Then, the sequence $\{x_k\}$ generated by iterative Tikhonov scheme (2) converges in expectation to the least-norm solution x^* of $\mathrm{VI}(X,F)$.

Proof. Let $\bar{\gamma}_k = k^{-a}$ 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 $\{\bar{\gamma}_k\}$ and $\{\eta_k\}$. Then from Lemma 3, Assumptions 1(c), and 2 are satisfied by this choice of $\{\eta_k\}$ and $\{\bar{\gamma}_k\}$ and therefore by Proposition 1, $\mathsf{E}\left[\|\bar{x}_{k+1}-y_k\|^2\right]$ goes to zero where $\{y_k\}$ is the sequence of exact solution to $VI(X,F+\eta_k\mathbf{I})$ 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 $k \geq 0$:

$$e_k(\gamma_1,\ldots,\gamma_{k-1}) - e_k(\gamma_1^*,\ldots,\gamma_{k-1}^*) \geq 0. \tag{14}$$

Since $\mathsf{E}\left[\|\bar{x}_{k+1}-y_k\|^2\right] - \mathsf{E}\left[\|x_{k+1}-y_k\|^2\right] \geq 0$ and by noting that $\mathsf{E}\left[\|\bar{x}_{k+1}-y_k\|^2\right] \to 0$ 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 $\mathrm{VI}(X,F)$. □

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

In recent work (Yousefian, Nedic, and Shanbhag 2011), we 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 (Yousefian, Nedic, and Shanbhag 2011, 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 arising from stochastic Nash games over continuous strategy sets. 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. 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 ith player solves the following problem, given $x_{-i} \triangleq (x_j)_{i\neq j\in\{1,\ldots,N\}}$:

$$\mathrm{Ag}(x_{-i}) \qquad \min_{x_i\in X_i} \quad \mathsf{E}[f_i(x_i;x_{-i},\xi)], \tag{15}$$

where for $i=1,\ldots,N$, $X_i \subseteq \mathbb{R}^{n_i}$ is a closed and convex set, $\sum_{i=1}^N n_i = n$ and $f_i : \mathbb{R}^n \times \Omega \to \mathbb{R}$ 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 $\mathrm{Ag}(x_{-i}^*)$ for $i=1,\ldots,N$. Then the resulting Nash equilibrium is given by a solution to a multi-valued variational inequality $\mathrm{VI}(X,\partial F)$ where $X \triangleq \prod_{i=1}^N X_i$ and $\partial F(x) \triangleq \partial_{x_i}\mathsf{E}[f_i(x;\xi)]$ for $i=1,\ldots,N$. Note that if $f_i(x;\xi)$ is differentiable in x_i, given x_{-i}, $\mathrm{VI}(X,\partial F)$ reduces to $\mathrm{VI}(X,F)$ where $F(x)=(\nabla_{x_i}\mathsf{E}[f_i(x;\xi)])_{i=1}^N$ is a single-valued mapping. Clearly, if $N=1$, then the Nash 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 ith player solves the following *smoothed* problem:

$$\min_{x_i\in X_i} \quad \mathsf{E}[\mathsf{E}[f_i(x_i+z_i;x_{-i},\xi)\mid\xi]], \tag{16}$$

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:

$$X_\varepsilon = \{y \mid y = x+z,\ x\in X,\ z\in\mathbb{R}^n,\ \|z\|\le\varepsilon\}.$$

We employ a uniform distribution for purposes of smoothing but a normal distribution may also work, as considered in (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:

$$p_u(z;\varepsilon) = \begin{cases} \frac{1}{c_n\varepsilon^n} & \text{for } \|z\|\le\varepsilon, \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

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

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

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. The resulting scheme takes on the following form:

$$x_{k+1} = \Pi_X \left(x_k - \gamma_k(\hat{F}(x_k) + w_k)\right) \qquad \text{for all } k \geq 0,$$
$$w_k = s_k - \hat{F}(x_k), \quad \text{where } s_k \in \prod_{i=1}^{N} \partial_{x_i} f_i(x_i + z_i^k; x_{-i}, \xi), \tag{18}$$

where z_k is drawn from a uniform distribution $p_u(z;\varepsilon_k)$ and $\varepsilon_k \to 0$. In effect, the step computation is modified by a randomness drawn from a uniform distribution with steadily decreasing support. While this seems intuitive form the standpoint of recovering convergence, the rate at which ε_k is reduced remains crucial from the standpoint of recovering convergence. In (Yousefian, Nedic, and Shanbhag 2011), we 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

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

where $\kappa = \frac{2}{\pi}$ if n is even and $\kappa = 1$ otherwise. Therefore, as $\varepsilon \to 0$, 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 $0 \leq \gamma_j \leq \frac{\eta}{(\eta + L(\varepsilon_j))^2}$ for $j = 1, \ldots, k$. 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 $\sum_k \gamma_k = \infty$ and is a focus of ongoing research.

4 Numerical result

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 over a Linux OS. Furthermore, the true solutions were computed by solving a sample-average approximation (SAA) problem.

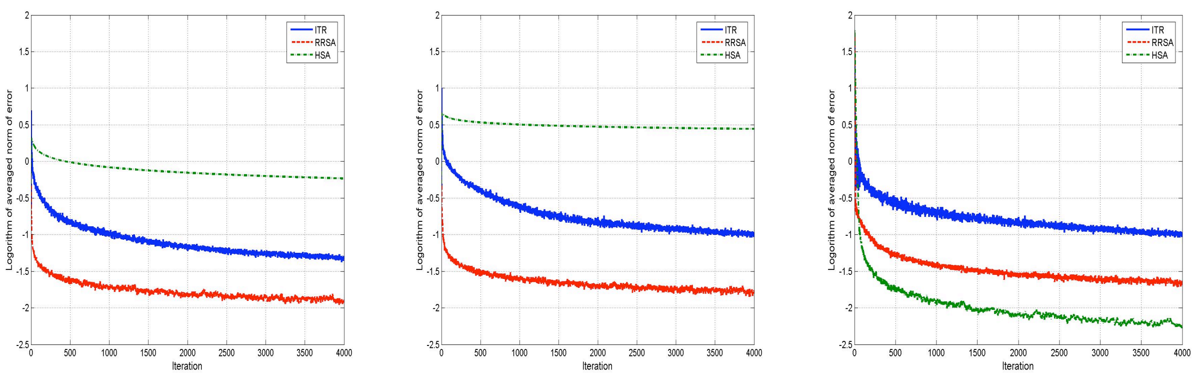
4.1 A stochastic network utility problem

We consider a spatial network with n users competing over L_1 links. Suppose the ith 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 $\mathrm{VI}(X,F)$ 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.



(a) $\beta = 5, C = C_1$ (b) $\beta = 10, C = C_1$ (c) $\beta = 10, C = C_3$

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

-	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]

Table 1: Parametric Sensitivity of Confidence Intervals

4.2 A nonsmooth stochastic optimization problem

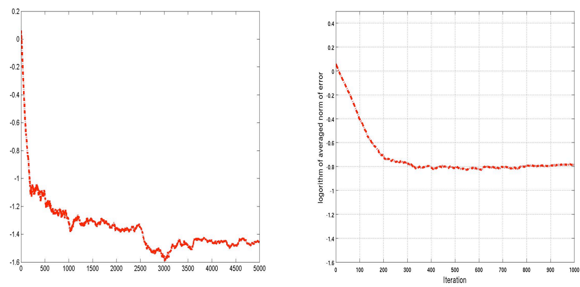
Next, we examine the following nonsmooth optimization problem:

$$\min_{x \in X} \left\{ f(x) = \mathsf{E}\left[\phi\left(\sum_{i=1}^{n}\left(\frac{i}{n} + \xi_i\right)x_i\right)\right] + \frac{\eta}{2}\|x\|^2 \right\}, \tag{19}$$

where $X = \{x \in R^n | x \geq 0, \sum_{i=1}^{n} x_i = 1\}$ and ξ_i are independent and identically distributed random variables with mean zero and variance one for $i = 1, \ldots, n$. 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

$$\min_{x\in X}\left\{\hat{f}(x) \triangleq \mathsf{E}\left[\phi(\sum_{i=1}^{n}(\frac{i}{n}+\xi_i)(x_i+z_i)) + \frac{\eta}{2}\|x+z\|^2\right]\right\}, \tag{20}$$

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 $1/\sqrt{k}$.



(a) $n = 10, \varepsilon_k = 1/k$ (b) $n = 5, \varepsilon_k = 1/\sqrt{k}$

Figure 2: The stochastic utility problem: randomized smoothing technique with adaptive stepsizes

REFERENCES

Broadi, M., D. Cicek, and A. Zeevi. “General Bounds and Finite-Time Improvement for the Kiefer-Wolfowitz Stochastic Approximation Algorithm”. *Submitted to Operations Research manuscript OPRE-2009-02-058.R4*.

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 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), 2010*.

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.

Ravat, U., and U. V. Shanbhag. 2011. “On the Characterization of Solution Sets in Smooth and Nonsmooth Stochastic Convex Nash Games”. *To appear in SIAM Journal of Optimization*.

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. Nedic, and U. Shanbhag. 2011. “On Stochastic Gradient and Subgradient Methods with Adaptive Steplength Sequences”. *To appear in Automatica*.