Adaptive System Optimization Using Random Directions Stochastic Approximation

Prashanth L. A., Shalabh Bhatnagar, *Senior Member, IEEE*, Michael Fu, *Fellow, IEEE*, and Steve Marcus, *Fellow, IEEE*

***Abstract*—We present new algorithms for *simulation optimization* using random directions stochastic approximation (RDSA). These include first-order (gradient) as well as second-order (Newton) schemes. We incorporate both continuous-valued as well as discrete-valued perturbations into both types of algorithms. The former are chosen to be independent and identically distributed (i.i.d.) symmetric uniformly distributed random variables (r.v.), while the latter are i.i.d. asymmetric Bernoulli r.v.s. Our Newton algorithm, with a novel Hessian estimation scheme, requires N-dimensional perturbations and three loss measurements per iteration, whereas the simultaneous perturbation Newton search algorithm of [1] requires $2N$-dimensional perturbations and four loss measurements per iteration. We prove the asymptotic unbiasedness of both gradient and Hessian estimates and asymptotic (strong) convergence for both first-order and second-order schemes. We also provide asymptotic normality results, which in particular establish that the asymmetric Bernoulli variant of Newton RDSA method is better than 2SPSA of [1]. Numerical experiments are used to validate the theoretical results.**

***Index Terms*—Random directions stochastic approximation (RDSA), simultaneous perturbation stochastic approximation (SPSA), stochastic approximation, stochastic optimization.**

I. Introduction

PROBLEMS of optimization under uncertainty arise in many areas of engineering and science, such as signal processing, operations research, computer networks, and manufacturing systems. The problems themselves may involve system identification, model fitting, optimal control, or performance evaluation based on observed data. We consider the following optimization problem in N dimensions:

$$\text{Find } x = \arg\min_{x \in \mathbb{R}^N} f(x). \tag{1}$$

We operate in a *simulation optimization* setting [2], i.e., we are given only noise-corrupted measurements of $f(\cdot)$ and the goal is to devise an iterative scheme that is robust to noise. We assume that, at any time instant n, the conditional expectation of the noise, say ξ_n, given all the randomness up to n is zero (see assumption (A2) in Section II-B for the precise requirement).

A general stochastic approximation algorithm [3] finds the zeros of a given objective function when only noisy estimates are available. Such a scheme can also be applied to gradient search provided one has access to estimates of the objective function gradient. Schemes based on sample gradients include perturbation analysis (PA) [4] and likelihood ratio (LR) [5] methods, which typically require only one simulation run per gradient estimate, but are not universally applicable.

Algorithms based on gradient search perform the following iterative update:

$$x_{n+1} = x_n - a_n \widehat{\nabla} f(x_n) \tag{2}$$

where a_n are *step-sizes* that are set in advance and satisfy standard stochastic approximation conditions (see (A5) in Section II) and $\widehat{\nabla} f(\cdot)$ is an estimate of the gradient $\nabla f(\cdot)$ of the objective function f.

The finite difference Kiefer-Wolfowitz (KW) [6] estimates require $2N$ system simulations for a gradient estimate $\widehat{\nabla} f$. This makes the scheme, also called finite difference stochastic approximation (FDSA), disadvantageous for large parameter dimensions. The random directions stochastic approximation (RDSA) approach [7, pp. 58–60] alleviates this problem by requiring two system simulations regardless of the parameter dimension. It does this by randomly perturbing all the N parameter component directions using independent random vectors that are uniformly distributed over the surface of the N-dimensional unit sphere. It has been observed in [8], see also [9]–[11], that RDSA also works in the case when the component perturbations are independent Gaussian or Cauchy distributed.

RDSA with Gaussian perturbations has also been independently derived in [12] by approximating the gradient of the expected performance by its convolution with a multivariate Gaussian that is then seen (via an integration-by-parts argument) as a convolution of the objective function with a scaled Gaussian. This procedure requires only one simulation (regardless of the parameter dimension) with a perturbed parameter (vector) whose component directions are perturbed using independent

Manuscript received April 22, 2016; revised August 1, 2016; accepted August 8, 2016. Date of publication August 16, 2016; date of current version April 24, 2017. This work was supported by the National Science Foundation under Grants CMMI-1434419, CNS-1446665, and CMMI-1362303, by AFOSR under Grant FA9550-15-10050, and by the Robert Bosch Centre for Cyber-Physical Systems, IISc. Recommended by Associate Editor L. H. Lee.

Prashanth L. A. is with the Institute for Systems Research, University of Maryland, College Park, Maryland Switzerland (e-mail: prashanth@isr.umd.edu).

S. Bhatnagar is with the Department of Computer Science and Automation, Indian Institute of Science, Bangalore (e-mail: shalabh@csa.iisc.ernet.in).

M. Fu is with the Robert H. Smith School of Business & Institute for Systems Research, University of Maryland, College Park, Maryland (e-mail: mfu@isr.umd.edu).

S. Marcus is with the Department of Electrical and Computer Engineering & Institute for Systems Research, University of Maryland, College Park, Maryland(e-mail: marcus@umd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2016.2600643

standard Gaussian random variables. A two-simulation finite difference version that has lower bias than the aforementioned RDSA scheme is studied in [13], [8], [10].

Among all gradient-based random perturbation approaches involving sample measurements of the objective function, the simultaneous perturbation stochastic approximation (SPSA) of [14] has been the most popular and widely studied in applications, largely due to its ease of implementation and observed numerical performance when compared with other approaches. Here each component direction of the parameter is perturbed using independent, zero mean, symmetric Bernoulli distributed random variables. In [8], performance comparisons between RDSA, SPSA, and FDSA have been studied through both analytical and numerical comparisons of the mean square error metric, and it is observed that SPSA outperforms both RDSA with Gaussian perturbations and FDSA.

Within the class of simulation-based search methods, there are also methods that estimate the Hessian in addition to the gradient. Such methods perform the following update:

$$x_{n+1} = x_n - a_n(\overline{H}_n)^{-1}\widehat{\nabla} f(x_n) \tag{3}$$

where $\widehat{\nabla} f(x_n)$ is an estimate of the gradient $\nabla f(\cdot)$ as before, while $\overline{H}_n$ is an $N \times N$-dimensional matrix estimating the true Hessian $\nabla^2 f(x^*)$. Thus, (3) can be seen to be the stochastic version of the well-known Newton method for optimization.

Stochastic Newton methods are often more accurate than simple gradient search schemes, which are sensitive to the choice of the constant a_0 in the canonical step-size, $a_n = a_0/n$. The optimal (asymptotic) convergence rate is obtained only if $a_0 > 1/3\lambda_0$, where λ_0 is the minimum eigenvalue of the Hessian of the objective function (see [15]). However, this dependency is problematic, as λ_0 is unknown in a *simulation optimization* setting. Hessian-based methods get rid of this dependency, while attaining the optimal rate (one can set $a_0 = 1$). An alternative approach to achieve the same effect is to employ Polyak-Ruppert averaging [16], [17], which uses larger step-sizes and averages the iterates. However, iterate averaging is optimal only in an asymptotic sense. Finite-sample analysis (see [18, Theorem 2.4]) shows that the initial error (that depends on the starting point x_0 of the algorithm) is not forgotten subexponentially fast, but at the rate $1/n$ where n is the number of iterations. Thus, the effect of averaging kicks in only after enough iterations have passed and the bulk of the iterates are centered around the optimum.

In [19], the Hessian is estimated using $O(N^2)$ samples of the cost objective at each iterate, while in [20] the Hessian is estimated assuming knowledge of objective function gradients. During the course of the last fifteen years, there has been considerable research activity aimed at developing adaptive Newton-based random search algorithms for stochastic optimization. In [1], the first adaptive Newton algorithm using the simultaneous perturbation method was proposed. The latter algorithm involves the generation of $2N$ independent symmetric Bernoulli distributed random variables at each update epoch. The Hessian estimator in this algorithm requires four parallel simulations with different perturbed parameters at each update epoch. Two of these simulations are also used for gradient estimation. The Hessian estimator is projected to the space of positive definite and symmetric matrices at each iterate for the algorithm to progress along a descent direction. In [21], three other simultaneous perturbation estimators of the Hessian that require three, two, and one simulation(s) have been proposed in the context of long-run average cost objectives. The resulting algorithms incorporate two-timescale stochastic approximation, see Chapter 6 of [22]. Certain three-simulation balanced simultaneous perturbation Hessian estimates have been proposed in [23]. In addition, certain Hessian inversion procedures that require lower computational effort have also been proposed, see Section 7.4 of [11]. In [24], a similar algorithm as in [1] is considered except that for computational simplicity, the geometric mean of the eigenvalues (projected to the positive half line) is used in place of the Hessian inverse in the parameter update step. In [25], certain enhancements to the four-simulation Hessian estimates of [1] using some feedback and weighting mechanisms have been proposed. In [10], Newton-based smoothed functional algorithms based on Gaussian perturbations have been proposed. An overview of random search approaches (both gradient and Newton-based) involving both theory and application of these techniques is available in [11].

A related body of work in the machine learning community is bandit convex optimization. Gradient estimates using the principle of first-order RDSA schemes have been used in [26]. In [27], the authors explore smoothing using Gaussian perturbations for stochastic convex optimization problems and establish optimal convergence rates. In [28], the authors establish optimal (non-asymptotic) convergence rates for a mirror descent scheme with RDSA gradient estimate.

The principal aim of this paper is to develop a second-order method that achieves a convergence rate similar to 2SPSA [1], but at a lower per-iteration cost. In other words, the proposed scheme should achieve the same accuracy as 2SPSA, but with lower number of system simulations. For this purpose, we employ the RDSA approach and propose two schemes that differ in the choice of random perturbations. The first scheme employs i.i.d. uniform $[-\eta, \eta]$ (for some $\eta > 0$) perturbations, while the second scheme employs i.i.d. asymmetric Bernoulli perturbations. The latter takes values -1 and $1+\epsilon$ (for some small $\epsilon > 0$) with probabilities $\left(\frac{1+\epsilon}{2+\epsilon}\right)$ and $\left(\frac{1}{2+\epsilon}\right)$, respectively.

As evident from the update rule (3), the performance of any second-order method is affected by the choices of both the gradient and the Hessian estimation schemes. This motivates our study of first-order RDSA methods with the aforementioned two perturbation schedules prior to analysing the second-order RDSA algorithms. While our asymptotic normality results establish that 1RDSA-Unif/1RDSA-AB are inferior to 1SPSA, they are crucial to show that the corresponding second-order schemes 2RDSA-Unif/AB achieve a higher sample efficiency in comparison to 2SPSA. The sample efficiency of 2RDSA-Unif/AB is due to the fact that, in each iteration, these algorithms use 3 system simulations for estimating the Hessian and uses 2 of these same simulations for estimating the gradient. Using 1SPSA in place of 1RDSA would increase the number of simulations per iteration to 5 as 1SPSA uses a different choice for perturbations.

Table I presents a classification of our algorithms based on their order and the perturbations employed.

We summarize our contributions below.

a) Stochastic Newton Method: We propose an adaptive Newton-based RDSA algorithm. The benefits of using our procedure are two-fold. First, the algorithm requires generating only N perturbation variates at each iteration even for the Newton scheme (N being the parameter dimension), unlike the simultaneous perturbation Newton algorithms of [1], [21], [25], [23], [24], which require generation of $2N$ perturbation

TABLE I
A TAXONOMY OF PROPOSED ALGORITHMS

Order → **Perturbations** ↓	First-order	Second-order
Uniform $[-\eta, \eta]$	1RDSA-Unif	2RDSA-Unif
Asymmetric Bernoulli $\{-1, 1+\epsilon\}$	1RDSA-AB	2RDSA-AB

variates. Second, the number of system simulations required per iteration in our procedure is three, whereas the procedure in [1] requires four.

b) Stochastic Gradient Method: We also propose a gradient RDSA scheme that can be used with either uniform or asymmetric Bernoulli perturbations. As in the case of 1SPSA algorithm of [14], our gradient estimates require two parallel simulations per iteration but with a different choice for perturbation compared with 1SPSA.

c) Convergence and Rates: We prove asymptotic unbiasedness of our gradient and Hessian estimators and prove almost sure convergence of our algorithms to local minima of the objective function. We also present asymptotic normality results that help us analyse our algorithms from an asymptotic mean square error (AMSE) viewpoint. From this analysis, we observe that

(i) On *all* problem instances, 2RDSA-AB results in an AMSE that is better than 2SPSA, for the same number of iterations. It is computationally advantageous to employ our adaptive RDSA scheme, as it requires three simulations per iteration (2SPSA requires four per iteration).

(ii) The question of whether 2SPSA or 2RDSA-Unif is better depends on the values of problem-dependent (equivalently, algorithm-independent) quantities, and we provide a characterization of the ratio of AMSE of 2RDSA-Unif to that of 2SPSA in terms of problem-dependent quantities (see (A) and (B) in Section II-C).

(iii) The gradient algorithm 1RDSA-AB exhibits an AMSE that is nearly comparable to that of SPSA, while 1RDSA variants with Gaussian [7], [8] and uniform perturbations perform worse.

d) Experiments: Numerical results using two objective functions, one quadratic and the other fourth-order, show that

(i) asymmetric Bernoulli variant of 1RDSA performs on par with 1SPSA of [14]; and

(ii) our Newton algorithm 2RDSA-AB provides better accuracy levels than the Newton algorithm 2SPSA in [1] for the same number of iterations, despite 2RDSA requiring only 75% of the cost per-iteration as compared to 2SPSA.

We also demonstrate the practicality of our algorithms in a traffic signal control application, where also our algorithms are found to be superior in performance.

The rest of the paper is organized as follows: In Section II, we describe the first-order RDSA algorithm with the accompanying asymptotic theory. In Section III, we present the second-order RDSA algorithm along with proofs of convergence and asymptotic rate results. We present the results from numerical experiments in Section IV and provide concluding remarks in Section V.

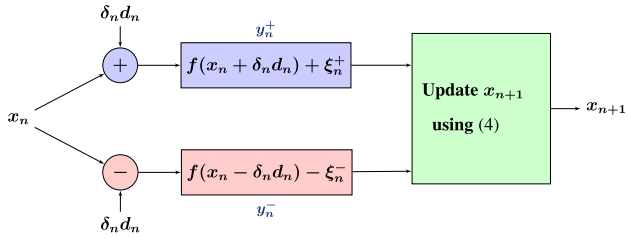

Fig. 1. Overall flow of 1-RDSA algorithm.

II. FIRST-ORDER RANDOM DIRECTIONS SA (1RDSA)

Recall that a first-order gradient search scheme for solving (1) has the following form:

$$x_{n+1} = x_n - a_n \widehat{\nabla} f(x_n), \tag{4}$$

where $\widehat{\nabla} f(x_n)$ is an estimate of $\nabla f(x_n)$. As illustrated in Fig. 1 the idea behind an RDSA scheme is to obtain noisy measurements of f at parameter values $x_n + \delta_n d_n$ and $x_n - \delta_n d_n$. Denote these respective values by y_n^+ and y_n^-, i.e.,

$$y_n^+ = f(x_n + \delta_n d_n) + \xi_n^+, \quad y_n^- = f(x_n - \delta_n d_n) + \xi_n^-.$$

In the above, the noise tuple $\{\xi_n^+, \xi_n^-, n \geq 0\}$ is a martingale difference sequence, the sequence of the perturbation constants $\{\delta_n, n \geq 0\}$ is a positive and asymptotically vanishing sequence and the random perturbations $d_n = (d_n^1, \ldots, d_n^N)^\mathsf{T}$ are such that $\{d_n^i, i = 1, \ldots, N, n = 1, 2, \ldots\}$ are i.i.d. and independent of the noise sequence. These quantities are assumed to satisfy the conditions in (A2)–(A5) in Section III-B below.

In the next section, we specify two different choices for d_n for obtaining the gradient estimate using the noisy function measurements y_n^+ and y_n^-. The first choice uses (continuous-valued) uniform random variables, while the second is based on (discrete-valued) asymmetric Bernoulli random variates.

A. Gradient Estimate

Uniform perturbations: Choose d_n^i, $i = 1, \ldots, N$ to be i.i.d. $U[-\eta, \eta]$ for some $\eta > 0$, where $U[-\eta, \eta]$ denotes the uniform distribution on the interval $[-\eta, \eta]$. The RDSA estimate of the gradient is given by

$$\widehat{\nabla} f(x_n) = \frac{3}{\eta^2} d_n \left[\frac{y_n^+ - y_n^-}{2\delta_n} \right]. \tag{5}$$

Asymmetric Bernoulli perturbations: Choose d_n^i, $i = 1, \ldots, N$, i.i.d. as follows:

$$d_n^i = \begin{cases} -1 & \text{w.p.} \dfrac{(1+\epsilon)}{(2+\epsilon)}, \\ 1+\epsilon & \text{w.p.} \dfrac{1}{(2+\epsilon)} \end{cases} \tag{6}$$

where $\epsilon > 0$ is a constant that can be chosen to be arbitrarily small. Note that, for any $i = 1, \ldots, N$, $Ed_n^i = 0$, $E(d_n^i)^2 = 1 + \epsilon$ and $E(d_n^i)^4 = \frac{(1+\epsilon)(1+(1+\epsilon)^3)}{(2+\epsilon)}$. Henceforth, we will use τ to denote $E(d_n^i)^4$.

Then, the RDSA estimate of the gradient is given by

$$\widehat{\nabla} f(x_n) = \frac{1}{1+\epsilon} d_n \left[\frac{y_n^+ - y_n^-}{2\delta_n} \right]. \tag{7}$$

For notational simplicity, we use $\widehat{\nabla} f(x_n)$ to denote the gradient estimate for both uniform and asymmetric Bernoulli distributions, where the underlying perturbations should be clear from the context.

Motivation for the gradient estimates: Lemma 1 below establishes that the gradient estimates in (5) and (7) are biased by a term of order $O(\delta_n^2)$, and this bias vanishes since $\delta_n \to 0$ (see (A5) below). The proof uses suitable Taylor's series expansions (as in [14]) to obtain the following for both uniform and asymmetric Bernoulli perturbations: $f(x_n \pm \delta_n d_n) = f(x_n) \pm \delta_n d_n^{\mathsf{T}} \nabla f(x_n) + \frac{\delta_n^2}{2} d_n^{\mathsf{T}} \nabla^2 f(x_n) d_n + O(\delta_n^3)$. Hence, as is shown in the proof of Lemma 1

$$\mathbb{E}\left[d_n \left(\frac{f(x_n + \delta_n d_n) - f(x_n - \delta_n d_n)}{2\delta_n} \right) \middle| \mathcal{F}_n \right] = \mathbb{E}\left[d_n d_n^{\mathsf{T}} \right] \nabla f(x_n) + O(\delta_n^2)$$

where $\mathcal{F}_n = \sigma(x_m, m \leq n)$ denotes the underlying sigma-field. For the case of uniform perturbations, it is easy to see that $\mathbb{E}\left[d_n d_n^{\mathsf{T}} \right] = \frac{\eta^2}{3}$ and since we have a scaling factor of $\frac{3}{\eta^2}$ in (5), the correctness of gradient estimate follows as $\delta_n \to 0$. A similar argument holds for the case of asymmetric Bernoulli perturbations.

Remark 1. (Why Uniform/Asymmetric Bernoulli Perturbations?): Previous studies, see [7, Section 2.3.5] and [8]), assumed the perturbation vector d_n to be uniformly distributed over the surface of the unit sphere, whereas we consider alternative uniform/asymmetric Bernoulli perturbations for the following reasons:

Sample efficiency: Let $\hat{n}_{\mathrm{RDSA-Unif}}, \hat{n}_{\mathrm{RDSA-AB}}$ and $\hat{n}_{\mathrm{RDSA-Gaussian}}$ denote the number of function measurements required to achieve a given accuracy using uniform, asymmetric Bernoulli and Gaussian distributed perturbations in 1RDSA, respectively. Further, let $\hat{n}_{\mathrm{SPSA}}$ denote a similar number for the regular SPSA scheme with symmetric Bernoulli perturbations. Then, as discussed in detail in Section II-C, we have the following ratio:

$$\hat{n}_{\mathrm{RDSA-Unif}} : \hat{n}_{\mathrm{RDSA-AB}} : \hat{n}_{\mathrm{RDSA-Gaussian}} : \hat{n}_{\mathrm{SPSA}} = 1.8 : (1+\epsilon) : 3 : 1.$$

Notice that RDSA with Gaussian perturbations requires many more measurements (three times) than regular SPSA. Uniform perturbations bring down this ratio, but they are still significantly suboptimal in comparison to SPSA. On the other hand, asymmetric Bernoulli perturbations exhibit the best ratio that can be made arbitrarily close to 1, by choosing the distribution parameter ϵ to be a very small positive constant.

Computation: Generating perturbations uniformly distributed over the surface of the unit sphere involves simulating N Gaussian random variables, followed by normalization [29]. In comparison, uniform/asymmetric Bernoulli perturbations are easier to generate and do not involve normalization.

Remark 2. (Convex Optimization): In [28], an RDSA-based gradient estimate has been successfully employed to obtain the optimal $O(n^{-1/2})$ rate for stochastic convex optimization under the controlled noise setting, i.e., $\xi_n^+ = \xi_n^-, \forall n$. As in [8], the authors in [28] impose the condition that $\mathbb{E}\left[d_n d_n^{\mathsf{T}} \right] = I$, where I is the identity matrix and suggest Gaussian random variables as one possibility. It is easy to see that uniform and asymmetric Bernoulli perturbations work well in the stochastic convex optimization setting, as well. Recent results in [30] show that the convergence rate of a stochastic gradient descent scheme based on simultaneous perturbations (including RDSA) is $\Omega(n^{-1/3})$ in the uncontrolled noise (i.e., ξ_n^+ is not necessarily equal to $\xi_n^-, \forall n$) regime.

B. Main Results

Recall that $\mathcal{F}_n = \sigma(x_m, m < n)$ denotes the underlying sigma-field. We make the following assumptions[1]:

(A1) $f : \mathbb{R}^N \to \mathbb{R}$ is three-times continuously differentiable[2] with $\left| \nabla^3_{i_1 i_2 i_3} f(x) \right| < \alpha_0 < \infty$, for $i_1, i_2, i_3 = 1, \ldots, N$ and for all $x \in \mathbb{R}^N$.

(A2) $\{\xi_n^+, \xi_n^-, n = 1, 2, \ldots\}$ satisfy $\mathbb{E}\left[\xi_n^+ - \xi_n^- \,|\, d_n, \mathcal{F}_n \right] = 0$.

(A3) For some $\alpha_1, \alpha_2 > 0$ and for all $n, \mathbb{E}\left|\xi_n^{\pm}\right|^2 \leq \alpha_1, \mathbb{E}\left| f(x_n \pm \delta_n d_n) \right|^2 \leq \alpha_2$.

(A4) $\{d_n^i, i = 1, \ldots, N, n = 1, 2, \ldots\}$ are i.i.d. and independent of $\mathcal{F}_n$.

(A5) The step-sizes a_n and perturbation constants δ_n are positive, for all n and satisfy

$$a_n, \delta_n \to 0 \text{ as } n \to \infty,$$

$$\sum_n a_n = \infty \text{ and } \sum_n \left(\frac{a_n}{\delta_n} \right)^2 < \infty.$$

(A6) $\sup_n \|x_n\| < \infty$ w.p. 1.

The above assumptions are standard in the analysis of simultaneous perturbation methods, cf. [11]. In particular:

1) (A1) is required to ensure the underlying ODE is well-posed and also for establishing the asymptotic unbiasedness of the RDSA-based gradient estimates. A similar assumption is required for regular SPSA, as well (see Lemma 1 in [14]).
2) (A2) requires that the noise tuple $\{\xi_n^+, \xi_n^-\}$ is a martingale difference for all n, while the second moment bounds in (A3) are necessary to ensure that the effect of noise can be ignored in the (asymptotic) analysis of the 1RDSA recursion (4).
3) (A4) is crucial in establishing that the gradient estimates in (5) and (7) are asymptotically unbiased with the bias vanishing at the rate $O(\delta_n^2)$, because one obtains terms of the form $\mathbb{E}(d_n \xi_n^{\pm} \mid \mathcal{F}_n)$ after separating the function value $f(x_n \pm \delta_n d_n)$ and the noise $\xi_n^{\pm}$ in (5)/(7). The independence requirement in (A4) ensures that $\mathbb{E}(d_n(\xi_n^+ - \xi_n^-) \mid \mathcal{F}_n) = \mathbb{E}(d_n \mathbb{E}((\xi_n^+ - \xi_n^-) \mid d_n, \mathcal{F}_n)) = 0$. See Lemma 1 for the proof details that utilize (A4).
4) The step-size conditions in (A5) are standard stochastic approximation requirements, while the condition that $\sum_n \left(\frac{a_n}{\delta_n} \right)^2 < \infty$ is necessary to bound a certain martingale difference term that arises in the analysis of (4). See the proof of Theorem 2.
5) (A6) is a stability assumption required to ensure that (4) converges and is common to the analysis of stochastic approximation algorithms, which include simultaneous perturbation

[1] All norms are taken to be the Euclidean norm.

[2] Here, $\nabla^3 f(x) = \frac{\partial^3 f(x)}{\partial x^{\mathsf{T}} \partial x^{\mathsf{T}} \partial x^{\mathsf{T}}}$ denotes the third derivate of f at x and $\nabla^3_{i_1 i_2 i_3} f(x)$ denotes the $(i_1 i_2 i_3)$th entry of $\nabla^3 f(x)$, for $i_1, i_2, i_3 = 1, \ldots, N$,

schemes (cf. [14], [1], [11]). Note that (A6) is not straightforward to show in many scenarios. However, a standard trick to ensure boundedness is to project the iterate x_n onto a compact and convex set—see the discussion in [7, pp. 40–41] and also [11, Remark E.1].

We next present three results that hold for uniform, as well as for asymmetric Bernoulli perturbations: First, Lemma 1 establishes that the bias in the gradient estimates (5) and (7) is of the order $O(\delta_n^2)$. Second, Theorem 2 proves that the iterates x_n governed by (4) converge a.s. and finally, Theorem 3 provides a central limit theorem-type result.

Lemma 1. (Bias in the Gradient Estimate): Under (A1)–(A6), for $\widehat{\nabla} f(x_n)$ defined according to either (5) or (7), we have a.s. that[3]

$$\left|\mathbb{E}\left[\widehat{\nabla}_i f(x_n)\middle|\mathcal{F}_n\right] - \nabla_i f(x_n)\right| = O(\delta_n^2), \quad \text{for } i = 1, \ldots, N. \tag{8}$$

Proof: See Appendix I-A. ■

Theorem 2. (Strong Convergence): Let x^* be an asymptotically stable equilibrium of the following ordinary differential equation (ODE): $\dot{x}_t = -\nabla f(x_t)$, with domain of attraction $D(x^*)$, i.e., $D(x^*) = \{x_0 \mid \lim_{t\to\infty} x(t \mid x_0) = x^*\}$, where $x(t \mid x_0)$ is the solution to the ODE with initial condition x_0. Assume (A1)–(A6) and also that there exists a compact subset $\mathcal{D}$ of $D(x^*)$ such that $x_n \in \mathcal{D}$ infinitely often. Here x_n is governed by (4) with the gradient estimate $\widehat{\nabla} f(x_n)$ defined according to either (5) or (7). Then, we have

$$x_n \to x^* \text{ a.s. as } n \to \infty.$$

Proof: See Appendix I-B ■

We now present an asymptotic normality result for 1RDSA, for which we require the following variant of (A3):

(**A3'**) The conditions of (A3) hold. In addition, $\mathbb{E}(\xi_n^+ - \xi_n^-)^2 \to \sigma^2$ a.s. as $n \to \infty$.

The main result is as follows:

Theorem 3. (Asymptotic Normality): Assume (A1), (A2), (A3'), (A4)–(A6). Let $a_n = a_0/n^\alpha$ and $\delta_n = \delta_0/n^\gamma$, where $a_0, \delta_0 > 0, \alpha \in (0, 1]$ and $\gamma \geq 1/6$. Let $\beta = \alpha - 2\gamma > 0$ and P be an orthogonal matrix with $P\nabla^2 f(x)P^\mathsf{T} = \frac{1}{a_0}\text{diag}\,(\lambda_1, \ldots, \lambda_N)$. Then,

$$n^{\beta/2}(x_n - x^*) \xrightarrow{dist} \mathcal{N}(\mu, PMP^\mathsf{T}) \text{ as } n \to \infty, \tag{9}$$

where $\mathcal{N}(\mu, PMP^\mathsf{T})$ denotes the multivariate Gaussian distribution with mean μ and covariance matrix PMP^T. The mean μ is defined as follows: $\mu = 0$ if $\gamma > \alpha/6$ and $\mu = k_\mu(a_0\delta_0^2(2a_0\nabla^2 f(x^*) - \beta^+ I)^{-1}T)$ if $\gamma = \alpha/6$, where

$$k_\mu = \begin{cases} 3.6 & \text{for U}[-1,1] \text{ perturbations,} \\ \dfrac{2\tau}{(1+\epsilon)^2} & \text{for asymmetric Bernoulli} \\ & \text{perturbations.} \end{cases}$$

In the above, I is the identity matrix of size $N \times N$, $\beta^+ = \beta$ if $\alpha = 1$ and 0 if $\alpha < 1$ and $T = (T^1, \ldots, T^N)^\mathsf{T}$ with

$$T^l = -\frac{1}{6}\left[\nabla^3_{lll} f(x^*) + 3\sum_{i=1, i\neq l}^{N} \nabla^3_{iil} f(x^*)\right], l = 1, \ldots, N.$$

[3]Here, $\widehat{\nabla}_i f(x_n)$ and $\nabla_i f(x_n)$ denote the ith coordinates in the gradient estimate $\widehat{\nabla} f(x_n)$ and true gradient $\nabla f(x_n)$, respectively

The covariance matrix M is defined as follows:

$$M = \frac{a_0^2\sigma^2}{4\delta_0^2}\text{diag}((2\lambda_1 - \beta^+)^{-1}, \ldots, (2\lambda_N - \beta^+)^{-1}).$$

Proof: See Appendix I-C ■

C. (Asymptotic) Convergence Rates

The result in Theorem 3 shows that $n^{\beta/2}(x_n - x^*)$ is asymptotically Gaussian for 1RDSA under both perturbation choices. The asymptotic mean square error of $n^{\beta/2}(x_n - x^*)$, denoted by $\mathcal{AMSE}_{1\mathcal{RDSA}}(a, c)$, is given by

$$\mathcal{AMSE}_{1\mathcal{RDSA}}(a_0, \delta_0) = \mu^\mathsf{T}\mu + \text{trace}(PMP^\mathsf{T})$$

where a_0 is the step-size constant, δ_0 is the constant in the perturbation sequence δ_k and μ, P and M are as defined in Theorem 3. Under certain assumptions (cf. [31]), it can be shown that $\mathcal{AMSE}_{1\mathcal{RDSA}}(a, c)$ coincides with $n^\beta \mathbb{E}\left\|x_n - x^*\right\|^2$. From Theorem 3, it is easy to deduce from the conditions on step-size exponent α and perturbation constant exponent γ that the range of β is 0 to 2/3. Following the discussion in Section III-A of [8], a common value of $\beta = 2/3$ is optimal for all first-order algorithms, with $\alpha = 1$ and $\gamma = 1/6$.

With step-size $a_n = a_0/n$, setting a_0 optimally requires knowledge of the minimum eigenvalue λ_0 of the Hessian $\nabla^2 f(x^*)$, i.e., $a_0 > \beta/2\lambda_0$. Under this choice, with T as defined in Theorem 3, we obtain:

$$\begin{aligned} &\mathcal{AMSE}_{1\mathcal{RDSA}-Unif}(a_0, \delta_0) \\ &= \left(3.6\delta_0^2 a_0 \left\|\Psi T\right\|\right)^2 + \delta_0^{-2}\text{trace}\left(\Psi S\right), \end{aligned} \tag{10}$$

$$\begin{aligned} &\mathcal{AMSE}_{1\mathcal{RDSA}-AB}(a_0, \delta_0) \\ &= \left(\frac{2\delta_0^2 a_0 \tau}{(1+\epsilon)^2}\left\|\Psi T\right\|\right)^2 + \delta_0^{-2}\text{trace}\left(\Psi S\right), \end{aligned} \tag{11}$$

$$\text{where } \Psi = (2a_0\nabla^2 f(x^*) - \beta)^{-1} \text{ and } S = \frac{\sigma^2}{4}I. \tag{12}$$

Remark 3. (On Step-Size Dependency): Since λ_0 is unknown, obtaining the above rate is problematic and one can get rid of the dependency of a_0 on λ_0 either by averaging of iterates or employing an adaptive (second-order) scheme. The former would employ step-size $a_n = a_0/n^\alpha$, with $\alpha \in (1/2, 1)$ and couple this choice with averaging of iterates as $\bar{x}_n = 1/n\sum_{m=1}^n x_m$. The latter adaptive scheme would correspond to 2RDSA, which performs a Newton step to update x_n in (13). Section III presents 2RDSA along with an AMSE analysis that compares to 2SPSA.

Comparing AMSE of 1RDSA-Unif to that of 1SPSA: Taking the ratio of AMSE of 1RDSA-Unif to that of 1SPSA with symmetric Bernoulli ± 1-valued perturbations, we obtain the following:

$$\begin{aligned} &\frac{\mathcal{AMSE}_{1\mathcal{RDSA}-Unif}(a_0, \delta_0)}{\mathcal{AMSE}_{1\mathcal{SPSA}}(a_0, \delta_0)} \\ &= \frac{\left(2\delta_0^2 a_0 \left\|\Psi T\right\| 1.8\right)^2 + a_0\delta_0^{-2}\text{trace}\left(\Psi S\right)}{\left(2\delta_0^2 a_0 \left\|\Psi T\right\|\right)^2 + a_0\delta_0^{-2}\text{trace}\left(\Psi S\right)} \\ &= 1 + \frac{2.24}{1 + \left(a_0\delta_0^{-2}\text{trace}\left(\Psi S\right)\right) / \left(2\delta_0^2 a_0 \left\|\Psi T\right\|\right)^2} \end{aligned}$$

where Ψ, S are given by (12) and T is as defined in Theorem 3. From the above, we observe that 1SPSA has a better AMSE in comparison to 1RDSA-Unif, but it is not clear if the difference is 'large'. This is because the ratio in the denominator above depends on the objective function (via $\nabla^2 f(x^*)$ and T) and a high ratio value would make the difference between 1RDSA and 1SPSA negligible. Contrast this with the $1.8:1$ ratio obtained if one knows the underlying objective function (see Remark 4 below).

Comparing AMSE of 1RDSA-AB to that of 1SPSA: For 1RDSA-Unif, $\frac{3}{\eta^2}\mathbb{E}[d_n d_n^{\mathsf{T}}] = I$ and so $\frac{9}{\eta^4}\mathbb{E}[(d_n^i)^4] = 1.8$ appears in the first term in (10) causing 1RDSA to be much inferior to 1SPSA. For 1RDSA-AB, $\frac{1}{(1+\epsilon)}\mathbb{E}[d_n d_n^{\mathsf{T}}] = I$, so corresponding first term multiplier in (11) is $\frac{\tau}{(1+\epsilon)^2}$. For the sake of illustration, choosing $\epsilon = 0.01$ for 1RDSA-AB, we obtain $\frac{\tau}{(1+\epsilon)^2} = 1.000099$. Plugging this value into the AMSE calculation, we obtain

$$\frac{\mathcal{AMSE}_{1\mathcal{RDSA}-AB}(a_0,\delta_0)}{\mathcal{AMSE}_{1\mathcal{SPSA}}(a_0,\delta_0)} = \frac{\left(2\delta_0^2 a_0 \left\|\Psi T\right\| 1.000099\right)^2 + a_0\delta_0^{-2}\mathrm{trace}\left(\Psi S\right)}{\left(2\delta_0^2 a_0 \left\|\Psi T\right\|\right)^2 + a_0\delta_0^{-2}\mathrm{trace}\left(\Psi S\right)}$$

where Ψ, S are given by (12) and T is as defined in Theorem 3. From the above, we observe that 1RDSA-AB has an AMSE that is almost comparable to that of 1SPSA. One could choose a small ϵ to get this ratio arbitrarily close to 1.

Remark 4. (1RDSA With Gaussian Perturbations): In [8], the author simplifies the AMSE for 1RDSA by solving $\mathcal{AMSE}_{1\mathcal{RDSA}}(a_0,\delta_0)$ for δ_0 after setting a_0 optimally using λ_0. Using $N(0,1)$ for d_n and comparing the resulting AMSE of 1RDSA to that of first-order SPSA with symmetric Bernoulli distributed perturbations, they report a ratio of 3:1 for the number of measurements to achieve a given accuracy. Here, 3 is a result of the fact that for $N(0,1)$ distributed $d_n, \mathbb{E}d_n^4 = 3$, while 1 for SPSA comes from a bound on the second and inverse second moments, both of which are 1 for the Bernoulli case. Using $U[-\eta,\eta]$ distributed d_n in 1RDSA would bring down this ratio to $1.8:1$. However, this result comes with a huge caveat-that a_0 and δ_0 are set optimally. Setting these quantities requires knowledge of the objective, specifically, $\nabla^2 f(x^*)$ and the vector T.

III. Second-Order Random Directions SA (2RDSA)

As in the case of the first-order scheme, we present two variants of the second-order simulation optimization method: one using uniform perturbations and the other using asymmetric Bernoulli perturbations. Recall that a second-order adaptive search algorithm has the following form [1]:

$$x_{n+1} = x_n - a_n \Upsilon(\overline{H}_n)^{-1}\widehat{\nabla} f(x_n), \tag{13}$$

$$\overline{H}_n = \frac{n}{n+1}\overline{H}_{n-1} + \frac{1}{n+1}\widehat{H}_n. \tag{14}$$

In the above:

1) $\widehat{\nabla} f(x_n)$ is the estimate of $\nabla f(x_n)$ and this corresponds to (5) for the uniform variant and (7) for the asymmetric Bernoulli variant.
2) $\widehat{H}_n$ is an estimate of the true Hessian $\nabla^2 f(\cdot)$, with $\widehat{H}_0 = I$.
3) $\overline{H}_n$ is a smoothed version of $\widehat{H}_n$, which is crucial to ensure convergence.
4) Υ is an operator that projects a matrix onto the set of positive definite matrices. Update (14) does not necessarily ensure that $\overline{H}_n$ is invertible and without Υ, the parameter update (13) may not move along a descent direction—see conditions (C7) and (C12) in Section III-B below for the precise requirements on the matrix projection operator.

The basic algorithm in (13), (14) is similar to the adaptive scheme analyzed by [1]. However, we use RDSA for the gradient and Hessian estimates, while [1] employs SPSA.

Remark 5. (Matrix Projection): A simple way to define $\Upsilon(\overline{H}_n)$ is to first perform an eigen-decomposition of $\overline{H}_n$, followed by projecting all the eigenvalues onto the positive real line by adding a positive scalar δ_n-see [32], [1] for a similar operator. This choice for Υ satisfies the assumptions (C7) and (C12), which are required to ensure asymptotic unbiasedness of the Hessian scheme presented in the next section. Note that the scalar δ_n used for Υ is also used as a perturbation constant for function evaluations.

A. Hessian Estimate

We use three measurements per iteration in (13) to estimate both the gradient and the Hessian of the objective f. These measurements correspond to parameter values $x_n, x_n + \delta_n d_n$ and $x_n - \delta_n d_n$. Let us denote these values by y_n, y_n^+ and y_n^- respectively, i.e., $y_n = f(x_n) + \xi_n, y_n^+ = f(x_n + \delta_n d_n) + \xi_n^+$ and $y_n^- = f(x_n - \delta_n d_n) + \xi_n^-$, where the noise terms ξ_n, ξ_n^+, ξ_n^- satisfy $\mathbb{E}\left[\xi_n^+ + \xi_n^- - 2\xi_n \,|\, \mathcal{F}_n\right] = 0$. We next present two constructions for the perturbations d_n—one based on i.i.d. uniform r.v.s and the other using asymmetric Bernoulli r.v.s. Unlike the construction in [1], which entails generating $2N$ Bernoulli r.v.s in each iteration, our construction requires N r.v.s that follow either a uniform or an asymmetric Bernoulli distribution.

Uniform perturbations: Using the three measurements and the random directions obtained from d_n, we form the Hessian estimate $\widehat{H}_n$ as follows:

$$\widehat{H}_n = \frac{9}{2\eta^4} M_n \left(\frac{y_n^+ + y_n^- - 2y_n}{\delta_n^2}\right), \text{where} \tag{15}$$

$$M_n = \begin{bmatrix} \frac{5}{2}\left((d_n^1)^2 - \frac{\eta^2}{3}\right) & \cdots & d_n^1 d_n^N \\ d_n^2 d_n^1 & \cdots & d_n^2 d_n^N \\ d_n^N d_n^1 & \cdots & \frac{5}{2}\left((d_n^N)^2 - \frac{\eta^2}{3}\right) \end{bmatrix}.$$

Lemma 4 establishes that the above estimator is of order $O(\delta_n^2)$ away from the true Hessian. The first step of the proof is to use a suitable Taylor's series expansion of f to obtain

$$\frac{f(x_n + \delta_n d_n) + f(x_n - \delta_n d_n) - 2f(x_n)}{\delta_n^2} = \sum_{i=1}^{N}(d_n^i)^2 \nabla_{ii}^2 f(x_n) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_n^i d_n^j \nabla_{ij}^2 f(x_n) + O(\delta_n^2).$$

Taking conditional expectations on both sides above, it can be seen that the RHS does not simplify to the true Hessian and includes bias terms. By multiplying the term $\frac{9}{2\eta^4}M_n$ with the RHS above, we obtain an asymptotically unbiased Hessian

estimate, with the bias vanishing at the rate $O(\delta_n^2)$—see the passage starting from (18) in the proof of Lemma 4 below for details.

Asymmetric Bernoulli perturbations: Using the three measurements and the random directions obtained from d_n, we form the Hessian estimate $\widehat{H}_n$ as follows:

$$\widehat{H}_n = M_n \left(\frac{y_n^+ + y_n^- - 2y_n}{\delta_n^2} \right), \text{where} \tag{16}$$

$$M_n = \begin{bmatrix} \frac{1}{\kappa}\left((d_n^1)^2 - (1+\epsilon)\right) & \cdots & \frac{1}{2(1+\epsilon)^2} d_n^1 d_n^N \\ \frac{1}{2(1+\epsilon)^2} d_n^2 d_n^1 & \cdots & \frac{1}{2(1+\epsilon)^2} d_n^2 d_n^N \\ \frac{1}{2(1+\epsilon)^2} d_n^N d_n^1 & \cdots & \frac{1}{\kappa}\left((d_n^N)^2 - (1+\epsilon)\right) \end{bmatrix}$$

where $\kappa = \tau\left(1 - \frac{(1+\epsilon)^2}{\tau}\right)$. Recall that $\tau = E(d_n^i)^4 = \frac{(1+\epsilon)(1+(1+\epsilon)^3)}{(2+\epsilon)}$, for any $i = 1, \ldots, N$.

Remark 6. (Need for Asymmetry): The Hessian estimate that we construct is such that it disallows using symmetric Bernoulli r.v.s for perturbations. In particular, in establishing the asymptotic unbiasedness of the Hessian estimate, the proof requires that the second and fourth moments of the perturbation r.v.s be different. Asymmetric Bernoulli r.v.s (with only a slight asymmetry) meet this condition and can be used in deriving a gradient estimate as well.

B. Main Results

Recall that $\mathcal{F}_n = \sigma(x_m, m \leq n)$ denotes the underlying sigma-field. We make the following assumptions that are similar to those in [1]:

(C1) The function f is four-times differentiable[4] with $\left|\nabla^4_{i_1 i_2 i_3 i_4} f(x)\right| < \infty$, for $i_1, i_2, i_3, i_4 = 1, \ldots, N$ and for all $x \in \mathbb{R}^N$.

(C2) For each n and all x, there exists a $\rho > 0$ not dependent on n and x, such that $(x - x^*)^{\mathsf{T}} \bar{f}_n(x) \geq \rho \|x_n - x\|$, where $\bar{f}_n(x) = \Upsilon(\overline{H}_n)^{-1} \nabla f(x)$.

(C3) $\{\xi_n, \xi_n^+, \xi_n^-, n = 1, 2, \ldots\}$ satisfy $\mathbb{E}[\xi_n^+ + \xi_n^- - 2\xi_n | \; d_n, \mathcal{F}_n] = 0$, for all n.

(C4) Same as (A4).

(C5) Same as (A5).

(C6) For each $i = 1, \ldots, N$ and any $\rho > 0$, $P(\{\bar{f}_{ni}(x_n) \geq 0 \text{ i.o}\} \cap \{\bar{f}_{ni}(x_n) < 0 \text{ i.o}\} \mid \{|x_{ni} - x_i^*| \geq \rho \quad \forall n\}) = 0$.

(C7) The operator Υ satisfies $\delta_n^2 \Upsilon(H_n)^{-1} \to 0$ a.s. and $E(\left\|\Upsilon(H_n)^{-1}\right\|^{2+\zeta}) \leq \rho$ for some $\zeta, \rho > 0$.

(C8) For any $\varsigma > 0$ and nonempty $S \subseteq \{1, \ldots, N\}$, there exists a $\rho'(\varsigma, S) > \varsigma$ such that

$$\limsup_{n \to \infty} \left| \frac{\sum_{i \notin S} (x - x^*)_i \bar{f}_{ni}(x)}{\sum_{i \in S} (x - x^*)_i \bar{f}_{ni}(x)} \right| < 1 \text{ a.s.}$$

for all $|(x - x^*)_i| < \varsigma$ when $i \notin S$ and $|(x - x^*)_i| \geq \rho'(\varsigma, S)$ when $i \in S$.

(C9) For some $\alpha_0, \alpha_1 > 0$ and for all $n, \mathbb{E}{\xi_n}^2 \leq \alpha_0, \mathbb{E}\xi_n^{\pm 2} \leq \alpha_0, \mathbb{E}f(x_n)^2 \leq \alpha_1$ and $\mathbb{E}f(x_n \pm \delta_n d_n)^2 \leq \alpha_1$.

(C10) $\sum_n \frac{1}{(n+1)^2 \delta_n^4} < \infty$.

For a detailed interpretation of the above conditions, the reader is referred to Section III and Appendix B of [1]. In particular, (C1) holds if the objective f is twice continuously differentiable with a bounded second derivative and (C2) ensures the objective f has enough curvature. (C3)–(C5) are standard requirements on noise and step-sizes and can be motivated in a similar manner as in the case of 1RDSA (see Section II-B). (C6) and (C8) are not necessary if the iterates are bounded, i.e., $\sup_n \|x_n\| < \infty$ a.s. (C7) can be ensured by having Υ defined as mentioned earlier, i.e., $\Upsilon(A)$ performs an eigen-decomposition of A followed by projecting the eigenvalues to the positive side by adding a large enough scalar. Finally, (C9) and (C10) are necessary to ensure convergence of the Hessian recursion, in particular, to invoke a martingale convergence result (see Theorem 6 and its proof below).

Lemma 4. (Bias in Hessian Estimate): Under (C1)–(C10), with $\widehat{H}_n$ defined according to either (15) or (16), we have a.s. that[5], for $i, j = 1, \ldots, N$,

$$\left| \mathbb{E}\left[\widehat{H}_n(i,j) \middle| \mathcal{F}_n \right] - \nabla^2_{ij} f(x_n) \right| = O(\delta_n^2). \tag{17}$$

From the above lemma, it is evident that the bias in the Hessian estimate above is of the same order as 2SPSA of [1].

Proof (Lemma 4) Case 1: Uniform perturbations:
By a Taylor's series expansion, we obtain

$$f(x_n \pm \delta_n d_n) = f(x_n) \pm \delta_n d_n^{\mathsf{T}} \nabla f(x_n) + \frac{\delta_n^2}{2} d_n^{\mathsf{T}} \nabla^2 f(x_n) d_n \pm \frac{\delta_n^3}{6} \nabla^3 f(x_n)(d_n \otimes d_n \otimes d_n) + \frac{\delta_n^4}{24} \nabla^4 f(\tilde{x}_n^+)(d_n \otimes d_n \otimes d_n \otimes d_n).$$

The fourth-order term in each of the expansions above can be shown to be of order $O(\delta_n^4)$ using (C1) and arguments similar to that in Lemma 1 (see (33) there). Hence,

$$\begin{aligned} &\frac{f(x_n + \delta_n d_n) + f(x_n - \delta_n d_n) - 2f(x_n)}{\delta_n^2} \\ &= d_n^{\mathsf{T}} \nabla^2 f(x_n) d_n + O(\delta_n^2) \\ &= \sum_{i=1}^N \sum_{j=1}^N d_n^i d_n^j \nabla^2_{ij} f(x_n) + O(\delta_n^2) \\ &= \sum_{i=1}^N (d_n^i)^2 \nabla^2_{ii} f(x_n) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_n^i d_n^j \nabla^2_{ij} f(x_n) + O(\delta_n^2). \end{aligned}$$

Now, taking the conditional expectation of the Hessian estimate $\widehat{H}_n$ and observing that $\mathbb{E}[\xi_n^+ + \xi_n^- - 2\xi_n \mid \mathcal{F}_n] = 0$ by

[4] Here $\nabla^4 f(x) = \frac{\partial^4 f(x)}{\partial x^{\mathsf{T}} \partial x^{\mathsf{T}} \partial x^{\mathsf{T}} \partial x^{\mathsf{T}}}$ denotes the fourth derivate of f at x and $\nabla^4_{i_1 i_2 i_3 i_4} f(x)$ denotes the $(i_1 i_2 i_3 i_4)$th entry of $\nabla^4 f(x)$, for $i_1, i_2, i_3, i_4 = 1, \ldots, N$.

[5] Here $\widehat{H}_n(i,j)$ and $\nabla^2_{ij} f(\cdot)$ denote the (i,j)th entry in the Hessian estimate $\widehat{H}_n$ and the true Hessian $\nabla^2 f(\cdot)$, respectively.

(C3), we obtain the following:

$$\mathbb{E}[\widehat{H}_n \mid \mathcal{F}_n] = \mathbb{E}\left[M_n \left(\sum_{i=1}^{N-1} (d_n^i)^2 \nabla_{ii}^2 f(x_n) + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} d_n^i d_n^j \nabla_{ij}^2 f(x_n) + O(\delta_n^2) \right) \middle| \mathcal{F}_n \right]. \tag{18}$$

Note that the $O(\delta_n^2)$ term inside the conditional expectation above remains $O(\delta_n^2)$ even after the multiplication with M_n. We analyse the diagonal and off-diagonal terms in the multiplication of the matrix M_n with the scalar above, ignoring the $O(\delta_n^2)$ term.

Diagonal terms in (18):

Consider the lth diagonal term inside the conditional expectation in (18):

$$\begin{aligned} &\frac{45}{4\eta^4}\left((d_n^l)^2 - \frac{\eta^2}{3}\right)\left(\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_n^i d_n^j \nabla_{ij}^2 f(x_n)\right) \\ &= \frac{45}{4\eta^4}(d_n^l)^2 \sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n) + \frac{45}{2\eta^4}(d_n^l)^2 \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_n^i d_n^j \nabla_{ij}^2 f(x_n) \\ &\quad - \frac{15}{4\eta^2}\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n) - \frac{15}{2\eta^2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_n^i d_n^j \nabla_{ij}^2 f(x_n). \end{aligned} \tag{19}$$

From the distributions of d_n^i, d_n^j and the fact that d_n^i is independent of d_n^j for $i < j$, it is easy to see that $\mathbb{E}\left((d_n^l)^2 \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_n^i d_n^j \nabla_{ij}^2 f(x_n) \middle| \mathcal{F}_n \right) = 0$ and $\mathbb{E}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_n^i d_n^j \nabla_{ij}^2 f(x_n) \middle| \mathcal{F}_n \right) = 0$. Thus, the conditional expectations of the second and fourth terms on the RHS of (19) are both zero.

The first term on the RHS of (19) with the conditional expectation can be simplified as follows:

$$\begin{aligned} &\frac{45}{4\eta^4}\mathbb{E}\left((d_n^l)^2 \sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n) \middle| \mathcal{F}_n \right) \\ &= \frac{45}{4\eta^4}\mathbb{E}\left((d_n^l)^4\nabla_{ll}^2 f(x_n) + \sum_{i=1, i\neq l}^{N}(d_n^l)^2(d_n^i)^2\nabla_{ii}^2 f(x_n) \right) \\ &= \frac{45}{4\eta^4}\left(\frac{\eta^4}{5}\nabla_{ll}^2 f(x_n) + \frac{\eta^4}{9}\sum_{i=1, i\neq l}^{N}\nabla_{ii}^2 f(x_n) \right), \text{a.s.} \end{aligned}$$

For the second equality above, we have used the fact that $\mathbb{E}[(d_n^l)^4] = \frac{\eta^4}{5}$ and $\mathbb{E}[(d_n^l)^2(d_n^i)^2] = \mathbb{E}[(d_n^l)^2]\mathbb{E}[(d_n^i)^2] = \frac{\eta^4}{9}, \forall l \neq i$.

The third term in (19) with the conditional expectation and without the negative sign can be simplified as follows:

$$\begin{aligned} &\frac{15}{4\eta^2}\mathbb{E}\left(\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n) \middle| \mathcal{F}_n \right) \\ &= \frac{15}{4\eta^2}\sum_{i=1}^{N}\mathbb{E}\left[(d_n^i)^2\right]\nabla_{ii}^2 f(x_n) = \frac{5}{4}\sum_{i=1}^{N}\nabla_{ii}^2 f(x_n), \text{a.s.} \end{aligned}$$

Combining the above followed by some algebra, we obtain

$$\frac{45}{4\eta^4}\mathbb{E}\left[\left((d_n^l)^2 - \frac{\eta^2}{3}\right)\left(\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_n^i d_n^j \nabla_{ij}^2 f(x_n)\right) \middle| \mathcal{F}_n \right] = \nabla_{ll}^2 f(x_n), \text{a.s.}$$

Off-diagonal terms in (18):

We now consider the (k, l)th term in (18): Assume w.l.o.g that $k < l$. Then

$$\begin{aligned} &\frac{9}{2\eta^4}\mathbb{E}\left[d_n^k d_n^l \left(\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_n^i d_n^j \nabla_{ij}^2 f(x_n) \right) \middle| \mathcal{F}_n \right] \\ &= \frac{9}{2\eta^4}\sum_{i=1}^{N}\mathbb{E}\left(d_n^k d_n^l (d_n^i)^2 \right)\nabla_{ii}^2 f(x_n) + \frac{9}{\eta^4}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbb{E}\left(d_n^k d_n^l d_n^i d_n^j \right)\nabla_{ij}^2 f(x_n) \\ &= \nabla_{kl}^2 f(x_n). \end{aligned} \tag{20}$$

The last equality follows from the fact that the first term in (20) is 0 since $k \neq l$, while the second term in (20) can be seen to be equal to $\frac{9}{\eta^4}\mathbb{E}\left((d_n^k)^2 (d_n^l)^2 \right)\nabla_{kl}^2 f(x_n) = \nabla_{kl}^2 f(x_n)$. The claim follows for the case of uniform perturbations.

Case 2: Asymmetric Bernoulli perturbations:

Note that the proof up to (18) is independent of the choice of perturbations. The proof differs in the analysis of the diagonal and off-diagonal terms in (18). In the case of asymmetric Bernoulli perturbations, the normalizing scalars in the definition of M_n in (16) are different.

Diagonal terms in (18)

Recall that τ denotes the fourth moment $E(d_n^i)^4$, for any $i = 1, \ldots, N$. An analog of (19) is as follows:

$$\begin{aligned}
&\frac{1}{\tau(1-\frac{(1+\epsilon)^2}{\tau})}\mathbb{E}\left(\left((d_n^l)^2-(1+\epsilon)\right)\left(\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n)\right.\right.\\
&\qquad\left.\left.+2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}d_n^i d_n^j\nabla_{ij}^2 f(x_n)\right)\Bigg|\mathcal{F}_n\right)\\
&=\frac{1}{\tau(1-\frac{(1+\epsilon)^2}{\tau})}\mathbb{E}\left((d_n^l)^2\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n)\Big|\mathcal{F}_n\right)\\
&\quad-\frac{(1+\epsilon)}{\tau(1-\frac{(1+\epsilon)^2}{\tau})}\mathbb{E}\left(\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n)\Big|\mathcal{F}_n\right). \qquad (21)
\end{aligned}$$

We have used the fact that the second term in the LHS above is conditionally zero (see argument below (19) for a justification). The first term on the RHS of (21) be simplified as follows:

$$\begin{aligned}
&\frac{1}{\tau(1-\frac{(1+\epsilon)^2}{\tau})}\mathbb{E}\left((d_n^l)^2\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n)\Big|\mathcal{F}_n\right)\\
&=\frac{1}{\tau(1-\frac{(1+\epsilon)^2}{\tau})}\mathbb{E}\Bigg((d_n^l)^4\nabla_{ll}^2 f(x_n)\\
&\qquad+\sum_{i=1,i\neq l}^{N}(d_n^l)^2(d_n^i)^2\nabla_{ii}^2 f(x_n)\Bigg)\\
&=\frac{1}{(1-\frac{(1+\epsilon)^2}{\tau})}\left(\nabla_{ll}^2 f(x_n)+\frac{(1+\epsilon)^2}{\tau}\sum_{i=1,i\neq l}^{N}\nabla_{ii}^2 f(x_n)\right). \qquad (22)
\end{aligned}$$

For the second equality above, we have used the fact that $\mathbb{E}[(d_n^l)^4]=\tau$ and $\mathbb{E}[(d_n^l)^2(d_n^i)^2]=\mathbb{E}[(d_n^l)^2]\mathbb{E}[(d_n^i)^2]=(1+\epsilon)^2, \forall l\neq i$.

The second term in (21) with the conditional expectation and without the negative sign can be simplified as follows:

$$\begin{aligned}
&\frac{(1+\epsilon)}{\tau(1-\frac{(1+\epsilon)^2}{\tau})}\mathbb{E}\left(\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n)\Big|\mathcal{F}_n\right)\\
&=\frac{(1+\epsilon)}{\tau(1-\frac{(1+\epsilon)^2}{\tau})}\sum_{i=1}^{N}\mathbb{E}\left[(d_n^i)^2\right]\nabla_{ii}^2 f(x_n)\\
&=\frac{(1+\epsilon)^2}{\tau(1-\frac{(1+\epsilon)^2}{\tau})}\sum_{i=1}^{N}\nabla_{ii}^2 f(x_n). \qquad (23)
\end{aligned}$$

Combining (22) and (23), the correctness of the Hessian estimate follows for the diagonal terms.

Off-diagonal terms in (18)

Consider the (k,l)th term in (18), with $k<l$. We obtain

$$\begin{aligned}
&\frac{1}{2(1+\epsilon)^2}\mathbb{E}\left[d_n^k d_n^l\left(\sum_{i=1}^{N}(d_n^i)^2\nabla_{ii}^2 f(x_n)\right.\right.\\
&\qquad\left.\left.+2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}d_n^i d_n^j\nabla_{ij}^2 f(x_n)\right)\Bigg|\mathcal{F}_n\right]\\
&=\frac{1}{2(1+\epsilon)^2}\sum_{i=1}^{N}\mathbb{E}\left(d_n^k d_n^l (d_n^i)^2\right)\nabla_{ii}^2 f(x_n)\\
&\quad+\frac{1}{(1+\epsilon)^2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbb{E}\left(d_n^k d_n^l d_n^i d_n^j\right)\nabla_{ij}^2 f(x_n) \qquad (24)\\
&=\nabla_{kl}^2 f(x_n).
\end{aligned}$$

Note that the first term on the RHS of (24) equals zero since $k\neq l$. The claim follows for the case of asymmetric Bernoulli perturbations. ■

Theorem 5. (Strong Convergence of Parameter): Assume (C1)–(C8). Then $x_n\to x^*$ a.s. as $n\to\infty$, where x_n is given by (13). ■

Proof: From Lemma 1, we observe that the gradient estimate $\widehat{\nabla} f(x_n)$ in (13) satisfies

$$\mathbb{E}\left[\widehat{\nabla} f(x_n)\Big|\mathcal{F}_n\right]=\nabla f(x_n)+\beta_n$$

where the bias term β_n is such that $\delta_n^{-2}\|\beta_n\|$ is uniformly bounded for sufficiently large n. The rest of the proof follows in a manner similar to the proof of [1, Theorem 1a][6]. ■

Theorem 6. (Strong Convergence of Hessian): Assume (C1)–(C10). Then $\overline{H}_n\to\nabla^2 f(x^*)$ a.s. as $n\to\infty$, where $\overline{H}_n$ is governed by (14) and $\widehat{H}_n$ defined according to either (15) or (16).

Proof: We first use a martingale convergence result to show that $\frac{1}{n+1}\sum_{m=0}^{n}(\widehat{H}_m-\mathbb{E}[\widehat{H}_m|x_m])\to 0$ a.s. Next, using Lemma 4, we can conclude that $\frac{1}{n+1}\sum_{m=0}^{n}\mathbb{E}[\widehat{H}_m|x_m]\to\nabla^2 f(x^*)$ a.s. and the claim follows. The reader is referred to Appendix II-A for the detailed proof. ■

We next present an asymptotic normality result for 2RDSA under the following additional assumptions:

(C11) For some $\zeta,\alpha_0,\alpha_1>0$ and for all $n, \mathbb{E}{\xi_n}^{2+\zeta}\leq\alpha_0, \mathbb{E}\xi_n^{\pm 2+\zeta}\leq\alpha_0, \mathbb{E}f(x_n)^{2+\zeta}\leq\alpha_1$ and $\mathbb{E}f(x_n\pm\delta_n d_n)^{2+\zeta}\leq\alpha_1$.

(C12) The operator Υ is chosen such that $\left\|\Upsilon(\overline{H}_n)-\overline{H}_n\right\|\to 0$ a.s. as $n\to\infty$.

Assumption (C11) is required to ignore the effects of noise, while (C12) together with Theorem 6 ensures that $\Upsilon(\overline{H}_n)$ converges to the true Hessian a.s. It is easy to see that the choice suggested in Remark 5 for Υ satisfies (C12).

The main result is as follows:

[6] Note that the proof of Theorem 5 does not assume any particular form of the Hessian estimate and only requires assumptions (C1)–(C7), which are similar to those in [1]. The only variation in our case, in comparison to [1], is the gradient estimation uses an RDSA scheme while [1] uses first-order SPSA. Thus, only the first step of the proof differs and in our case, Lemma 1 controls the bias with the same order as that of SPSA, leading to the final result.

Theorem 7. (Asymptotic Normality): Assume (C1)–(C12) and that $\nabla^2 f(x^*)^{-1}$ exists. Let $a_n = a_0/n^\alpha$ and $\delta_n = \delta_0/n^\gamma$, where $a_0, \delta_0 > 0, \alpha \in (0,1]$ and $\gamma \geq 1/6$. Let $\beta = \alpha - 2\gamma$. Let $\mathbb{E}(\xi_n^+ - \xi_n^-)^2 \to \sigma^2$ as $n \to \infty$. Then, we have

$$n^{\beta/2}(x_n - x^*) \xrightarrow{dist} \mathcal{N}(\mu, \Omega) \text{ as } n \to \infty, \tag{25}$$

where $\mathcal{N}(\mu, \Omega)$ is the multivariate Gaussian distribution with mean μ and covariance matrix Ω. The mean μ is defined as follows: $\mu = 0$ (an N-vector of all zeros) if $\gamma > \alpha/6$ and $\mu = k_\mu(a_0\delta_0^2(2a_0 - \beta^+)^{-1}\nabla^2 f(x^*)^{-1}T)$ if $\gamma = \alpha/6$, where k_μ, T and β^+ are as defined in Theorem 3. The covariance matrix Ω is defined as follows:

$$\Omega = \frac{a_0^2\sigma^2}{\delta_0^2(8a_0 - 4\beta^+)}\left(\nabla^2 f(x^*)^{-1}\right)^2.$$

Proof: As in the case of 2SPSA of [1], we verify conditions (2.2.1)–(2.2.3) of [33] to establish the result and the reader is referred to Appendix B of [34] for details. ■

C. (Asymptotic) Convergence Rates

Recall from Theorems 3 and Theorem 7 that we set $a_n = a_0/n^\alpha$ and $\delta_n = \delta_0/n^\gamma$, where $a_0, \delta_0 > 0$. Set $\alpha = 1$ and $\gamma = 1/6$. Let $\mathcal{AMSE}_{2\mathcal{RDSA}-Unif}(a_0, \delta_0)$ and $\mathcal{AMSE}_{2\mathcal{RDSA}-AB}(a_0, \delta_0)$ denote the AMSE for the uniform and asymmetric Bernoulli variants of 2RDSA, respectively. These quantities can be derived using Theorems 3 and 7 as follows:

$$\mathcal{AMSE}_{2\mathcal{RDSA}-Unif}(a_0, \delta_0) = \left(\frac{3.6\delta_0^2 a_0}{2a_0 - \frac{2}{3}}\left\|(\nabla^2 f(x^*)^{-1}T\right\|\right)^2 + \frac{a^2}{\delta_0^2(2a_0 - \frac{2}{3})}\text{trace}\left(\nabla^2 f(x^*)^{-1}S\nabla^2 f(x^*)^{-1}\right),$$

$$\mathcal{AMSE}_{2\mathcal{RDSA}-AB}(a_0, \delta_0) = \left(\frac{2\tau\delta_0^2 a_0}{(1+\epsilon)^2(2a_0 - \frac{2}{3})}\left\|(\nabla^2 f(x^*)^{-1}T\right\|\right)^2 + \frac{a^2}{\delta_0^2(2a_0 - \frac{2}{3})}\text{trace}\left(\nabla^2 f(x^*)^{-1}S\nabla^2 f(x^*)^{-1}\right),$$

where T is as defined in Theorem 3, $S = \frac{\sigma^2}{4}I$ and $\tau = E(d_n^i)^4 = \frac{(1+\epsilon)(1+(1+\epsilon)^3)}{(2+\epsilon)}$, for any $i = 1, \ldots, N$.

Recall from the discussion in Section II-C that 1RDSA has a problem dependence on the minimum eigenvalue λ_0 of $\nabla^2 f(x^*)$ for the step-size constant a_0. One can get rid of this dependence by using one of the 2RDSA variants and $a_0 = 1$. An alternative is to use iterate averaging, whose AMSE can be shown to be

$$\mathcal{AMSE}_{1\mathcal{RDSA}-Avg}(\delta_0) = \left(\frac{2.7}{\delta_0^2}\left\|(\nabla^2 f(x^*)^{-1}T\right\|\right)^2 + \frac{3}{4\delta_0^2}\text{trace}\left(\nabla^2 f(x^*)^{-1}S\nabla^2 f(x^*)^{-1}\right).$$

Notice that with either variant of 2RDSA one obtains the same rate as with iterate averaging and both these schemes do not have dependence on λ_0. Moreover, using arguments similar to [35] (see expressions (5.2) and (5.3) there), we obtain

$$\forall\delta_0, \mathcal{AMSE}_{2\mathcal{RDSA}-Unif}(1, \delta_0) < 2\min_{a_0 > \beta/(2\lambda_0)}\mathcal{AMSE}_{1\mathcal{RDSA}}(a_0, \delta_0),$$

$$\forall\delta_0, \mathcal{AMSE}_{2\mathcal{RDSA}-AB}(1, \delta_0) < 2\min_{a_0 > \beta/(2\lambda_0)}\mathcal{AMSE}_{1\mathcal{RDSA}}(a_0, \delta_0).$$

Note that the above bound holds for any choice of δ_0. Thus, 2RDSA is a robust scheme, as a wrong choice for a_0 would adversely affect the bound for 1RDSA, while 2RDSA has no such dependence on a_0.

Remark 7. (Iterate Averaging): Only from an "asymptotic" convergence rate viewpoint is it optimal to use larger step-sizes and iterate averaging. Finite-sample analysis (Theorem 2.4 in [18]) shows that the initial error (which depends on the starting point of the algorithm) is not forgotten subexponentially fast, but at the rate $1/n$, where n is the number of iterations. Thus, the effect of averaging kicks in only after enough iterations have passed and the bulk of the iterates are centered around the optimum. See Section 4.5 in [36] for a detailed discussion on this topic.

Comparing 2RDSA-Unif vs 2SPSA: Taking the ratio of AMSE of 2RDSA-Unif to that of 2SPSA, we obtain:

$$\frac{\mathcal{AMSE}_{2\mathcal{RDSA}-Unif}(1, \delta_0)}{\mathcal{AMSE}_{2\mathcal{SPSA}}(1, \delta_0)} = \frac{3.24(A) + (B)}{(A) + (B)}, \text{where} \tag{26}$$

$$(A) = \left(\frac{3\delta_0^2}{2}\left\|(\nabla^2 f(x^*)^{-1}T\right\|\right)^2, \tag{27}$$

$$(B) = \frac{3}{4\delta_0^2}\text{trace}\left(\nabla^2 f(x^*)^{-1}S\nabla^2 f(x^*)^{-1}\right). \tag{28}$$

However, 2SPSA uses four system simulations per iteration, while 2RDSA-Unif uses only three. So, in order to achieve a given accuracy, the ratio of the number of simulations needed for 2RDSA-Unif (denoted by $\hat{n}_{2\text{RDSA}-\text{Unif}}$) to that for 2SPSA (denoted by $\hat{n}_{2\text{SPSA}}$) is

$$\frac{\hat{n}_{2\mathcal{RDSA}-Unif}}{\hat{n}_{2\mathcal{SPSA}}} = \frac{3}{4} \times \frac{\mathcal{AMSE}_{2\mathcal{RDSA}-Unif}(1, \delta_0)}{\mathcal{AMSE}_{2\mathcal{SPSA}}(1, \delta_0)} = \frac{3}{4} \times \frac{3.24(A) + (B)}{(A) + (B)} = 1 + \frac{5.72(A) - (B)}{4(A) + 4(B)}.$$

Thus, if $5.72(A) - (B) < 0$, 2RDSA-Unif's AMSE is better than 2SPSA. On the other hand, if $5.72(A) - (B) \approx 0$, 2RDSA-Unif is comparable to 2SPSA and finally, in the case where $5.72(A) - (B) > 0$, 2SPSA is better, but the difference may be minor unless $5.72(A) \gg (B)$, as we have the term $4(A) + 4(B)$ in the denominator above. Note that the quantities (A) and (B) are problem-dependent, as they require knowledge of $\nabla^2 f(x^*)$ and T.

Unlike the first-order algorithms, one cannot conclude that 2SPSA is better than 2RDSA-Unif even when $\nabla^2 f(x^*)$ and T are known. 2RDSA-Unif uses fewer simulations per iteration, which may tilt the balance in favor of 2RDSA-Unif. We next show that asymmetric Bernoulli distributions are a better alternative, as they result in an AMSE for 2RDSA schemes that is lower than that for 2SPSA on *all* problem instances.

Comparing AMSE of 2RDSA-AB to that of 2SPSA: Taking the ratio of AMSE of 2RDSA-AB to that of 2SPSA, we obtain

$$\frac{\mathcal{AMSE}_{2\mathcal{RDSA}-AB}(1,\delta_0)}{\mathcal{AMSE}_{2\mathcal{SPSA}}(1,\delta_0)} = \frac{\frac{\tau^2}{(1+\epsilon)^4}(A)+(B)}{(A)+(B)},$$

where (A) and (B) are as defined in (27), (28). However, 2SPSA uses four system simulations per iteration, while 2RDSA-AB uses only three. So, in order to achieve a given accuracy, the ratio of the number of simulations needed for 2RDSA-AB (denoted by $\hat{n}_{2\mathrm{RDSA-AB}}$) to that for 2SPSA (denoted by $\hat{n}_{2\mathrm{SPSA}}$) is

$$\begin{aligned}\frac{\hat{n}_{2\mathcal{RDSA}-AB}}{\hat{n}_{2\mathcal{SPSA}}} &= \frac{3}{4}\times\frac{\mathcal{AMSE}_{2\mathcal{RDSA}-AB}(1,\delta_0)}{\mathcal{AMSE}_{2\mathcal{SPSA}}(1,\delta_0)}\\ &= \frac{3}{4}\times\frac{\frac{\tau^2}{(1+\epsilon)^4}(A)+(B)}{(A)+(B)}.\end{aligned}$$

For the sake of illustration, we set $\epsilon = 0.01$ in the asymmetric Bernoulli distribution (6) to obtain

$$\frac{\hat{n}_{2\mathcal{RDSA}-AB}}{\hat{n}_{2\mathcal{SPSA}}} = \frac{3.00059(A)+3(B)}{4(A)+4(B)} < 1.$$

Thus, the AMSE of 2RDSA-AB is better than 2SPSA on *all problem instances*, as (A) and (B) are positive (albeit unknown) quantities.

IV. Numerical Experiments

A. Setting

We use two functions, both in $N = 10$ dimensions for evaluating our algorithms.

Quadratic function: Let A be such that NA is an upper triangular matrix with each entry one and b is the N-dimensional vector of ones. Then, the quadratic objective function is defined as follows:

$$f(x) = x^{\mathsf{T}}Ax + b^{\mathsf{T}}x. \tag{29}$$

The optimum x^* for f is such that each coordinate of x^* is -0.9091, with $f(x^*) = -4.55$.

Fourth-order function: This is the function used for evaluating the second-order SPSA algorithm in [1] and is given as follows:

$$f(x) = x^{\mathsf{T}}A^{\mathsf{T}}Ax + 0.1\sum_{j=1}^{N}(Ax)_j^3 + 0.01\sum_{j=1}^{N}(Ax)_j^4, \tag{30}$$

where A is the same as that in the case of quadratic loss. The optimum $x^* = 0$ with $f(x^*) = 0$.

For any x, the noise is $[x^{\mathsf{T}}, 1]z$, where z is distributed as a multivariate Gaussian distribution in 11 dimensions with mean 0 and covariance $\sigma^2 I_{11\times 11}$. As remarked in [1], the motivation for this noise structure is to have most of the noise components depend on the iterate x_n and also a component z that ensures that the variance is at least σ^2.

B. Implementation

We implement the following algorithms[7]:

First-order: This class includes the RDSA schemes with uniform and asymmetric Bernoulli distributions—1RDSA-Unif and 1RDSA-AB, respectively, and regular SPSA with Bernoulli perturbations—1SPSA.

Second-order: This class includes 2RDSA-Unif and 2RDSA-AB—the second-order RDSA schemes with uniform and asymmetric Bernoulli distributions, respectively, and also regular second-order SPSA with Bernoulli perturbations—2SPSA.

For 1RDSA and 1SPSA, we set $\delta_n = 1.9/n^{0.101}$ and $a_n = 1/(n+50)$. For 2RDSA and 2SPSA, we set $\delta_n = 3.8/n^{0.101}$ and $a_n = 1/n^{0.6}$. As remarked in [1], these rates are approximately the slowest allowed in theory and have been used earlier for SPSA implementations—see [36]. For uniform perturbation variants, we set the distribution parameter $\eta = 1$ and for the asymmetric Bernoulli variants, we set the distribution parameter ϵ as follows: $\epsilon = 0.0001$ for 1RDSA-AB and $\epsilon = 1$ for 2RDSA-AB. These choices are motivated by a sensitivity study with different choices for ϵ—see Table VI in Appendix III. For all the algorithms, the initial point x_0 is the 10-dimensional vector of ones. To keep the iterates stable, each coordinate of the parameter θ is projected onto the set $[-2.048, 2.047]$. All results are averages over 1000 replications.

We use normalized mean square error (NMSE) as the performance metric for comparing algorithms. This quantity is defined as follows:

$$\mathrm{NMSE} = \frac{\|x_{n_{\mathrm{end}}} - x^*\|^2}{\|x_0 - x^*\|^2}$$

where $x_{n_{\mathrm{end}}}$ is the algorithm iterate at the end of the simulation. Note, n_{end} is algorithm-specific and a function of the number of measurements. For instance, with 2000 measurements, $n = 1000$ for both 1SPSA and both 1RDSA variants, as they use two measurements per iteration. On the other hand, for 2SPSA and both 2RDSA variants, an initial 20% of the measurements were used up by 1SPSA/1RDSA and the resulting iterates were used to initialize the corresponding second-order method. Thus, with 2000 measurements available, the initial 400 measurements are used for 1SPSA/1RDSA and the remaining 1600 are used up by 2SPSA/2RDSA. This results in n_{end} of $1600/4 = 400$ for 2SPSA and $1600/3 \approx 533$ for 2RDSA algorithms. Note that the difference here is due to the fact that 2RDSA uses 3 simulations per iteration, while 2SPSA needs 4.

C. Results: Quadratic Objective

Tables II, III present the normalized mean square error (NMSE) for both first and second-order algorithms with quadratic objective (29) and noise parameter σ set to 0.001 and 0.

Observation 1: Among first-order schemes, 1RDSA-AB performs on par with 1SPSA, while 1RDSA-Unif is subpar.

The NMSE of 1RDSA-AB is comparable to that of 1SPSA, while 1RDSA-Unif results in a higher NMSE. This is consistent with the asymptotic rate results discussed earlier in Section II-C.

Observation 2: Second-order schemes outperform their first-order counterparts, and 2RDSA-AB performs best in this class.

[7] The implementation is available at https://github.com/prashla/RDSA/archive/master.zip

TABLE II
NMSE FOR QUADRATIC OBJECTIVE (29) AND NOISE PARAMETER $\sigma = 0.001$: STANDARD ERROR FROM 1000 REPLICATIONS SHOWN IN PARENTHESES

First-order Algorithms		
# function measurements	1000	2000
1SPSA	$\mathbf{.0415}$ $(\mathbf{.00052})$	.0342 (.00047)
1RDSA-Unif	.0453 (.00057)	.0367 (.00053)
1RDSA-AB	.0418 (.00054)	$\mathbf{.0338}$ $(\mathbf{.00049})$
Second-order Algorithms		
# function measurements	1000	2000
2SPSA	.00105 $(2.25E^{-5})$	.0000036 $(7.62E^{-8})$
2RDSA-Unif	.000096 $(2.48E^{-6})$	.0000045 $(6.61E^{-8})$
2RDSA-AB	$\mathbf{.000084}$ $(\mathbf{2.25E^{-6}})$	$\mathbf{.0000022}$ $(\mathbf{3.35E^{-8}})$

TABLE III
NMSE FOR QUADRATIC OBJECTIVE (29) AND NOISE PARAMETER $\sigma = 0$: STANDARD ERROR FROM 1000 REPLICATIONS SHOWN IN PARENTHESES

First-order Algorithms		
# function measurements	1000	2000
1SPSA	$\mathbf{.0415}$ $(\mathbf{.00052})$	.0342 (.00047)
1RDSA-Unif	.0453 (.00057)	.0367 (.00053)
1RDSA-AB	.0418 (.00054)	$\mathbf{.0337}$ $(\mathbf{.00049})$
Second-order Algorithms		
# function measurements	1000	2000
2SPSA	.00076 $(1.59E^{-5})$	$6.77E^{-7}$ $(2.78E^{-8})$
2RDSA-Unif	.000093 $(2.48E^{-6})$	$\mathbf{2.42E^{-9}}$ $(\mathbf{1.11E^{-10}})$
2RDSA-AB	$\mathbf{.000083}$ $(\mathbf{2.25E^{-6}})$	$2.90E^{-9}$ $(1.41E^{-10})$

The first part of the observation is consistent with earlier results for the low noise regime (i.e., $\sigma = 0.01$), for instance, see [1]. Further, the gains of using second-order schemes are more noticeable in the zero-noise regime (see Table III). Moreover, 2RDSA-AB results in the best NMSE. In fact, running 2RDSA-AB for 400 iterations, which was the number used for 2SPSA with 2000 measurements available, the resulting NMSE values was found to be 2.34×10^{-6}, which is better than the corresponding 400-iteration result of 3.60×10^{-6} for 2SPSA (see Table II), while using only 75% as many simulations.

D. Results: Fourth-Order Objective

Table IV presents results similar to those in Table II for the fourth-order objective function (30) with the noise parameter σ set to 0.001. Considering that the fourth-order objective is more difficult to optimize in comparison to the quadratic one, we run all algorithms with a simulation budget of 10000 function evaluations. From the results in Table IV, one can draw conclusions similar to those in observations 1 and 2 above, except that 2RDSA-Unif shows the best performance among second-order schemes.

E. Traffic Light Control Application

As a real-world application, we consider the problem of traffic light control (TLC) at junctions of a road network. In practice, obtaining exact queue length information is difficult, but one can obtain coarse estimates along the lanes of the road network,

TABLE IV
NMSE FOR FOURTH-ORDER OBJECTIVE (30) AND NOISE PARAMETER $\sigma = 0.001$: STANDARD ERROR FROM 1000 REPLICATIONS SHOWN IN PARENTHESES

First-order Algorithms		
# function measurements	2000	10000
1SPSA	$\mathbf{.137}$ $(\mathbf{.0014})$	.114 (.0014)
1RDSA-Unif	.138 (.0013)	.118 (.0012)
1RDSA-AB	.135 (.0014)	$\mathbf{.114}$ $(\mathbf{.0012})$
Second-order Algorithms		
# function measurements	2000	10000
2SPSA	.032 $(5.38E^{-4})$	.0101 $(1.96E^{-4})$
2RDSA-Unif	$\mathbf{.015}$ $(\mathbf{2.64E^{-4}})$	$\mathbf{.0017}$ $(\mathbf{3.65E^{-5}})$
2RDSA-AB	.048 $(9.01E^{-4})$	.037 $(1.19E^{-3})$

for instance, by placing magnetic sensor loops at some distance from the junction. The challenge is to choose the optimal locations for placing sensor loops to infer congestion information for any lane in the road network considered—a problem studied earlier in [37] and [23].

Setting 1: The simulation setup is identical to that in [23] and is described briefly in the following: Let $x = (L_1, \ldots, L_5, T_1, \ldots, T_3)^{\mathsf{T}}$, where $L_i, i = 1, \ldots, 5$ are the thresholds used to obtain coarse queue length estimates, with $L_1 < \ldots < L_5$ and $T_i, i = 1, \ldots, 3$ are the thresholds on the elapsed times that measure the time since a lane has turned red. For any given x, we setup a parameterized Markov process with state at any instant j defined as

$$X_j(x) := (\tilde{q}_j^1(x), \ldots, \tilde{q}_j^K(x), \tilde{t}_j^1(x), \ldots, \tilde{t}_j^K(x))$$

where K is the total number of lanes in the road network, $\tilde{q}_j^k(x)$ (resp. $\tilde{t}_j^k(x)$) denote the coarse queue length (respectively, elapsed time) estimates for lane k. The actual queue length on any lane could either be $< L_1$ or between L_1 and L_2 and so on until $> L_5$, leading to the corresponding graded assignment of $0, 0.2, 0.4, 0.6, 0.8$ or 1 to $\tilde{q}_j^k(x)$, for each $k = 1, \ldots, K$. The coarse elapsed time $\tilde{t}_j^k(x)$ is set similarly using $T_1, \ldots, T_3$.

The goal is to solve a problem similar to (1), except that the objective function is a long-run average cost, i.e.,

$$f(x) = \lim_{m \to \infty} \frac{1}{m} \sum_{j=0}^{m-1} g(X_j(x)), \text{where}$$

$$g(X_j(x)) = \sum_{k \in I_p} \alpha_1 \left(\tilde{q}_j^k(x) + \tilde{t}_j^k(x)\right) + \sum_{k \notin I_p} \beta_1 \left(\tilde{q}_j^k(x) + \tilde{t}_j^k(x)\right)$$

where $0.5 < \alpha_1 < 1$ and $\beta_1 = (1 - \alpha_1)$. This assignment to α_1 ensures that the priority lanes (specified by the set I_p) get more importance in $g(\cdot)$. We set $\alpha_1 = 0.6$ in our experiments.

The underlying TLC algorithm is PTLC [23]. At any instant j, PTLC assigns a priority value for each lane k in a graded fashion such that lanes with high congestion (i.e., $\tilde{q}_j^k(x) = 1$) and longer elapsed times (i.e., $\tilde{t}_j^k(x) = 1$) get the highest priority, those with low congestion/elapsed time get lowest priority and the configurations in-between are assigned intermediate priorities. The sign configuration chosen at instant j is the one that maximizes the sum of priorities over all lanes. The reader is referred to [23, Section 5] for further details.

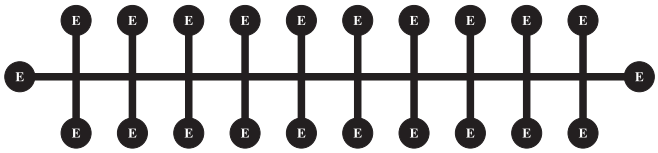

Fig. 2. Road Network used for our experiments.

TABLE V
EMPIRICAL AVERAGE COST $\hat{f}(x_{n_{\text{end}}})$ FOR SECOND-ORDER ALGORITHMS ON A TEN-JUNCTION CORRIDOR: STANDARD ERROR FROM 50 REPLICATIONS SHOWN IN PARENTHESES

Algorithm	$\hat{f}(x_{n_{\text{end}}})$
2SPSA	28.94 (3.27)
2RDSA-Unif	21.47 (2.08)
2RDSA-AB	**8.89 (0.10)**

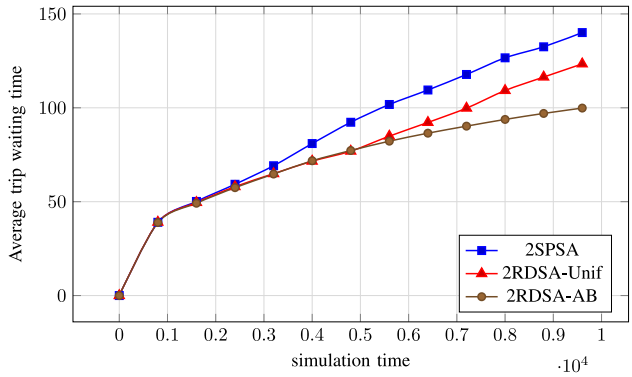

Fig. 3. Comparison of second-order algorithms using average trip waiting time on a ten-junction corridor.

The authors in [37], [23] employ first and second-order SPSA for optimizing the thresholds and run it in conjunction with PTLC. In this paper, we apply our 2RDSA schemes for optimizing the thresholds and also compare their performance against 2SPSA using GLD traffic simulator [38] on a ten-junction corridor (see Fig. 2). The parameters concerning 2RDSA/2SPSA algorithms are set as noted in Section IV-B.

The experiments involved training and testing phases. In the training phase, the second-order algorithm is run to obtain the parameter $x_{n_{\text{end}}}$. Here, n_{end} is a function of the simulation budget, which specifies the number of simulated trajectories of the underlying Markov process. Each trajectory was of $m = 250$ steps and the empirical average $\hat{f}(x_n) = \frac{1}{m}\sum_{j=0}^{m-1} g(X_j(x_n))$ was used to update the parameter x_n. In the testing phase, 50 independent simulations are run with the parameter $x_{n_{\text{end}}}$.

Results: For evaluating the algorithms, we use empirical average cost $\hat{f}(x_{n_{\text{end}}}) = \dfrac{1}{m}\displaystyle\sum_{j=0}^{m-1} g(X_j(x_{n_{\text{end}}}))$ as a proxy for the long-run average cost $f(\cdot)$, with $m = 10000$. Table V reports the mean and standard-error for $\hat{f}(x_{n_{\text{end}}})$ from the testing phase, with a simulation budget of 2000 function evaluations in the training phase. From a traffic application standpoint, we also measure the average trip waiting time for each of the algorithms in the testing phase and plot the waiting times in Fig. 3. Based on observed empirical average cost and average waiting times, it is apparent that 2RDSA schemes outperform 2SPSA, with 2RDSA-AB performing the best.

Remark 8. (Enhanced 2SPSA): In [25], enhancements to the 2SPSA algorithm incorporated adaptive feedback and weighting to improve Hessian estimates. However, preliminary numerical experiments that we conducted for enhanced 2SPSA with the parameters recommended in [25] indicate that the benefits of such a scheme kick in only after a large number of iterations [39]. Moreover, the numerical results presented in Tables II–IV make a fair comparison in the sense that, except for the perturbations every other parameter (e.g., step-sizes a_n, perturbation constants δ_n, initial point x_0) is kept constant across algorithms in each class (first/second-order). The results demonstrate that it is indeed advantageous to use uniform/asymmetric Bernoulli perturbations.

V. CONCLUSIONS

We considered a general problem of optimization under noisy observations and presented the first adaptive random directions Newton algorithm. Two sets of i.i.d. random perturbations were analyzed: symmetric uniformly distributed and asymmetric Bernoulli distributed. In addition, we also presented a simple gradient search scheme using two sets of perturbations. While our gradient search scheme requires the same number of perturbations and system simulations per iteration as the simultaneous perturbation gradient scheme of [14], our Newton scheme only requires half the number of perturbations and three-fourths the number of simulations as compared to the simultaneous perturbation Newton algorithm of [1]. We proved the convergence of our algorithms and analyzed their rates of convergence using the asymptotic mean square error (AMSE). From this analysis, we concluded that the asymmetric Bernoulli perturbation variants exhibit the best AMSE for both first- and second-order RDSA schemes. Furthermore, our numerical experiments show that our Newton algorithm requires only 75% of the number of function evaluations as required by the Newton algorithm of [1] while providing the same accuracy levels as the latter algorithm.

As future work, it would be of interest to extend our algorithms to scenarios where the noise random variables form a parameterized Markov process and to develop multiscale algorithms in this setting for long-run average or infinite horizon discounted costs. Such algorithms will be of relevance in the context of reinforcement learning, for instance, as actor-critic algorithms.

APPENDIX I
PROOFS FOR 1RDSA

A. Proof of Theorem 1

Proof: We use the proof technique of [14] (in particular, Lemma 1 there) in order to prove the main claim here.

Notice that

$$\mathbb{E}\left[\frac{y_n^+ - y_n^-}{2\delta_n}\middle|\mathcal{F}_n\right]$$
$$= \mathbb{E}\left[d_n\left(\frac{f(x_n+\delta_n d_n) - f(x_n-\delta_n d_n)}{2\delta_n}\right)\middle|\mathcal{F}_n\right].$$

The equality above follows from the fact that $\mathbb{E}\left[d_n\left(\frac{\xi_n^+-\xi_n^-}{2\delta_n}\right)\Big|\mathcal{F}_n\right]=0$ from (A2) and (A4). We now analyse the term on the RHS above for both uniformly distributed perturbations and asymmetric Bernoulli perturbations.

Case 1: Uniform Perturbations: Let $\nabla^2 f(\cdot)$ denote the Hessian of f. By Taylor's series expansions, we obtain, a.s.,

$$f(x_n \pm \delta_n d_n) = f(x_n) \pm \delta_n d_n^{\mathsf{T}} \nabla f(x_n) + \frac{\delta_n^2}{2} d_n^{\mathsf{T}} \nabla^2 f(x_n) d_n \pm \frac{\delta_n^3}{6} \nabla^3 f(\tilde{x}_n^+)(d_n \otimes d_n \otimes d_n),$$

where $\otimes$ denotes the Kronecker product and $\tilde{x}_n^+$ (respectively, $\tilde{x}_n^-$) are on the line segment between x_n and $(x_n+\delta_n d_n)$ (respectively, $(x_n-\delta_n d_n)$). Hence

$$\begin{aligned}&\mathbb{E}\left[d_n\left(\frac{f(x_n+\delta_n d_n)-f(x_n-\delta_n d_n)}{2\delta_n}\right)\Big|\mathcal{F}_n\right]\\&=\mathbb{E}\left[d_n d_n^{\mathsf{T}}\,\nabla f(x_n)|\,\mathcal{F}_n\right]\\&+\mathbb{E}\left[\frac{\delta_n^2}{12}d_n(\nabla^3 f(\tilde{x}_n^+)+\nabla^3 f(\tilde{x}_n^-))(d_n\otimes d_n\otimes d_n)\Big|\mathcal{F}_n\right].\end{aligned} \tag{31}$$

The first term on the RHS above can be simplified as follows:

$$\mathbb{E}\left[d_n d_n^{\mathsf{T}}\,\nabla f(x_n)|\,\mathcal{F}_n\right]=\mathbb{E}\left[d_n d_n^{\mathsf{T}}\right]\nabla f(x_n)=\frac{\eta^2}{3}\nabla f(x_n). \tag{32}$$

In the above, the first equality follows from (A4) and the last equality follows from $\mathbb{E}[(d_n^i)^2]=\frac{\eta^2}{3}$ and $\mathbb{E}[d_n^i d_n^j]=\mathbb{E}[d_n^i]\mathbb{E}[d_n^j]=0$ for $i\neq j$.

Now, the lth coordinate of the second term in the RHS of (31) can be upper-bounded as follows:

$$\begin{aligned}&\mathbb{E}\left[\frac{\delta_n^2}{12}d_n^l(\nabla^3 f(\tilde{x}_n^+)+\nabla^3 f(\tilde{x}_n^-))(d_n\otimes d_n\otimes d_n)\Big|\mathcal{F}_n\right]\\&\leq\frac{\alpha_0\delta_n^2}{6}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\mathbb{E}\left(d_n^l d_n^{i_1} d_n^{i_2} d_n^{i_3}\right)\leq\frac{\alpha_0\delta_n^2\eta^4 N^3}{6}.\end{aligned} \tag{33}$$

The first inequality above follows from (A1), while the second inequality follows from the fact that $\left|d_n^l\right|\leq\eta, l=1,\ldots,N$. The claim follows by plugging (32) and (33) into (31).

Case 2 Asymmetric Bernoulli Perturbations: The proof follows in an analogous fashion as above, after noting that the scaling factor of $\frac{1}{(1+\epsilon)}$ in (7) cancels out $\mathbb{E}[(d_n^i)^2]=(1+\epsilon)$ and the bound in (33) gets replaced by $\left(\frac{\alpha_0\delta_n^2(1+\epsilon)^4N^3}{6}\right)$. ■

B. Proof of Theorem 2

Proof: We first rewrite the update rule (4) as follows:

$$x_{n+1}=x_n-a_n(\nabla f(x_n)+\eta_n+\beta_n), \tag{34}$$

where $\eta_n=\widehat{\nabla}f(x_n)-\mathbb{E}(\widehat{\nabla}f(x_n)\mid\mathcal{F}_n)$ is a martingale difference term and $\beta_n=\mathbb{E}(\widehat{\nabla}f(x_n)\mid\mathcal{F}_n)-\nabla f(x_n)$ is the bias in the gradient estimate. Convergence of (34) can be inferred from [7, Theorem 2.3.1], provided we verify that [7, Assumptions A2.2.1–A2.2.3 and A2.2.4"] are satisfied. We list these assumptions as (B1)–(B4) below.

(B1) ∇f is a continuous $\mathbb{R}^N$-valued function.
(B2) The sequence $\beta_n, n\geq 0$ is almost surely bounded with $\beta_n\to 0$ almost surely as $n\to\infty$.
(B3) The step-sizes $a_n, n\geq 0$ satisfy $a(n)\to 0$ as $n\to\infty$ and $\sum_n a_n=\infty$.
(B4) $\{\eta_n, n\geq 0\}$ is a sequence such that for any $\epsilon>0$

$$\lim_{n\to\infty}P\left(\sup_{m\geq n}\left\|\sum_{i=n}^{m}a_i\eta_i\right\|\geq\epsilon\right)=0.$$

The above assumptions can be verified for (34) as follows:

1) (A1) implies (B1).
2) (A5) together with (33) in the proof of Lemma 1 imply that the bias β_n is almost surely bounded. Further, Lemma 1 implies that β_n is of the order $O(\delta_n^2)$ and since $\delta_n\to 0$ as $n\to\infty$ [see (A5)], we have that $\beta_n\to 0$. Thus, (B2) is satisfied.
3) (A5) implies (B3).
4) We now verify (B4) using arguments similar to those used in [36, Ch. 7.3]: We first recall a martingale inequality attributed to Doob (also given as (2.1.7) in [7, pp. 27])

$$P\left(\sup_{m\geq 0}\|W_m\|\geq\epsilon\right)\leq\frac{1}{\epsilon^2}\lim_{m\to\infty}\mathbb{E}\|W_m\|^2. \tag{35}$$

We apply the above inequality in our setting to the martingale sequence $\{W_n\}$, where $W_n:=\sum_{i=0}^{n-1}a_i\eta_i, n\geq 1$, to obtain

$$\begin{aligned}P\left(\sup_{m\geq n}\left\|\sum_{i=n}^{m}a_i\eta_i\right\|\geq\epsilon\right)&\leq\frac{1}{\epsilon^2}\mathbb{E}\left\|\sum_{i=n}^{\infty}a_i\eta_i\right\|^2\\&=\frac{1}{\epsilon^2}\sum_{i=n}^{\infty}a_i^2\mathbb{E}\|\eta_i\|^2.\end{aligned} \tag{36}$$

The last equality above follows by observing that, for $m<n$, $\mathbb{E}(\eta_m\eta_n)=\mathbb{E}(\eta_m\mathbb{E}(\eta_n\mid\mathcal{F}_n))=0$.

Letting $x_n^{\pm}=x_n\pm\delta_n d_n$ and using the identity $\mathbb{E}\|X-E[X\mid\mathcal{F}_n]\|^2\leq\mathbb{E}\|X\|^2$ for any random variable X, we bound $\mathbb{E}\|\eta_n\|^2$ as follows:

$$\begin{aligned}\mathbb{E}\|\eta_n\|^2&\leq\mathbb{E}\left\|\widehat{\nabla}f(x_n)\right\|^2\\&=\mathbb{E}\left(\|d_n\|^2\left(\left(\frac{\xi_n^+-\xi_n^-}{2\delta_n}\right)^2+\left(\frac{f(x_n^+)-f(x_n^-)}{2\delta_n}\right)^2\right.\right.\\&\quad\left.\left.+2\left(\frac{\xi_n^+-\xi_n^-}{2\delta_n}\right)\left(\frac{f(x_n^+)-f(x_n^-)}{2\delta_n}\right)\right)\right)\\&=\mathbb{E}\left(\|d_n\|^2\left(\frac{\xi_n^+-\xi_n^-}{2\delta_n}\right)^2\right)+\mathbb{E}\|d_n\|^2\left(\frac{f(x_n^+)-f(x_n^-)}{2\delta_n}\right)^2\end{aligned} \tag{37}$$

$$\leq\frac{C}{\delta_n^2},\text{for some } C<\infty. \tag{38}$$

The equality in (37) follows from (A2), whereas the inequality in (38) follows from (A3) and the fact that the perturbations d_n have finite moments.

Plugging (19) into (36), we obtain

$$\lim_{n\to\infty} P\left(\sup_{m\geq n}\left\|\sum_{i=n}^{m} a_i\eta_i\right\| \geq \epsilon\right) \leq \frac{C}{\epsilon^2}\lim_{n\to\infty}\sum_{i=n}^{\infty}\frac{a_i^2}{\delta_i^2} = 0.$$

The equality above follows from the fact that $\sum_n \left(\frac{a_n}{\delta_n}\right)^2 < \infty$ [see (A5)]. The claim follows from [7, Theorem 2.3.1]. ■

C. Proof of Theorem 3

Proof: Follows from Proposition 1 of [8] after observing the following facts:

Uniform perturbations: $\frac{3}{\eta^2}\mathbb{E}[d_n d_n^{\mathsf{T}}] = I$ and $\frac{9}{\eta^4}\mathbb{E}[(d_n^i)^4] = 1.8$ for any $i = 1, \ldots, N$.

Asymmetric Bernoulli perturbations: $\frac{1}{(1+\epsilon)}\mathbb{E}[d_n d_n^{\mathsf{T}}] = I$ and $\frac{1}{(1+\epsilon)^2}\mathbb{E}[(d_n^i)^4] = \frac{(1+\epsilon)(1+(1+\epsilon)^3)}{(2+\epsilon)(1+\epsilon)^2}$ for any $i = 1, \ldots, N$. ■

APPENDIX II PROOFS FOR 2RDSA

A. Proof of Theorem 6

Proof: The proof proceeds in exactly the same manner as the proof of [1, Theorem 2a] and we sketch below the main arguments involved.

Let $W_m = \widehat{H}_m - \mathbb{E}\left[\widehat{H}_m \middle| x_m\right]$. Then, we know that $\mathbb{E}W_m = 0$. In addition, we have $\sum_m \frac{\mathbb{E}\|W_m\|^2}{m^2} < \infty$. The latter follows by first observing that $\mathbb{E}\left[\delta_m^4 \left\|\widehat{H}_m\right\|^2\right] < \infty, \forall m$ uniformly as a consequence of (C9) and then coupling this fact with (C8). Now, applying a martingale convergence result from [40, p. 397] to W_m, we obtain

$$\frac{1}{n+1}\sum_{m=0}^{n}\left(\widehat{H}_m - \mathbb{E}\left[\widehat{H}_m \middle| x_m\right]\right) \to 0 \quad \text{a.s.} \qquad (39)$$

From Proposition 4, we know that $\mathbb{E}\left[\widehat{H}_n \middle| x_n\right] = \nabla^2 f(x_n) + O(\delta_n^2)$. Hence

$$\frac{1}{n+1}\sum_{m=0}^{n}\mathbb{E}\left[\widehat{H}_m \middle| x_m\right] = \frac{1}{n+1}\sum_{m=0}^{n}\left(\nabla^2 f(x_m) + O(\delta_m^2)\right)$$
$$\to \nabla^2 f(x^*) \quad \text{a.s.}$$

The final step above follows from the fact that the Hessian is continuous near x_n and Theorem 5 which implies x_n converges almost surely to x^*. Thus, we obtain $\frac{1}{n+1}\sum_{m=0}^{n}\widehat{H}_m \to \nabla^2 f(x^*)$ a.s. and the claim follows since $\overline{H}_m = \frac{1}{n+1}\sum_{m=0}^{n}\widehat{H}_m$.

APPENDIX III ADDITIONAL NUMERICAL RESULTS

TABLE VI
NMSE OF 1RDSA-AB AND 2RDSA-AB AS A FUNCTION OF DISTRIBUTION PARAMETER ϵ, WITH QUADRATIC OBJECTIVE (29), $\sigma = 0.001$ AND USING 2000 FUNCTION MEASUREMENTS

ϵ value	1RDSA-AB		2RDSA-AB	
0.000001	.0338	$(4.87E^{-4})$	.0721	$(8.95E^{-4})$
0.00001	.0338	$(4.87E^{-4})$	.0793	$(1.46E^{-3})$
0.0001	.0338	$(4.87E^{-4})$	.0624	$(1.34E^{-3})$
0.001	.0338	$(4.87E^{-4})$	.0839	$(2.36E^{-3})$
0.01	.0339	$(4.87E^{-4})$	.0832	$(4.04E^{-2})$
0.1	.0338	$(4.81E^{-4})$	.235	$(1.13E^{-2})$
0.2	.0337	$(4.87E^{-4})$	.102	$(9.10E^{-3})$
0.5	.0342	$(5.00E^{-4})$	.000145	$(1.43E^{-4})$
1	.0354	$(5.09E^{-4})$	.0000022	$(3.67E^{-8})$
2	.0387	$(5.76E^{-4})$	.0000022	$(3.35E^{-8})$
5	.0521	$(8.10E^{-4})$	.0000029	$(4.60E^{-8})$

REFERENCES

[1] J. C. Spall, "Adaptive stochastic approximation by the simultaneous perturbation method," *IEEE Trans. Autom. Control*, vol. 45, pp. 1839–1853, 2000.

[2] M. C. Fu, Ed., *Handbook of Simulation Optimization*. New York: Springer, 2015.

[3] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.

[4] Y. C. Ho and X. R. Cao, *Perturbation Analysis of Discrete Event Dynamical Systems*. Boston, MA: Kluwer, 1991.

[5] R. Y. Rubinstein and A. Shapiro, *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. New York: Wiley, 1993.

[6] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Statist.*, vol. 23, pp. 462–466, 1952.

[7] H. J. Kushner and D. S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer Verlag, 1978.

[8] D. C. Chin, "Comparative study of stochastic algorithms for system optimization based on gradient approximations," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 27, no. 2, pp. 244–249, 1997.

[9] R. Y. Rubinstein, *Simulation and the Monte Carlo Method*. New York: Wiley, 1981.

[10] S. Bhatnagar, "Adaptive Newton-based smoothed functional algorithms for simulation optimization," *ACM Trans. Model. Comput. Simul.*, vol. 18, no. 1, pp. 2:1–2:35, 2007.

[11] S. Bhatnagar, H. L. Prasad, and L. A. Prashanth, *Stochastic Recursive Algorithms for Optimization: Simultaneous Perturbation Methods (Lecture Notes in Control and Information Sciences)*. New York: Springer, 2013, vol. 434.

[12] V. Y. Katkovnik and Y. Kulchitsky, "Convergence of a class of random search algorithms," *Automat. Remote Control*, vol. 8, pp. 1321–1326, 1972.

[13] M. A. Styblinski and T.-S. Tang, "Experiments in nonconvex optimization: Stochastic approximation with function smoothing and simulated annealing," *Neur. Netw.*, vol. 3, pp. 467–483, 1990.

[14] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Autom. Control*, vol. 37, no. 3, pp. 332–341, 1992.

[15] V. Fabian, "Stochastic approximation of minima with improved asymptotic speed," *Annal. Mathemat. Statist.*, pp. 191–200, 1967.

[16] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM J. Control and Optimiz.*, vol. 30, no. 4, pp. 838–855, 1992.

[17] D. Ruppert, "Stochastic approximation," *Handbook of Sequential Analysis*, pp. 503–529, 1991.

[18] M. Fathi and N. Frikha, "Transport-entropy inequalities and deviation estimates for stochastic approximation schemes," *Electron. J. Probab*, vol. 18, no. 67, pp. 1–36, 2013.

[19] V. Fabian, "Stochastic approximation," in *Optimizing Methods in Statistics* (ed. J. J. Rustagi). New York: Academic Press, 1971, pp. 439–470.

[20] D. Ruppert, "A Newton-Raphson version of the multivariate Robbins-Monro procedure," *Annal. Statist.*, vol. 13, pp. 236–245, 1985.
[21] S. Bhatnagar, "Adaptive multivariate three-timescale stochastic approximation algorithms for simulation based optimization," *ACM TOMACS*, vol. 15, no. 1, pp. 74–107, 2005.
[22] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge, U.K.: Cambridge University Press, 2008.
[23] S. Bhatnagar and L. A. Prashanth, "Simultaneous perturbation Newton algorithms for simulation optimization," *J. Optimiz. Theory and Applic.*, vol. 164, no. 2, pp. 621–643, 2015.
[24] X. Zhu and J. C. Spall, "A modified second-order SPSA optimization algorithm for finite samples," *Int. J. Adapt. Control Signal Process.*, vol. 16, pp. 397–409, 2002.
[25] J. C. Spall, "Feedback and weighting mechanisms for improving Jacobian estimates in the adaptive simultaneous perturbation algorithm," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1216–1229, 2009.
[26] A. D. Flaxman, A. T. Kalai, and H. B. McMahan, "Online convex optimization in the bandit setting: Gradient descent without a gradient," in *ACM-SIAM Symp. Discrete Algorithms*, 2005, pp. 385–394.
[27] Y. Nesterov and V. Spokoiny, "Random gradient-free minimization of convex functions," *Found. Computat. Mathemat.*, 2011.
[28] J. C. Duchi, M. I. Jordan, M. J. Wainwright, and A. Wibisono, "Optimal rates for zero-order convex optimization: The power of two function evaluations," *arXiv Preprint arXiv:1312.2139*, 2013.
[29] G. Marsaglia, "Choosing a point from the surface of a sphere," *Annal. Mathemat. Statist.*, vol. 43, no. 2, pp. 645–646, 1972.
[30] X. Hu, L. A. Prashanth, A. György, and C. Szepesvári, "(Bandit) Convex Optimization with Biased Noisy Gradient Oracles," in *AISTATS*, 2016.
[31] L. Gerencsér, "Convergence rate of moments in stochastic approximation with simultaneous perturbation gradient approximation and resetting," *IEEE Trans. Autom. Control*, vol. 44, no. 5, pp. 894–905, 1999.
[32] P. Gill, W. Murray, and M. Wright, *Practical Optimization*. New York: Academic Press, 1981.
[33] V. Fabian, "On asymptotic normality in stochastic approximation," *Annal. Mathemat. Statist.*, pp. 1327–1332, 1968.
[34] L. Prashanth, S. Bhatnagar, M. Fu, and S. Marcus, "Adaptive system optimization using random directions stochastic approximation," *arXiv Preprint arXiv:1502.05577v2*, 2015.
[35] J. Dippon and J. Renz, "Weighted means in stochastic approximation of minima," *SIAM J. Control Optimiz.*, vol. 35, no. 5, pp. 1811–1827, 1997.
[36] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*., New York: Wiley, 2005, 65.
[37] L. A. Prashanth and S. Bhatnagar, "Threshold tuning using stochastic optimization for graded signal control," *IEEE Trans. Vehic. Technol.*, vol. 61, no. 9, pp. 3865–3880, Nov. 2012.
[38] M. Wiering, J. Vreeken, J. van Veenen, and A. Koopman, "Simulation and optimization of traffic in a city," in *Proc. IEEE Intell. Vehic. Symp.*, Jun. 2004, pp. 453–458.
[39] D. S. K. Reddy, L. Prashanth, and S. Bhatnagar, "Improved Hessian estimation for adaptive random directions stochastic approximation," in *Proc. CDC (to appear)*, 2016.
[40] R. G. Laha and V. K. Rohatgi, *Probability Theory*. New York: Wiley, 1979.

Prashanth L. A. received the Master's and Ph.D. degrees in computer science and automation from the Indian Institute of Science in 2008 and 2013, respectively.

He is currently a postdoctoral researcher at the Institute for Systems Research, University of Maryland–College Park. His research interests are in reinforcement learning, stochastic optimization and multi-armed bandits, with applications in transportation systems, wireless networks and recommendation systems.

Dr. L.A. was awarded the third prize for his Ph.D. dissertation by the IEEE Intelligent Transportation Systems Society (ITSS).

Shalabh Bhatnagar (SM'05) received the Bachelor's (Hons.) degree in physics from the University of Delhi, Delhi, India, in 1988 and the Master's and Ph.D. degrees in electrical engineering from the Indian Institute of Science, Bangalore, in 1992 and 1997, respectively.

He is a Professor with the Department of Computer Science and Automation, Indian Institute of Science, Bangalore. His research interests are in stochastic approximation theory and stochastic optimization and control, as well as applications in communication, wireless, and vehicular traffic networks.

Michael Fu (F'09) received degrees in mathematics and electrical engineering and computer science from MIT, Cambridge, MA, in 1985 and the Ph.D. degree in applied math from Harvard University in 1989.

Since 1989, he has been with the University of Maryland, College Park, currently as the Ralph J. Tyser Professor of Management Science. He has also served as the Operations Research Program Director at the National Science Foundation. His research interests include simulation optimization and stochastic gradient estimation.

Dr. Fu is a Fellow of the Institute for Operations Research and the Management Sciences.

Steve Marcus (F'86) received the B.A. degree from Rice University, Houston, TX, in 1971 and the S.M. and Ph.D. degrees from MIT, Cambridge, MA, in 1972 and 1975, respectively.

From 1975 to 1991, he was with the Department of Electrical and Computer Engineering, University of Texas at Austin. In 1991, he joined the University of Maryland, College Park, as a Professor in the Electrical and Computer Engineering Department and the Institute for Systems Research. He was Director of the Institute for Systems Research from 1991 to 1996 and Chair of the Electrical and Computer Engineering Department from 2000 to 2005. Currently, his research is focused on stochastic control, estimation, hybrid systems, and optimization.

Dr. Marcus is a Fellow of the Society for Industrial and Applied Mathematics.