Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

STOCHASTIC GRADIENT ESTIMATION USING A SINGLE DESIGN POINT

Jamie R. Wieland
Bruce W. Schmeiser

School of Industrial Engineering
Purdue University
West Lafayette, IN 47907, U.S.A.

ABSTRACT

Using concepts arising in control variates, we propose estimating gradients using Monte Carlo data from a single design point. Our goal is to create a statistically efficient estimator that is easy to implement, with no analysis within the simulation oracle and no unknown algorithm parameters. We compare a simple version of the proposed method to finite differences and simultaneous perturbation, assuming first and second-order linear logic models and response surfaces. Results of the analysis indicate that the proposed gradient estimator is unbiased with variance that is inversely related to the variance of the assumed input model. Compared to the only existing single design-point method, the proposed gradient estimator is advantageous in that its variance is not dependent on the magnitude of the response surface at the design point of interest and also decreases as the simulation run length increases.

1 INTRODUCTION

1.1 Problem Statement

Given a real-valued function $g:\mathbb{R}^q \rightarrow \mathbb{R}$ and a design point $\underline{x}_0$, we address the problem of providing an estimate of the gradient of g at $\underline{x}_0$,

$$\nabla g(\underline{x})\big|_{\underline{x}=\underline{x}_0} .$$

In a single dimension (q = 1) the gradient is the derivative of g at x_0, which is defined as

$$\lim_{\delta \to 0} \frac{g(x_0+\delta)-g(x_0)}{\delta} . \qquad (1)$$

We refer to $\underline{x}$ as input-model parameters and define the performance measure of interest θ as $g(\underline{x})=\theta$. For some stochastic models closed-form solutions of θ are available and gradients can be obtained analytically. The usual stochastic simulation context, however, is that only estimates of g are observable by means of a stochastic oracle, which is assumed to be a simulation model. In this context, gradients cannot be calculated using analytic or numerical methods. Instead, gradient-estimation methods must be used.

1.2 Background & Objective

Much work has been done in developing gradient-estimation methods, which are used in stochastic optimization, stochastic root-finding, sensitivity analysis, and input-model uncertainty. These methods can be categorized based on (1) whether they use information from within the simulation oracle in deriving the estimators and (2) the total number of design points observed.

Methods relying solely on information that can be output from the simulation oracle, so no knowledge of the underlying model is used in deriving the gradient estimators, are referred to as indirect methods. Finite differences, simultaneous perturbation, response surface methods, and frequency domain are classified as indirect methods. Direct methods, such as infinitesimal perturbation analysis and likelihood ratios, use information from within the simulation oracle. (Fu 2005)

For some gradient-estimation methods, the number of design points in which users are required to observe increases with the dimensionality of the problem. Spall's (1992) simultaneous perturbation method, however, can estimate gradients in any number of dimensions using data obtained from only one or two design points. Infinitesimal perturbation analysis and likelihood ratios are other examples of methods in which the number of design points observed does not increase with the dimensionality of the problem.

We present an indirect gradient-estimation method that is based on concepts from the control-variates variance-reduction method and uses data from a single design point. We compare a simple version of the proposed method with

1-4244-0501-7/06/$20.00 ©2006 IEEE

finite differences and simultaneous perturbation. Using results derived from an assumed second-order linear model, we illustrate that the statistical performance of the proposed method appears to be competitive with that of existing methods. We then discuss future work, with an end objective of developing a gradient-estimation method that can be completely automated, requiring no user-input.

1.3 Assumptions

We make the following assumptions regarding the gradient-estimation problem:

1. The function g is continuous and differentiable everywhere.
2. The oracle used for obtaining estimates of $g(\underline{x})$ is a black-box in that information inside the oracle is not considered.
3. The oracle provides data that can be used to obtain consistent estimators for both the unknown performance measure θ and the known design point $\underline{x}_0$.

Under assumption 3, problems for which users wish to estimate the gradient of one performance measure with respect to another performance measure (instead of an input parameter) are also considered. Input-model parameters can be viewed as a subset of the performance measures in which the expected value is controllable.

We refer to a call to the oracle at a given design point $\underline{x}_0$ as "obtaining an observation" and define the simulation run length, n, to be the total number of calls made to the oracle across all design points.

1.4 Criteria for Comparing Methods

Only indirect gradient-estimation methods are considered. The criterion for evaluating methods is the sum of the generalized mean squared errors across all q dimensions,

$$(n+1)\cdot \mathrm{E}\left(\sum_{i=1}^{q}\left(\hat{\nabla} g_i\left(\underline{x}_0\right)-\nabla g_i\left(\underline{x}_0\right)\right)^2\right),$$

where $\hat{\nabla} g_i\left(\underline{x}_0\right)$ is the estimate of the ith component of $\nabla g\left(\underline{x}_0\right)$. The costs of specifying algorithm parameters and the costs of changing the values of input parameters to observe multiple design points are also considered.

1.5 Organization

The remainder of this paper is organized as follows: Section 2 discusses the finite differences and simultaneous perturbation gradient-estimation methods, and Section 3 introduces a simple version of the proposed gradient-estimation method. In Section 4 we present a model that is used for analyzing and comparing methods. Section 5 further discusses simultaneous perturbation focusing on its statistical properties when the method is evaluated conditionally. Conclusions and plans future work are contained in Section 6.

2 EXISTING METHODS

We review two variations of finite differences and two variations of simultaneous perturbation. Here we ignore frequency-domain and response-surface methods. Refer to Schruben and Cogliano (1981) for further details on the frequency-domain method.

2.1 Finite Differences

Finite differences (FD) is based on the definition of the derivative of g at x_0. The forward finite differences (FFD) estimator for single dimensions is

$$\hat{g}'(x_0)=\frac{\hat{g}(x_0+\delta)-\hat{g}(x_0)}{\delta}, \tag{2}$$

where delta δ represents the perturbation in the input parameter, which must be chosen by the user. (L'Ecuyer 1991) An alternative is the central differences estimator (FCD), which in single dimension is

$$\hat{g}'(x_0)=\frac{\hat{g}(x_0+\delta)-\hat{g}(x_0-\delta)}{2\delta}. \tag{3}$$

There is a third FD estimator, finite backward differences, with properties similar to that of FFD.

In multiple dimensions FD requires users to specify the perturbation amount δ_i for $i=1,2,\ldots,q$. The ith component of the q-dimensional FCD gradient estimator is

$$\hat{g}_i'\left(\underline{x}_0\right)=\frac{\hat{g}\left(\underline{x}_0+\underline{\delta}_i\right)-\hat{g}\left(\underline{x}_0-\underline{\delta}_i\right)}{2\delta_i}, \tag{4}$$

where $\underline{\delta}_i$ is the vector with δ_i as the ith component and zeros elsewhere.

Because FD perturbs input-model parameters one at a time, in q dimensions FFD requires users to obtain observations from q+1 design points and FCD requires $2q$ design points.

2.2 Simultaneous Perturbation

Spall (1992) modified finite differences to reduce the number of design points required in higher dimensions. There are two versions of the simultaneous perturbation method. The first allows users to estimate gradients in any number of q dimensions with two design points. The second requires only one design point.

2.2.1 Two Design-Point Version

The two design-point version of simultaneous perturbation (SP2) is similar to finite differences in that users are required to specify a vector of input-parameter perturbations $\underline{\delta}$, one perturbation for each dimension. However, instead of perturbing only one parameter at a time, all input parameters are perturbed simultaneously sampling pairs of opposing vertices at random.

The *i*th component of this q-dimensional gradient estimator is

$$\hat{g}_i'\left(\underline{x}_0\right) = \frac{\hat{g}\left(\underline{x}_0 + \underline{\delta}\underline{\Lambda}\right) - \hat{g}\left(\underline{x}_0 - \underline{\delta}\underline{\Lambda}\right)}{2\delta_i \Lambda_i},$$

where δ_i is the perturbation for the *i*th dimension, $\underline{\Lambda}$ is a vector containing the signs (± 1) of the perturbations for each dimension, and Λ_i is the sign of the perturbation in the *i*th dimension. Unlike $\underline{\delta}$, which is user-specified, $\underline{\Lambda}$ is randomly generated, requiring no user input. The sign vector is typically generated as a Bernoulli(1/2) random variable taking on values {+1,-1} instead of {0,1}. Xiong et al. (2002) have also investigated deterministic sequences for choosing $\underline{\Lambda}$.

SP2 requires users to obtain observations from design points $\underline{x}_0 \pm \underline{\delta}\underline{\Lambda}$, rather than $2q$ design points as is required by FCD.

2.2.2 Single Design-Point Version

The *i*th component of the q-dimensional gradient estimator for the single design-point version of simultaneous perturbation (SP1) is

$$\hat{g}_i'\left(\underline{x}_0\right) = \frac{\hat{g}\left(\underline{x}_0 + \underline{\delta}\underline{\Lambda}\right)}{\delta_i \Lambda_i}.$$

Like SP2, the SP1 randomly generates the signs of the perturbations $\underline{\Lambda}$ for each dimension.

SP1 requires users to obtain observations from design point $\underline{x}_0 + \underline{\delta}\underline{\Lambda}$ to estimate the gradient at $\underline{x}_0$ so in some contexts it may be important to consider the default input-parameter values.

If the default input-parameter values are assumed to be $\underline{x}_0$, then this method would not be classified as a single design-point method because it would require the user to run the simulation with input-parameter setting $\underline{x}_0 + \underline{\delta}\underline{\Lambda}$. Considering the default input-parameter settings is usually irrelevant in simulation experiments. For physical experiments, however, assumptions about the default values may be important.

3 PROPOSED METHOD

We obtain observations from the oracle at a single design point $\underline{x}_0$. Recall that in Section 1.3 we assume that in addition to estimating the performance measure θ_0, users can also obtain estimates of the input parameters $\underline{x}_0$.

Let $\hat{\theta}_{0j}$ be the *j*th batch estimate of θ_0 based on a batch size of m and $\hat{\underline{X}}_{0j}$ be the *j*th batch estimate of $\underline{x}_0$ based on a batch size of m for $j = 1,2,\ldots,k$, where $k \cdot m$ is equal to the simulation run length n. Refer to Law and Kelton (2000) for details on batch statistics.

First consider a one-dimensional problem. In this case the proposed gradient estimator is

$$\hat{g}'\left(\underline{x}_0\right) = \frac{\widehat{\mathrm{Cov}}\left(\hat{\theta}_0, \hat{X}_0\right)}{\widehat{\mathrm{Var}}\left(\hat{X}_0\right)} = \frac{\sum_{j=1}^{k}\left(\hat{X}_{0j} - \bar{X}\right)\left(\hat{\theta}_{0j} - \bar{\theta}\right)}{\sum_{j=1}^{k}\left(\hat{X}_{0j} - \bar{X}\right)^2}, \quad (5)$$

where

$$\bar{X} = \frac{1}{k}\sum_{j=1}^{k}\hat{X}_{0j}$$

and

$$\bar{\theta} = \frac{1}{k}\sum_{j=1}^{k}\hat{\theta}_{0j}.$$

The basis for using (5) as a gradient estimator is that, when the conditional relationship between $\hat{\theta}_{0j}$ and $\hat{X}_{0j}$ is linear, then

$$\mathrm{E}\left[\hat{\theta}_{0j} \mid \hat{X}_{0j}\right] = \gamma_0 + \gamma_1 \hat{X}_{0j}. \quad (6)$$

Assuming that $\hat{\theta}_{0j}$ and $\hat{X}_{0j}$ are unbiased estimators of θ_0 and x_0, the unconditional expectation of $\hat{\theta}_{0j}$ is then

$$\mathrm{E}_{\hat{X}_0}\left[\mathrm{E}\left[\hat{\theta}_{0j} \mid \hat{X}_{0j}\right]\right] = \gamma_0 + \gamma_1 x. \quad (7)$$

Differentiating (7) with respect to x yields γ_1, which is estimated via least squares.

Lavenberg and Welch (1981) show that the conditional linear model in (6) is correct when $\hat{\theta}_{0j}$ and $\hat{X}_{0j}$ have a bivariate normal distribution. For most estimators, when the raw output data is batched into k batches of length m, as $m \to \infty$ $\left(\hat{\theta}_{0j}, \hat{X}_{0j}\right)$ have a bivariate normal distribution with means $\left(\hat{\theta}_{0j}, \hat{X}_{0j}\right)$, variances

$$\left(\frac{\sigma_\theta^2}{m}, \frac{\sigma_X^2}{m}\right),$$

and correlation ρ.

3.1.1 Connection to Control Variates

The proposed method for estimating $g'(x_0)$ is similar to the control-variates variance-reduction method. The objective of control variates, however, is to reduce the variance of the estimate of θ_0, not to estimate $g'(x_0)$. Linear control variates assume that

$$\mathrm{E}\left[\hat{\theta}_{0j} \mid \hat{X}_{0j}\right] = \theta_0 + \alpha_1\left(\hat{X}_{0j} - x_0\right),$$

which is a centered linear model. (Law and Kelton 2000)

Assuming a centered model implies that the expected value of the input parameter is known. In this model α_1 is the optimal control-variate weight, which is usually estimated with

$$\hat{\alpha}_1 = \frac{\widehat{\mathrm{Cov}}\left(\hat{\theta}_0, \hat{X}_0\right)}{\widehat{\mathrm{Var}}\left(\hat{X}_0\right)}.$$

So the estimator for α_1 is equivalent to our gradient estimator, $\hat{\gamma}_1$. This can also be shown by converting the centered model parameters (α_0, α_1) to the uncentered model parameters (γ_0, γ_1)

$$\begin{aligned} \alpha_0 &= \gamma_0 + \alpha_1 x \\ \alpha_1 &= \gamma_1 \end{aligned}.$$

Refer to Tamhane and Dunlop (2000) for details on conversion between centered and uncentered linear models.

We do not assume that the expected values of the input parameters are known, which is why we do not use a centered model. One benefit of the proposed method is that it can be used to estimate gradients with respect to variables with unknown expected values. Consider the case, for example, where one wishes to estimate the gradient of one performance measure with respect to another performance measure, rather than an input-model parameter.

3.2 Higher-Dimensional Problems

The proposed method can be extended to estimate gradients in multiple dimensions. For such problems we fit individual linear models

$$\mathrm{E}\left[\hat{\theta}_{0j} \mid \hat{X}_{ij}\right] = \gamma_{i,0} + \gamma_{i,1}\hat{X}_{ij},$$

for $i = (1, 2, \ldots, q)$ and estimate $\gamma_{i,1}$, the *i*th component of the *q*-dimensional gradient, with

$$\hat{g}_i'\left(\underline{x}_0\right) = \frac{\widehat{\mathrm{Cov}}\left(\hat{\theta}_0, \hat{X}_i\right)}{\widehat{\mathrm{Var}}\left(\hat{X}_i\right)} = \frac{\sum_{j=1}^{k}\left(\hat{X}_{ij} - \overline{X}\right)\left(\hat{\theta}_{0j} - \overline{\theta}\right)}{\sum_{j=1}^{k}\left(\hat{X}_{ij} - \overline{X}\right)^2},$$

where $\hat{X}_{ij}$ is the *j*th batch estimate of the *i*th input parameter based on a batch size of *m*.

An alternative to fitting individual linear models would be to fit a combined linear model,

$$\mathrm{E}\left[\hat{\theta}_{0j} \mid \hat{X}_{1j}, \hat{X}_{2j}, \ldots, \hat{X}_{qj}\right] = \gamma_0 + \sum_{i=1}^{q}\gamma_i \hat{X}_{ij},$$

and estimate γ_i for $i = 1, 2, \ldots, q$. This an area of future research discussed in Section 6.

3.3 Advantages

The proposed method is an indirect method so it is generally applicable across simulation models and can be used to estimate q-dimensional gradients using data from a single design point.

The only method parameter that users are required to specify is the number of batches k, which can be chosen and adjusted after the data have been collected. If users change decisions about how to batch the data, they are not required to run the experiment again. This is not the case with FD or SP where users must specify the input parameter perturbations $\underline{\delta}$ in advance. Choosing the number of batches affects only the analysis of the experiment, whereas choosing the input perturbations changes the experimental design.

3.4 Drawbacks and Considerations

The proposed method cannot be used for estimating gradients with respect to quantities that have zero or extremely little variance at a given design point. This situation commonly arises when users want to estimate gradients with respect to logic-model parameters rather than input-model parameters. See Figure 1.

$$G \rightarrow U \rightarrow \boxed{\text{Input Model}} \rightarrow X \rightarrow \boxed{\text{Logic Model}} \rightarrow Y \rightarrow \left(\hat{\theta}_0, \hat{X}_0\right)$$

Figure 1: Process of Obtaining Estimates Through a Simulation Oracle

Figure 1 illustrates the process of obtaining estimates of θ_0 and x_0 through a simulation oracle. This process begins with a random number generator, *G*, which is used to obtain Uniform(0,1) random numbers, *U*. Input models are then used to generate input observations *X*. The input observations are then converted to output observations, *Y*, via a logic model. Finally, output observations are used to obtain the desired estimates $\left(\hat{\theta}_0, \hat{X}_0\right)$.

Logic model parameters are not generated and hence have no variance at a given design point x_0. Consider the case, for example, where users wish to obtain the gradient of a performance measure with respect to the buffer size (assuming that the buffer size is continuous), which is a logic-model parameter. At a given design point, the number of buffer positions is deterministic. In such a case, the proposed gradient estimator introduced in (5) is undefined.

4 MODEL AND ANALYSIS

To compare the proposed method with FD and SP we construct a stochastic model. We have one input-parameter, X, with input model

$$N\left(x_0, \sigma_X^2\right). \quad (8)$$

Let X_h denote the *h*th observation of X for $h = 1,2,\ldots,n$. We assume a second-order linear logic model with additive error ε_h,

$$\theta_h = \beta_0 + \beta_1 X_h + \beta_2 X_h^2 + \varepsilon_h, \quad (9)$$

where $\varepsilon_h \sim N\left(0, \sigma_\varepsilon^2\right)$. We also assume that X_h's and ε_h's are independent and identically distributed and that X_h and ε_h are independent of each other for $h = 1,2,\ldots,n$. This implies that the θ_h's are also independent and identically distributed.

Assuming input model (8) and logic model (9), the response surface, defined as

$$g\left(x_0\right) = \mathrm{E}\left(\hat{\theta}_h\right),$$

where $\hat{\theta}_h$ estimates θ, is implied to be

$$g\left(x_0\right) = \beta_0 + \beta_1 x_0 + \beta_2\left(x_0{}^2 + \sigma_X^2\right).$$

Recall that the objective is to estimate

$$\left.\frac{dg}{dx}\right|_{x = x_0},$$

which is

$$\beta_1 + 2\beta_2 x_0.$$

4.1 First-Order Model Results

We analyze the bias and the variance of the gradient estimators for the proposed method, FFD, FCD, SP2, and SP1 for the assumed stochastic model where $\beta_2 = 0$. Table 1 contains results for this analysis.

Table 1: First-Order Model Results

Method	Bias	Variance
Proposed	0	$\dfrac{\sigma_\varepsilon^2}{(k-3)\sigma_X^2}$
FFD	0	$\dfrac{\beta_1{}^2\sigma_X^2 + \sigma_\varepsilon^2}{n\delta^2}$
FCD	0	$\dfrac{\beta_1{}^2\sigma_X^2 + \sigma_\varepsilon^2}{n\delta^2}$
SP2	0	$\dfrac{\beta_1{}^2\sigma_X^2 + \sigma_\varepsilon^2}{n\delta^2}$
SP1	0	$\left(\dfrac{g(x_0)}{\delta}\right)^2 + \dfrac{\beta_1{}^2\sigma_X^2 + \sigma_\varepsilon^2}{n\delta^2}$

All of the gradient estimation methods yield unbiased estimates for the assumed first-order model so we consider only variance.

The variance term for the proposed method is inversely related to the number of batches *k* so, assuming that the number of batches used increases proportionally with the simulation run length *n*, variance decreases as *n* increases. The variance of the proposed estimator differs from that of the other estimators in that it does not depend on β_1 and is inversely related to σ_X^2.

The variance is the same for the FFD, FCD, and SP2 gradient estimators. Variance for the SP1 estimator not only depends on β_1, but also on $g(x_0)$. Furthermore, it is not decreasing as *n* increases.

For a fixed *n*, the variance of the proposed method is minimized by choosing *k* = *n*. For all other methods variance is minimized by choosing δ to be as large as possible. Thus, δ can always be chosen such that the variance of the estimators for the other methods is lower than that of the proposed method.

4.2 Variance Reduction

The results obtained in Table 1 are derived assuming that observations from the input model are independent and identically distributed. Common random numbers could be used for FFD, FCD, and SP2, which would result in these methods having zero variance. Antithetic variates could be used for the proposed method, which would result in the proposed estimator having zero variance.

4.3 Second-Order Model Results

We now analyze the bias and the variance of the gradient estimators for the proposed method, FFD, FCD, SP2, and SP1 for the assumed stochastic model for an arbitrary β_2. Table 2 contains results for this analysis.

Table 2: Second-Order Model Results

Method	Bias	Variance
Proposed	0	$\frac{1}{(k-3)\sigma_X^2}\left(\frac{\beta_2^2(m-1)2\sigma_X^4}{m}+\sigma_\varepsilon^2\right)$
FFD	$\beta_2\delta$	$\frac{2\left(\sigma_\theta^2(x_0+\delta)+\sigma_\theta^2(x_0)\right)}{\delta^2 n}$
FCD	0	$\frac{\sigma_\theta^2(x_0+\delta)+\sigma_\theta^2(x_0-\delta)}{2\delta^2 n}$
SP2	0	$\frac{\sigma_\theta^2(x+\delta)+\sigma_\theta^2(x-\delta)}{2\delta^2 n}$
SP1	0	$\frac{\left(g(x_0+\delta)\right)^2+\left(g(x_0-\delta)\right)^2}{2\delta^2}$ $-\left(\beta_1^2+4\beta_2^2x_0^2+4\beta_1\beta_2x_0\right)$ $+\frac{\sigma_\theta^2(x_0+\delta)+\sigma_\theta^2(x_0-\delta)}{2\delta^2 n}$

The variance of the FFD, FCD, SP2, and SP1 estimators is a function of $\sigma_\theta^2(x)$, which is

$$4\beta_2 x\left(\beta_2\,\sigma_X^2 x+\beta_1\sigma_X^2\right)+2\beta_2{}^2\sigma_X^4+\beta_1{}^2\sigma_X^2+\sigma_\varepsilon^2\,.$$

All of the gradient estimation methods except FFD yield unbiased gradient estimates for the second-order model. Results for the variance of the estimators are similar to that of the first-order model. The variance for the estimator of proposed method, however, is dependent on β_2, whereas in the first-order model variance was not dependent on the response-surface coefficients.

Because bias for FFD is directly related to δ, there is a tradeoff between choosing delta small enough to reduce bias and large enough to reduce variance.

5 CONDITIONAL SIMULTANEOUS PERTURBATION

Both SP2 and SP1 require users to sample the signs of the input parameter perturbations $\underline{\Lambda}$ randomly. Statistical performance of the SP2 and SP1 estimators is then evaluated across multiple realizations of $\underline{\Lambda}$. (Spall 1992, 1997)

Evaluating the performance in this manner, both the SP2 and SP1 estimators are unbiased, as shown in Tables 1 and 2. Such an evaluation is useful when these methods are used within a stochastic approximation algorithm, which was the context in which these methods were originally developed. Stochastic approximation is an iterative process where users make multiple replications at each point. This type of analysis, however, may not be as useful for the stand-alone gradient estimation problem.

Evaluating the performance across multiple realizations of $\underline{\Lambda}$ implies that users must use more than one or two design points to obtain unbiased estimates. Both SP2 and SP1 will observe 2^q design points as the number of realizations increases, where as FCD only observes $2q$ design points. Therefore, in the stand-alone gradient estimation problem, it needs to be determined whether SP2 and SP1 are truly two design-point and one design-point methods, respectively, or whether the number of design points required is determined by the number of realizations users perform.

Using the model presented in Section 4, we analyze the performance of both SP2 and SP1 conditional on a given set of signs $\underline{\Lambda}=\underline{\lambda}$, requiring that users observe only two design points for conditional SP2 (CSP2) and one design point for conditional SP1 (CSP1).

Table 3 lists the bias and the variance of the CSP2 and CSP1 estimators for a first-order model ($\beta_2 = 0$).

Table 3: CSP2 and CSP1 Results

Method	Bias	Variance
CSP2	0	$\dfrac{\beta_1^2\sigma_X^2+\sigma_\varepsilon^2}{n\delta^2}$
CSP1	$\dfrac{\beta_0+\beta_1 x_0}{\lambda\delta}$	$\dfrac{\left(\beta_1^2\sigma_X^2+\sigma_\varepsilon^2\right)}{n\delta^2}$

The results listed in Table 3 show that the variance of CSP2 is equivalent to that of SP2. CSP2 is unbiased for one-dimensional problem because the estimators are equivalent for both $\Lambda=+1$ and $\Lambda=-1$. For higher-dimensional, first-order problems, CSP2 is biased. Extending the model presented in Section 4 such that the number of inputs is $q > 1$, the logic model would then be

$$\theta_h=\beta_0+\sum_{i=1}^{q}\beta_{1,i}X_{h,i}+\varepsilon_h ,$$

for $i=1,2,\ldots,q$, and the response surface is

$$g\left(x_0\right)=\beta_0+\sum_{i=1}^{q}\beta_{1,i}x_{0i} ,$$

which is a linear first-order model with gradient

$$\left(\beta_{1,1},\beta_{1,2},\ldots,\beta_{1,q}\right).$$

The bias for the *i*th component of the CSP2 estimator is

$$\sum_{\substack{l=1,\\ l\neq i}}^{q}\frac{\beta_{1,l}\lambda_l\delta_l}{\lambda_i\delta_i} .$$

Recall from the results in Table 1 that both the FFD and FCD estimators are unbiased for linear first-order response surfaces.

Unlike SP1, CSP1 is biased even for a one-dimensional problem. The variance of CSP1, however, is lower than it is for SP1, but the mean squared error, which is the sum of squared bias and the variance, is equivalent for SP1 and CSP1.

6 CONCLUSIONS AND FUTURE WORK

We compared a simplistic version of the proposed method to existing methods assuming first and second-order linear logic models and respective first and second-order implied response surfaces. Results of the analysis indicate that the proposed gradient estimator is unbiased with variance inversely related to σ_X^2 .

Compared to the only existing single design-point method, SP1, the proposed gradient estimator is advantageous in that its variance is not dependent on the magnitude of the response surface at the design point of interest $g\left(x_0\right)$ and it is decreasing as the simulation run length *n* increases.

We also analyzed the statistical performance of SP conditional on the sign vector $\underline{\Lambda}$ for a first-order linear model. Results indicate CSP1 is biased and that CSP2 is biased for problems with dimensionality greater than one.

Future work for extending the proposed method includes:

1. investigating conditions under which it is better to a combined model, rather than individual models to estimate the regression coefficients
2. developing an algorithm to automate the proposed method without any user-input
3. developing a sequential method for choosing the magnitude of the perturbations in the input parameters such that the proposed method can be used in cases where the input-model parameters have zero variance.

ACKNOWLEDGMENTS

The authors would like to thank Raghu Pasupathy, Christopher Mulligan, and Mica Hutchison for their contributions in editing this paper.

REFERENCES

Fu, M. 2005. Stochastic Gradient Estimation. Pre-print version of Chapter 19 in Handbook on Operations Research and Management Science: Simulation, S.G. Henderson and B.L. Nelson, editors, Elsevier, forthcoming.

Lavenberg, S.S., and P. D. Welch. 1981. A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. Management Science 27: 322–335.

Law, A. M., and W. D. Kelton. 2000. Simulation Modeling and Analysis. Third edition. New York, NY: McGraw-Hill

L'Ecuyer, P. 1991. An overview of derivative estimation. Proceedings of the 1991 Winter Simulation Conference, eds. Nelson, B. L., Kelton, D, K., and Clark, G. M., 207-217.

Schruben, L.W., and V.J. Cogliano. 1981. Simulation sensitivity analysis: a frequency domain approach. Proceedings of the 1981 Winter Simulation Conference, eds. Delfosse, C.M., Shub, C.M., and Ören, T.I., 455-459.

Spall, J.C. 1992. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Transactions on Automatic Control 37: 332-341.

Spall, J.C. 1997. A one-measurement form of simultaneous perturbation stochastic approximation. Automatica 33: 109-112.

Tamhane, A.C. and D.D. Dunlop. 2000. Statistics and Data Analysis. Upper Saddle River, NJ: Prentice-Hall, Inc.

X. Xiong, I-J. Wang, and M.C. Fu. 2002. Randomized-direction stochastic approximation algorithms using deterministic sequences. Proceedings of the 2002 Winter Simulation Conference, eds. Snowdon, J.L., Charnes, J.M. Yücesan, E., and Chen, C-H., 285-291.

AUTHOR BIOGRAPHIES

JAMIE R. WIELAND is a Ph.D. student in the School of Industrial Engineering at Purdue University. She received a B.S. in Industrial Engineering & Management Sciences from Northwestern University in 2001 and a M.S. in Industrial Engineering from Purdue University in 2003. Her research interests are in stochastic operations research and economics. Her email address is <jwieland@purdue.edu>.

BRUCE W. SCHMEISER His research interests center on developing methods for better simulation experiments. He is a Fellow of INFORMS, is a Fellow of IIE, and has been active within the Winter Simulation Conference for many years, including being the 1983 Program Chair and chairing the Board of Directors from 1988–1990. His email address is <bruce@purdue.edu>.