Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

SOME ISSUES IN MULTIVARIATE STOCHASTIC ROOT FINDING

Raghu Pasupathy
Bruce W. Schmeiser

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

ABSTRACT

The stochastic root finding problem (SRFP) involves finding points in a region where a function attains a prespecified target value, using only a consistent estimator of the function. Due to the properties that the SRFP contexts entail, the development of good solutions to SRFPs has proven difficult, at least in the multi-dimensional setting. This paper discusses certain key issues, insights and complexities for SRFPs. Some of these are important in that they point to phenomena that contribute to the difficulties that arise in the development of efficient algorithms for SRFPs. Others are simply observations, sometimes obvious, but important for providing useful insight into algorithm development.

1 INTRODUCTION

The deterministic root finding problem is a well-researched problem in mathematics. It involves a known function $g : \Re^q \to \Re^q$, a known target $\gamma \in \Re^q$, and an unknown root $x^* \in \Re^q$. The problem is to determine the unique root $x = x^*$, where g attains the target value γ. The target γ usually represents the desired level of a system's performance to be obtained by controlling the vector of inputs x. Several famous and efficient numerical methods such as bisection search, Newton's method and regula falsi (Conte and De Boor, 1980) have been devised for the single-dimensional case ($q = 1$).

A generalization of the deterministic root finding problem is the comparatively less-known stochastic root finding problem (SRFP) (Chen, 1994; Chen and Schmeiser, 1994a, 2001). Unlike the deterministic root finding problem where the function g is known to the researcher, in the SRFP a consistent estimator $\overline{Y}_m(x)$ of $g(x)$ is all that is available. This implies that the researcher may have to sample a large number of times to obtain an accurate value $g(x)$ for any x in the domain of g. SRFPs arise often in the control of stochastic systems where a performance function (g) of the system is known only through an oracle (e.g a simulator) or has a complex analytical form.

Formally, the SRFP is stated as follows (Chen and Schmeiser, 1994b, 2001):

Given: (a) a constant vector $\gamma \in \Re^q$ and (b) a (computer) procedure for generating, for any $x \in \Re^q$, a q-dimensional consistent estimate $\overline{Y}_m(x)$ of $g(x)$.

Find: the unique root x^* satisfying $g(x) = \gamma$ using only the procedure.

Five useful criteria by which algorithms for SRFPs can be evaluated are (1) numerical stability, (2) robustness, (3) convergence, (4) computational efficiency and (5) the ability to report solution accuracy (Chen, 1994). *Numerical stability* is a qualitative measure of the actual performance of the algorithm when implemented on a digital computer. This is by contrast to theoretical performance measures that discount issues such as computer arithmetic and floating-point representation. *Robustness* is another qualitative measure signifying how sensitive the algorithm is to the starting values of the different parameters in the algorithm. A good algorithm should not be overly sensitive to parameter starting values so that the need for user tuning of parameters becomes minimal. It is often easy to develop heuristics for setting algorithmic parameters so that a small class of problems are solved efficiently. The objective however should be the development of a 'black-box' algorithm that will perform well across a wide range of g functions without having to specially set algorithmic parameters. *Convergence* refers to asymptotic convergence in some probability measure. Although convergence is a measure that may not be indicative of the finite-time performance of the algorithm, an algorithm that guarantees convergence is more desirable than one that does not. *Computational efficiency* is the tradeoff function that will be used to relate solution quality with the amount of computing effort involved. It can be measured as E(work $\times$ squared error) with smaller values indicating higher efficiencies. This measure is useful in that it measures both the finite and asymptotic performance of the algorithm in terms of the solution quality and the effort expended (in terms of computer time) in obtaining the solution. A related issue is the estimation of the solution accuracy during each iteration. On account of randomness in the estimator of g, the solution at each stage is itself random. If the current solution is $\overline{X}_i$ and the true root is x^*, a useful measure of accuracy is the mean squared error $\mathrm{E}(\overline{X}_i - x^*)^2$. Of the two components, variance $\mathrm{Var}(\overline{X}_i)$ and squared bias $\mathrm{E}^2(\overline{X}_i - x^*)$ that make up the mean squared error, methods for estimating $\mathrm{Var}(\overline{X}_i)$ can be designed. The bias, however, is often inestimable. Researchers thus routinely use $\mathrm{Var}(\overline{X}_i)$ as the sole measure of accuracy of the current solution. If the bias is substantial (often due to the non-linear nature of g) and persists over an extended number of iterations, variance may be a poor measure of accuracy.

SRFPs form an important class of problems that has direct application in a wide variety of disciplines including statistics, operations research, transportation systems engineering, telecommunication systems engineering and aerospace engineering. The context is usually performance control or optimization in a stochastic setting (Spall, 1999). To this extent, efficient algorithms, as measured by the criteria stated, are important, but they have been difficult to design in the multi-dimensional context.

2 ISSUES

The objective of this paper is to discuss certain key issues in the context of SRFPs. Some of these issues are important in that they point to phenomena that contribute to the difficulties that arise in the development of efficient algorithms for SRFPs. Other issues discussed are simply observations, sometimes obvious, but important nevertheless because they may provide useful insight into algorithm development. The issues are broadly classified as (i) Problem Interpretation, (ii) Nature of g, (iii) Dimensionality and (iv) Stochasticity.

2.1 Problem Interpretation

One useful way of looking at the general SRFP equation $g(x) = \gamma, g : \Re^q \to \Re^q, \gamma \in \Re^q$ is as a nonlinear system of equations:

$$\begin{aligned} g_1(x_1, x_2, \ldots, x_q) &= \gamma_1 \\ g_2(x_1, x_2, \ldots, x_q) &= \gamma_2 \\ &\vdots \\ g_q(x_1, x_2, \ldots, x_q) &= \gamma_q \end{aligned} \tag{1}$$

where $g_i : \Re^q \to \Re, \gamma_i \in \Re$. Let E_i represent the set of solutions x that solve the ith equation $g_i(x) = \gamma_i, x \in \Re^q, i \in \{1, 2, \ldots, q\}$. Clearly, $E_i \subset \Re^q$. Then the set of solutions to the non-linear system is

$$E^* = \bigcap_{i=1}^{q} E_i.$$

The sets E_i present an interesting contrast between linear and nonlinear systems. If the g_i are linear, then each E_i is either empty, has exactly one element or has infinite number of elements (in fact, E_i can have one element only if $q = 1$). From linear algebra, it is also true that their intersection E^* possesses the same property — E^* is either empty, has exactly one element or has infinite elements.

By contrast, when the g_i are nonlinear, the constituent solution sets E_i can each have any number of elements. Likewise, their intersection E^* can also have any number of elements.

SRFPs require the solution of the non-linear system (1) but the functions g_i in (1) are not observable. This is why many current methods (Chen and Schmeiser, 2001; Spall, 1999; Andradóttir, 1990) for SRFPs repeatedly construct and solve an approximate non-linear system during each iteration:

$$\begin{aligned} Y_{1,m}(x_1, x_2, \ldots, x_q) &= \gamma_1 \\ Y_{2,m}(x_1, x_2, \ldots, x_q) &= \gamma_2 \\ &\vdots \\ Y_{q,m}(x_1, x_2, \ldots, x_q) &= \gamma_q \end{aligned} \tag{2}$$

where each $Y_{i,m}$ is a random variable.

It is instructive to think of (2) as a deformation of (1). The extent of this deformation is explicitly dependent on the sample size m. Larger sample sizes deform (1) lesser in the sense that for each x

$$\lim_{m \to \infty} Y_m(x) = g(x), a.s.$$

Therefore, under weak conditions,

$$\lim_{m \to \infty} Y_m^{-1}(\gamma) = g^{-1}(\gamma), a.s.$$

It is in this sense that SRFP solution methods that progressively solve (2) using increasing sample sizes (e.g. retrospective approximation methods) can be thought of as *homotopy* methods (Allgower and Georg, 1990; Todd, 1976) for stochastic root finding.

2.2 Nature of g

In all of algorithm design there is a trade-off to be made between algorithm efficiency and applicability. Applicability is decided by the nature of assumptions made on the constituent function(s) defining the problem class. Stronger assumptions afford the development of better algorithms but often decrease applicability. Likewise, weaker assumptions, while increasing the size of the applicable problem class inhibit the development of efficient algorithms.

In the context of SRFPs, the assumptions about g predominantly decide the size of the applicable problem class. Some form of continuity assumption (ranging from continuity at the root to continuous-everywhere) on g seems necessary to devise algorithms that are applicable in practice. This is because lack of any guarantee about continuity means functional estimates at different points in the domain of g leave no clues about the location of the root.

Even continuity over the entire domain is not sufficient to guarantee that any specific algorithm will have good finite-time performance on every instance in its problem class. In other words, given any algorithm, it is easy to construct continuous-everywhere g functions for which the algorithm converges arbitrarily slowly.

Several works on the single-dimensional SRFP, including the original work by Robbins and Munro (1951), assume monotonicity of g. Monotonicity, an assumption that appears strong, is actually quite reasonable for many settings in which SRFPs arise. For example, g can be argued to be monotone in many single-dimensional SRFP motivating contexts (Chen and Schmeiser, 1994a, 1994b). Monotonicty, in combination with some sort of continuity assumption, is useful because it affords the design of efficient search algorithms that make intelligent movement decisions based on functional estimates at different points.

Differentiability of g, on the other hand, seems like a debatable assumption and depends on the specific context in which the SRFP arises. If g is differentiable, it is easy to improve the asymptotic efficieny of any convergent search algorithm. This is often done by switching the search scheme in the algorithm to a technique such as the secant method as soon as searching evolves to a region close to the root. ‘Close to the root’ is often difficult to quantify, however, in any rigorous and verifiable fashion.

2.3 Dimensionality

Developing algorithms for general SRFPs ($q \geq 1$) seems to present some additional challenges as compared to the single-dimensional context. This is primarily because searching for solutions in a single dimension ($\Re$) may proceed only in one of two possible directions (right or left). By contrast, there are infinitely many potential directions when searching in $\Re^q, q \geq 2$. This issue lies at the heart of many convergence and algorithm-efficiency related complications that arise in the multi-dimensional context.

For example, interpolation is a concept that is routinely used in single-dimensional search algorithms. Interpolation in a single dimension is usually preceeded by the identification of an interval (called a *bounding* region) whose end-points straddle the target γ. This concept of bounding the target using functional values is useful in the following sense: if $g(x_1)$ and $g(x_2)(g : \Re \to \Re, x_1, x_2 \in \Re)$ bound the target $\gamma \in \Re$, i.e if

$$(g(x_1) - \gamma)(g(x_2) - \gamma) < 0,$$

then assuming g is continuous, x_1 and x_2 bound the root x^* (from the intermediate value theorem). An interval (x_1, x_2) containing the root is thus identified in the process.

For the multi-dimensional case, however, the form of the bounding region is unclear. It could be any closed region such as a cube, a simplex, a sphere or any irregular closed object that contains the target.

More importantly, bounding the target is not as helpful in the multi-dimensional case as it is in the single-dimensional case. As Figure 1 suggests, bounding the target using an object such as a simplex does not automatically imply that the root x^* has been bounded. This is because functions have much more *directional freedom* in $\Re^q$ as compared to $\Re$.

Another example that illustrates the lack of seamless transition from single to multiple dimensions arises in the context of the assumptions on g. It was argued in Section 2.2 that in a single dimension g can be assumed to be monotone. While being reasonable, the assumption helps immensely in the development of efficient search algorithms. It is unclear, however, as to what the corresponding assumption would be in a multi-dimensional setting.

2.4 Stochasticity

In an SRFP (unlike deterministic root-finding problems), accurate functional estimates of $g(x)$ are not readily available (without computational cost) to the algorithm. As a result, an algorithm that attempts to search for the true root by obtaining clues from functional estimates must explicitly account for randomness. Thus, an intelligent algorithm would base its search direction decisions on a probability model that includes the nature of the estimator of g.

Another feature that warrants careful consideration in the design of algorithms for the solutions to SRFPs is the choice of sample size across iterations. The sample-sizing issue is central to the classic trade-off between computational efficiency and solution accuracy. The sequence of sample sizes across iterations, $\{m_i\}$, should be such that $\lim_{i\to\infty} m_i = \infty$ (otherwise, the sampling error associated with a finite sample size will restrict asymptotic solution quality). Furthermore, it is also natural that the sequence $\{m_i\}$ should be chosen to be increasing so that the variances of the solutions obtained reduce across iterations. What is not obvious, however, but critical in terms of computing effort, is the actual sequence of sample sizes $\{m_i\}$ to be used.

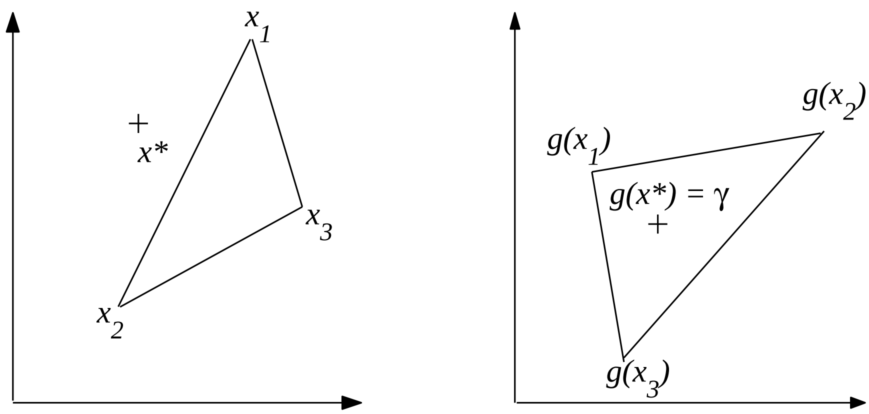

Figure 1: Two Dimensional Bounding Using Simplices (Cross Hairs Represent the Unknown Root x^* and Known Target γ)

3 SUMMARY

SRFPs are an important class of problems that has found application in a wide range of areas. Recently, several algorithms have been developed and solve SRFPs efficiently in the single-dimensional context. The problem in multiple dimensions, however, seems to pose some challenges. This is in part because there seem to be no obvious multi-dimensional analogues to some assumptions about g (such as monotonicity) that single-dimensional algorithms routinely make. In addition, there seems to be a lack of analogues to useful single-dimensional search concepts such as bounding and bisection. Further investigation in these areas may be useful in the design of numerically stable, robust and computationally efficient algorithms for general SRFPs.

REFERENCES

Allgower, E. L. and K. Georg. 1990. *Numerical Continuation Methods. An Introduction*, Springer-Verlag.

Andradóttir, S. 1991. A projected stochastic approximation algorithm. In *Proceedings of the 1991 Winter Simulation Conference*, ed. B.L. Nelson, W. D. Kelton, and G. M. Clark, 945–957, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Chen, H. 1994. *Stochastic Root Finding in System Design*, Ph.D. Dissertation, School of Industrial Engineering, Purdue University, West Lafayette, Indiana, USA.

Chen, H. and B.W. Schmeiser. 1994a. Stochastic root finding: problem definition, examples, and algorithms. In *Proceedings of the Third Industrial Engineering Research Conference*, ed. L. Burke and J. Jackman, 605–610, Institute of Industrial Engineers, Norcross, Georgia.

Chen, H. and B.W. Schmeiser. 1994b. Retrospective approximation algorithms for stochastic root finding. In *Proceedings of the 1994 Winter Simulation Conference*, ed. J.D. Tew, S. Manivannan, D.A. Sadowski, and A.F. Seila, 255–261, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Chen, H. and B.W. Schmeiser. 2001. Stochastic root finding via retrospective approximation. *IIE Transactions* **33**: 259–275.

Conte, S.D. and C. de Boor. 1980. *Elementary Numerical Analysis: An Algorithmic Approach*, McGraw-Hill.

Robbins, H. and S. Monro. 1951. A stochastic approximation method. *Annals of Mathematical Statistics* **22**: 400–407.

Spall, J. C. 1999. Stochastic optimization and the simultaneous perturbation method. In *Proceedings of the 1999 Winter Simulation Conference*, ed. P.A. Farrington, H.B. Nembhard, D.T. Sturrock, and G.W. Evans, 101–109, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Todd, M. J. 1976. *The computation of fixed points and applications*, Springer-Verlag.

AUTHOR BIOGRAPHIES

RAGHU PASUPATHY is a Ph.D. student in the School of Industrial Engineering at Purdue University. His dissertation has focused on designing efficient algorithms for the solution of general stochastic root-finding problems. His broad research interests are in the area of stochastic operations research. His e-mail address is `<pasupath@purdue.edu>`.

BRUCE W. SCHMEISER is a professor in the School of Industrial Engineering at Purdue University. His interests lie in applied operations research, with emphasis in stochastic models, especially the probabilistic and statistical aspects of stochastic simulation. He is an active participant the in Winter Simulation Conference, including being Program Chair in 1983 and chairing the Board of Directors from 1988-1990. His e-mail address is `<bruce@purdue.edu>`.