A Hybrid Algorithm for Global Optimization Problems[*]

Miguel Argáez, Leticia Velázquez, Carlos Quintero[†]

Department of Mathematical Sciences and Computational Science Program, The University of Texas at El Paso, El Paso, TX 79968, USA

leti@utep.edu

Hector Klie, and Mary Wheeler[‡]

Center for Subsurface Modeling, The University of Texas at Austin, Austin, TX 78712, USA.

Abstract

We propose a hybrid algorithm for solving global optimization problems that is based on the coupling of the Simultaneous Perturbation Stochastic Approximation (SPSA) and Newton-Krylov Interior-Point (NKIP) methods via a surrogate model. There exist verified algorithms for finding approximate global solutions, but our technique will further guarantee that such solutions satisfy physical bounds of the problem. First, the SPSA algorithm conjectures regions where a global solution may exist. Next, some data points from the regions are selected to generate a continuously differentiable surrogate model that approximates the original function. Finally, the NKIP algorithm is applied to the surrogate model subject to bound constraints for obtaining a feasible approximate global solution. We present some numerical results on a set of five small problems and two medium to large-scale applications from reservoir simulations.

1 Introduction

Finding a global optimal solution is a challenging task in many applications due to the fact that data and models are usually nonlinear and subject to different sources of error. As an example, major state environmental problems include obtaining subsurface

properties from pressure head and saturation observations in contaminant remediation sites, estimating temporal and spatial coefficients for shallow water models from observed bathymetry data, and fish behavior prediction. Many of these problems are computationally demanding, and have motivated the necessity for developing novel optimization approaches. Moreover, in many cases the number of these parameters may be significantly large since they generally depend on a prescribed level of spatial and temporal resolution. Most successful approaches in global optimization use some stochastic or heuristic strategies that do not depend on derivative information, e.g., genetic algorithms, tabu searches, neural networks, and simulated annealing, among others [1]. Despite inherent costs, their success resides in making a broad and systematic exploration of the search domain, although this may vary with the nature of the problem. Therefore, it is useful to obtain several local optimal solutions and compare them before selecting a global one. However, the performance of these algorithms depends significantly on the choice of parameters and this task tends to be less obvious on large scale since the number of function evaluations grows dramatically with the number of parameters. On the other hand, the availability of derivative information, such as gradient, and Jacobian or Hessian operators, allows for achieving higher rates of convergence towards a local optimum. Among several derivative-dependent methods, Newton's method has been proven to be the most effective. However, the use of derivatives can be prohibitively expensive and even unfeasible to compute. In that sense, nonlinear methods such as Broyden's and Levenberg-Marquardt's have been found to be very effective [2]. In any case, the inclusion of physical bound constraints is a powerful means to reduce the search space and speed up the computation. In this setting, interior-point methods are perhaps the methods of choice to generate iterates in the feasible region.

The focus of this paper is to extend the capabilities of two of these approaches by combining global stochastic and local deterministic optimization methods. The former allows for a better exploration of the parameter space without relying on derivatives. Also, it yields to conjecture the regions where the global solution may be located. Then, a local method is used to obtain a feasible optimal solution in the regions to speed up the process. The global stochastic optimization is performed by means of the simultaneous-perturbation-stochastic approximation (SPSA) algorithm [3]. This algorithm is efficient in high-dimensional problems in providing a good approximate solution while using few function values. The philosophy of this algorithm is to perform random simultaneous perturbations of all model parameters to generate a descent direction at each iteration. This algorithm has been successfully used in several application areas [1, 4, 3]. We use SPSA to explore the parameter space, i.e., the domain of the function. The result of this exploration consists of target regionsq where the function values are low and the global minimum may be located. Using the data in this target regions, we create a surrogate model that approximates the function in this neighborhoods. Since first-order information can be calculated from the surrogate model, then we are able to apply a local optimization algorithm that we call Newton-Krylov Interior-Point(NKIP) to find the minimum of the surrogate model and incorporate some additional inequality constraints. We test the hybrid approach on a set of small, medium and large scale problems.

2   Problem Formulation

We consider the global optimization problem in the form:

$$\text{minimize} \quad f(x) \tag{1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$, and the global solution x^* is such that

$$f(x^*) \leq f(x) \text{ for all } x. \tag{2}$$

In particular, we are interested in problem (1) that has many local minima.

A general parameter estimation problem can be written as a nonlinear least squares problem:

$$\text{minimize } ||G(x) - d||_2 \tag{3}$$

where $G : \mathbb{R}^n \to \mathbb{R}^m$ is a nonlinear function that may involve the discrete solution of a set of ordinary or partial differential equations, $x \in \mathbb{R}^n$ is the vector of parameters to be characterized, and $d \in \mathbb{R}^m$ is a set of discrete data observations. The variables n and m represent the number of parameters and data observations with $n \ll m$, respectively. This problem is ill-posed since there are infinitely solutions x that satisfy problem (3).

In the next section, we describe a hybrid approach for obtaining feasible global solutions.

3   Hybrid Approach

Our interest is to combine a global and local strategies for solving problem (1). We use SPSA as a global method to explore the domain of the function by starting with different initial guesses. This increases the chances for finding regions where a global optimal solution may exist, and allows a rich sampling of the parameter space. We select the regions with the lowest function values and filter the data by eliminating points outside such regions. This filtering is done by using a predefined radius. Using the selected data inside the target region, we create a surrogate model that approximates the function in this neighborhood. In our particular case, we are interested in finding an analytical representation of the original function of quadratic or cubic order. The surrogate model created provides us with a smooth approximation of the objective function within the region of the most promising optimal regions explored by SPSA.

Once the surrogate model is constructed, we can apply our local strategy NKIP. The derivatives necessary to apply this strategy are available from the analytical representation of the surrogate model. The NKIP approach allows for further refinement of the solution yield by SPSA. Moreover, the solution points obtained by NKIP algorithm are validated against the original model.

Figure 1 shows our hybrid optimization scheme. The next subsections describe the major optimization components used in the hybrid approach.
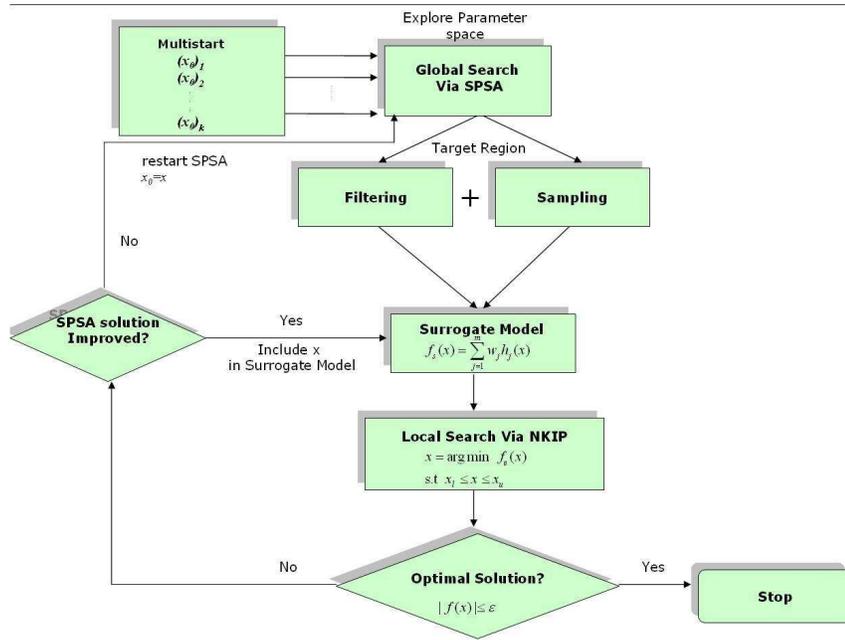
Figure 1: Hybrid algorithm scheme.

3.1 Simultaneous Perturbation Stochastic Approximation

Simultaneous Perturbation Stochastic Approximation (SPSA) is a stochastic steepest descent direction algorithm introduced by Spall [3], where a current point is improved by moving randomly in the direction of the negative gradient of the objective function. The algorithm does not depend explicitly on derivative information. A major advantage of SPSA is that it uses only two objective function evaluations at each iteration to obtain the update of the parameters. The disadvantage is that the algorithm has a slow rate of convergence and it cannot incorporate, in general, the constraints associated with the problem.

The essential feature of SPSA, which provides its power and relative ease of use in difficult multivariate optimization problems, is the underlying gradient approximation that requires only two objective function evaluations per iteration independent of the dimension of the optimization problem. Such evaluations are made by simultaneously varying all of the variables in the problem (the simultaneous perturbation), in contrast with the classical finite-difference method where the variables are varied one at a time. If the number of parameters being optimized is n, then the finite-difference method takes $2n$ evaluations of the objective function at each iteration to obtain a gradient approximation. Klie et al. [4] have used SPSA to solve large scale reservoir simulation problems obtaining encouraging results, but such results can be improved by the inclusion of constraints.

Next, we present a pseudocode of the standard SPSA algorithm. The user initializes the starting point x, the number of parameters being optimized n, the maximum number of iterations k_{max}, and the SPSA parameters a, A, c, α, and γ. The algorithm

calls an external function f to evaluate the objective function.

Algorithm 1 SPSA
Input: $x, n, k_{max}, a, A, c, \alpha, \gamma$
Output: x
1: **For** $k = 1 : k_{max}$
2:     Set $a_k = a/(k + A)^\alpha$ and $c_k = c/k^\gamma$
3:     Set the random perturbation vector $\Delta_k = 2(round(rand(n, 1)) - 1)$
4:     Simultaneous perturbation $x^+ = x + c_k\Delta_k$ and $x^- = x - c_k\Delta_k$
5:     Function evaluation $y^+ = f(x^+)$ and $y^- = f(x^-)$
6:     Set step $\Delta x = (y^+ - y^-)./(2c_k\Delta_k)$
7:     Update $x = x - a_k\Delta x$
8:     Check upper and lower constraints on x
9: **end**

The choice of the input values a, A, c, α, γ plays a fundamental role for the performance of SPSA. In our numerical experimentation, we set the values of α and γ to 0.602 and 1.101, respectively. Typically in a high-noise setting (i.e., poor quality measurements), it is necessary to pick a smaller a and larger c than in a low-noise setting.

3.2   Surrogate Model

Most problems require experiments and/or simulations to evaluate objective and constraint functions in terms of design variables. Frequently, optimization strategies require thousands or even millions of evaluations, and often the associated high cost and time requirements render this infeasible.

We can overcome the computational cost of the optimization by constructing analytical expressions, known as surrogate models, which locally approximate the original simulation or forward model with a limited amount of data. These approaches are very appealing when the simulation or forward model is based on the solution of a set of differential equations consisting of hundreds of thousands to millions of elements and with a significant number of time steps. The accuracy of the surrogate model depends strongly on the number and location of samples.

The most popular surrogate models are provided by polynomial response surfaces, kriging, artificial neural networks and radial basis function (see, e.g., [6]). In particular, our hybrid optimization approach seems to provide more reliable results when using radial basis functions to create the surrogate model.

Using the data provided by SPSA $(x_k, f(x_k))$, $k = 1, ..., p$, we find the surrogate model $f_s(x)$ using an interpolation method given by

$$f_s(x) = \sum_{j=1}^{m} w_j h_j(x), \tag{4}$$

where the multiquadric basis functions are given by

$$h_j(x) = \sqrt{1 + \sum_{i=1}^{n} \frac{(x_i - c_{ij})^2}{r_{ij}}},$$

where c_{ij}, r_{ij}, w_j are determined by the method to create the surrogate model and $i = 1, ..., n, j = 1, ..., m$. We also consider the Gaussian basis functions defined by

$$h_j(x) = e^{-\sum_{i=1}^{n} \frac{(x_k - c_{ij})^2}{r_{ij}^2}}. \tag{5}$$

A radial basis function (RBF) is typically parameterized by two sets of parameters: the center c, which defines its position, and a second set of parameters r that determines the shape (width or form) of a RBF.

3.3   Newton-Krylov Interior-Point (NKIP) Algorithm

Interior-point methods have been a successful technique for solving problems that involve inequality constraints. We describe the NKIP algorithm which is a linesearch Newton-Krylov method for solving general nonlinear programming problems. This algorithm was developed for obtaining an optimal solution for large scale and/or degenerate problems by Argáez and Tapia [7]. The most expensive part of the algorithm is the need for solving iteratively a least squares problem, but the rest of the operations are just matrix-vector multiplications.

Our strategy is to find an optimal solution of the surrogate model (4) by using NKIP. Since NKIP can handle inequalities constraints we can solve the following constrained optimization problem:

$$\begin{array}{c} \text{minimize } f_s(x) \\ \text{subject to } x \geq 0. \end{array} \tag{6}$$

The Lagrangian function associated with problem (6) is given by

$$l(x, y, z) = f_s(x) - x^T z, \tag{7}$$

where $z \geq 0 \in \mathbb{R}^n$ are the Lagrange multipliers associated with the inequality constraints.

In order to promote global convergence of the Newton interior-point method and keep iterates away from the boundary of the feasible region, except at the solution of the problem, the algorithm is based on the perturbed Karush-Kuhn-Tucker (KKT) conditions given by

$$F_\mu(x, z) \equiv \begin{pmatrix} \nabla_x \ell(x, z) \\ XZe - \mu e \end{pmatrix} = 0 \tag{8}$$
$$(x, z) \geq 0,$$

where $\nabla_x \ell(x, z) = \nabla f_s(x) - z$ is the gradient of the Lagrangian function, $X = \text{diag}(x), Z = \text{diag}(z), e = (1, ..., 1)^T \in \mathbb{R}^n$, and $\mu \geq 0$ stands for the perturbation parameter. For $\mu = 0$, these conditions are the KKT conditions associated with problem (6).

The methodology of interior-point methods is based on keeping $(x, z) > 0$ during the entire procedure except at an optimal solution. A point (x, z) is said to be an interior point if and only if $(x, z) > 0$.

For a given $\mu > 0$, the Newton step $(\Delta x, \Delta z)$ at each interior point (x, z) is obtained as the solution of the following non-symmetric linear system:

$$\begin{pmatrix} \nabla^2 f_s(x) & -I \\ Z & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta z \end{pmatrix} = \begin{pmatrix} z - \nabla f_s(x) \\ \mu e - XZe \end{pmatrix}. \tag{9}$$

To guarantee that the updated iterate is an interior point, a step-length $\alpha \in (0, 1]$ is selected such that $(x + \alpha \Delta x, z + \alpha \Delta z) > 0$.

Under this framework, we state a path-following strategy for solving (6).

Path-Following Strategy: *For $\mu > 0$, and working from the interior, apply a linesearch inexact Newton's method to the perturbed KKT conditions. Then decrease μ and repeat the process. Under appropriate conditions, an optimal solution will be obtained as μ approaches zero.*

Since our interest is solving large scale problems, we use inexact Newton directions for solving (9), and coupling it with the globalized strategy. In particular, we present an algorithm for obtaining a Newton-Krylov direction. In line with this objective, the above linear system is decoupled in the following manner: Once the direction Δx is known, then from the second block of equations we have

$$\Delta z = X^{-1}(\mu e - XZe - Z\Delta x). \tag{10}$$

Then substituting Δz in the first block of equations we have

$$(\nabla^2 f_s(x) + X^{-1}Z)\Delta x = \mu x^{-1} - \nabla f_s(x) \tag{11}$$

where x^{-1} denotes the vector in $\mathbb{R}^n$ whose components are the reciprocals of the components of x. Then, once Δx is obtained from (11), Δz is computed using simple arithmetic operations in (10).

Even though we still require to solve a linear system, this reduced system of equations has several advantages over system (9). First of all, the size of this system is half the size of system (9). Second, the error of (11) is equivalent to the total error for solving (9). And third, the coefficient matrix in system (11) is always symmetric and by adding a diagonal positive matrix, $X^{-1}Z$, we increase the likelihood that the matrix $\nabla^2 f_s(x) + X^{-1}Z$ will be positive definite. This fact is of great importance because it yields the implementation of an iterative procedure such as the conjugate gradient method for obtaining an approximate solution.

4   Numerical Results

In this section, a numerical study of the hybrid optimization algorithm is conducted on five small cases and on a set of large scale parameter estimation problems that arise in reservoir simulations.

4.1   Small Problems

A set of five small test global optimization problems was selected from the literature where the number of variables ranges from 2 to 100.

Problem 1 (Pinter's Problem [9]). The function to be minimized is

$$f(x, y) = -2(sin(x + 4y) - 2cos(2x + 3y) - 3sin(2x - y) + 4cos(x - 2y))$$

where $x, y \in (-5, 5)$. This function has two global solutions (x^*, y^*) at

$(-3.4333, 1.2852)$ and $(2.84988, 1.2852)$ with $f(x^*, y^*) = -1.937$.

Problem 2 (Rastringin's Problem [10]). We minimize the function:

$$f(x_1, x_2) = 20 + \sum_{i=1}^{2}(x_i^2 - 10cos(2\pi x_i))$$

subject to $-5.12 \leq x_i \leq 5.12$ and $i = 1, 2$. The global minimum is located at $x^* = 0$ with $f(x^*) = 0$.

Problem 3 (The Modified Bratu Problem [13, 14]). This problem is given by

$$\nabla^2 u + \alpha\frac{\partial u}{\partial x} + \lambda e^u = f \quad \text{in} \quad \Omega, u = 0 \quad \text{on} \quad \partial\Omega.$$

This problem plays an important role in combustion modeling and semiconductor design and processing, representing a simplified model for nonlinear diffusion phenomena. In the absence of the convection term, this operator is monotone with respect to u and hence it always has a solution for $\lambda < 0$. When $\lambda > 0$, there is a threshold value λ^* for which the equation has no solution for $\lambda > \lambda^*$ and at least one solution for $\lambda \leq \lambda^*$.

Problem 4-5 (Chandrasekhar H-Equation Problem [11, 12]). The function to be minimized is

$$F(u) = H(u) - \frac{1}{1 - \frac{\hat{c}}{2}\int_0^1 \frac{uH(\xi)}{u+\xi}d\xi} = 0,$$

with $u \in [0, 1]$. There are two solutions known for $\hat{c} \in (0, 1)$, and as this value approaches one, the problem becomes harder to solve. Here, we closely follow the specifications given in [12]; that is, $H(u) \equiv u$, $u^{(0)} = (0, \ldots, 0)^T$ and the composite midpoint rule to discretize the integral. The two variants of the H-equation are determined by setting $\hat{c} = .9$ and $\hat{c} = .999999$. For these two cases we specify 100 unknown variables obtained from the discretization of the integral equation.

For each problem, we use five different random initial points to start SPSA. Table 1 summarizes the setup of SPSA parameters a, c and A for test problems 1-5. We use the Gaussian RBF to create the surrogate model for Pinter, Rastringin, and the reservoir problem. The multiquadric RBF is used for Chandrasekhar and Bratu problems.

Table 1: Setup of the SPSA parameters.

Prob.	a	c	A
1	1.7e-2	3.9	1000
2	1.7e-1	1.9	100
3	1.7e-3	1.9	50
4	1.7	1e-1	50
5	1.7	1e-1	50

Table 2 summarizes the numerical results obtained by the hybrid approach for solving test problems 1-5. The first four columns give the problem number, the number

of variables, the value of the objective function at the global solution, and the initial value of the objective function respectively. The next two columns state the value of objective function obtained by SPSA and its number of iterations to reach this value. The last two columns show the ability of NKIP to reduce the objective function and its number of iterations to reach this value.

Table 2: Performance of the hybrid approach for solving test problems 1-5.

Prob.	n	$f(x^*)$	$f(x_0)$	$f_{\mathrm{SPSA}}(x)$	SPSA iterations	$f_{\mathrm{NKIP}}(x)$	NKIP Iterations
1	2	-1.937	6.27e-01	-1.409	400	-1.937	17
2	2	0	3.10e+01	1.070e-01	118	3.55e-15	2
3	2	0	7.29e+00	6.42e+00	200	4.88e-01	7
4	100	0	2.03e+01	9.57e+00	200	4.47e-01	6
5	100	0	2.22e+01	7.58e+00	200	5.34e-01	5

Figure 2 shows the results in terms of function values versus the number of iterations. The blue and red lines represent the iterations done by SPSA and NKIP, respectively. Here the blue line represents the best solution found by SPSA from one of the five initial points. Each iteration of SPSA requires two function evaluations of the objective function. The red line represents the iterations of NKIP when minimizing the surrogate model $f_s(x)$. Each NKIP iteration requires one function evaluation. If an optimal solution was not found, such function evaluation is added to update the surrogate model and NKIP is run again.

4.2 Two-Phase Flow Parameter Estimation Problem

We experimented with a large scale parameter estimation problem where the goal is to estimate the permeability field based on pressure data measurements. This analysis was included to show that, despite having full information of the fluid flow pressure field, the inverse problem is still highly ill-posed and has multiple local minima. Therefore, strategies to regularize and re-parameterize the inverse problem need to be employed.

We formulate the parameter estimation as a nonlinear least squares problem:

$$\min f(x) = \frac{1}{2}(G(x) - d)^T W(G(x) - d),$$

where $x \in \mathbb{R}^n$ represents the unknown permeability field, the term $G(x) - d$ measures the mismatch between the simulated pressure $G(x)$, and the observed pressure d, and W is a positive diagonal matrix.

We run the hybrid optimization algorithm in conjunction with the simulator framework Integrated Parallel Accurate Reservoir Simulator (IPARS) [15] using high-performance computing (HPC). The problem consists of finding a permeability field that involves 2000 parameters. The field is parameterized by the singular value decomposition (SVD) [4]. We chose several initial points, adding a percentage of noise to the true singular values: 5%, 10%, and 15%. Table 3 summarizes the results obtained by the hybrid approach.

The first row indicates the level of noise used in the experimentation. The second and third rows indicate the number of initial points used for each level and the range of the objective function for the initial points. The next three rows indicate the total number of iterations, the number of function evaluations used by SPSA and the range
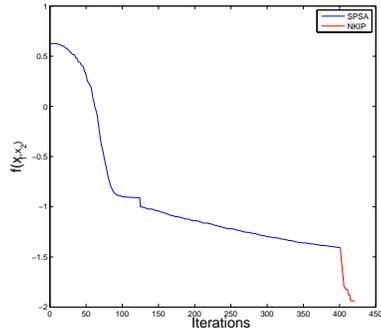
(a) Problem 1 with $n = 2$.
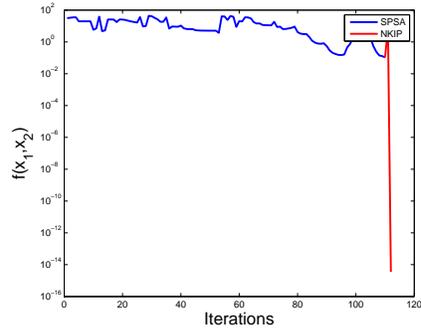
(b) Problem 2 with $n = 2$.
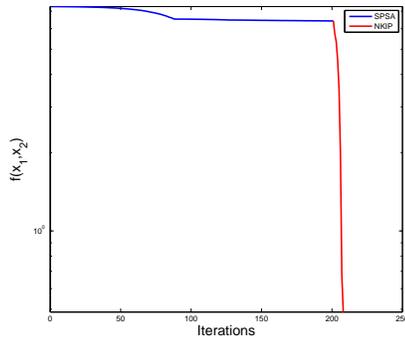
(c) Problem 3 with $n = 2$.
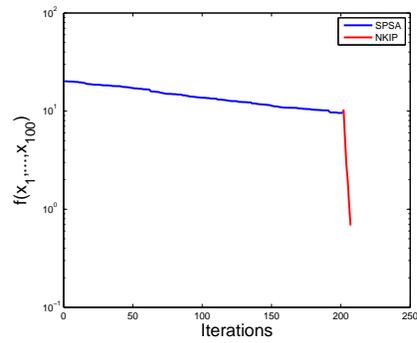
(d) Problem 4 with $n = 100$ and $\hat{c} = 0.9$.
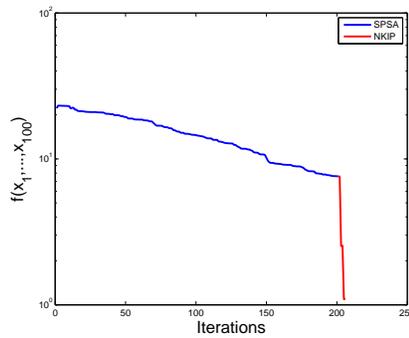
(e) Problem 5 with $n = 100$ and $\hat{c} = 0.999999$.

Figure 2: Convergence history for test problems 1-5.

Table 3: Numerical results for the two-phase parameter estimation problem.

Percentage of noise	5%	10%	15%
Number of initial points x_0	30	20	20
Range of $f(x_0)$	8.1 - 10.2	18.72 - 53.9	29.46 - 120.56
SPSA iterations	3645	2248	1647
SPSA total tunction evaluations	7290	4496	3294
Range of $f_{SPSA}(x)$	0.568 - 2.5	.858 - 6.785	1.541 - 15.89
Data points used to create $f_s(x)$	156	81	57
NKIP iterations	94	67	110
f_{KNIP}	0.478	0.747	1.465

of the objective function reached by SPSA. The next row indicates the number of data points used to create the surrogate model. The last two rows show the total number of iterations and the valued reach by NKIP.

5   Conclusions

The proposed hybrid approach exploits the best of SPSA and NKIP methods for achieving maximum efficiency and robustness in the search of a feasible global solution. It is clear that a broad exploration of the parameter space should be done to increase the chance of finding a global optimum when using SPSA. On the other hand, this search has to be performed in a controlled way to keep the number of function evaluations within reasonable computational cost. In this paper, we showed how this can be achieved by generating surrogate models and adjusting the degree of exploration as the iterations proceed.

The hybrid algorithm presented in this paper is a non-verified method since it cannot necessarily provide the global solution. However, the existing global optimization techniques benefit from the use of the hybrid algorithm since it eliminates regions in which the function value is high. Moreover the solution provided by the hybrid method is guaranteed to be feasible with respect to physical bound constrained set by the problem. Also, the multi-start approach of SPSA improves the likelihood of obtaining the feasible global solution.

We show promising numerical results that illustrate the efficiency of the hybrid code for solving small test-cases and large scale parameter estimation problems. Further experiments should be conducted using high-performance computing to exploit the hybrid scheme for solving very large scale problems. Further levels of parallelism may be exploited when computing the SPSA step and function evaluations. This paper opens new research questions related to the ill-posed associated with parameter estimation problems: regularization, parameterization, preconditioning, and uncertainty and sensitivity analysis. The regularization term can be added to the objective function in order to stabilize the ill-conditioning of the problem. The parameterization schemes seek to reduce the parameter space and, yet, be able to come up with better estimations. Preconditioning aims at improving the efficiency of iterative solvers. The uncertainty and sensitivity assessments can determine reliability bounds of the solutions. Further, the surrogate model framework offers interesting possibilities for

developing answers to these open questions.

References

[1] P. Pardalos, M. Resende, Handbook of Applied Optimization, Oxford, University Press (2002)

[2] J.E. Dennis, R.B. Schnabel, Numerical methods for unconstrained optimization and nonlinear equations, Prentice–Hall, Englewood Cliffs, New Jersey (1983)

[3] J.C. Spall, Introduction to stochastic search and optimization: Estimation, simulation and control, John Wiley & Sons, Inc., New Jersey (2003)

[4] A. Rodriguez, H. Klie, S. Thomas, M. Wheeler, A multiscale and metamodel simulation-based method for history matching, in 10th European Conference on the Mathematics of Oil Recovery (ECMOR), EAGE, Sept. 4–7 (2006)

[5] J.C. Spall, Implementation of the Simultaneous Perturbation Algorithm for Stochastic Optimization, IEEE Transactions on Aerospace and Electronic Systems **34**(3), 817–823 (1998)

[6] V. Keeman, Learning and Soft Computing: Support Vector Machines, Neural Networks and Fuzzy Logic Machines, The MIT Press (2001)

[7] M. Argáez, R. Tapia, On the global convergence of a modified augmented Lagrangian linesearch interior-point Newton method for nonlineaer programming, J. Optim. Theory Appl. **114**, 1, 1–25 (2002)

[8] M. Argáez, H. Klie, C. Quintero, L. Velazquez, M.F. Wheeler, Hybrid optimization algorithm for solving parameter estimation problems, PAMM Journal, ICIAM Proceedings, Vol. 7, 1062507–1062508, December 2007, DOI: 10.1002/pamm.20070094 (2007)

[9] J. Pinter, Continuous global optimization software: A brief review, Optima, the Newsletter of the Mathematical Programming Society **52**, 1, 1–8 (1996)

[10] A. Keane, P. Nair., Computational Approaches for Aerospace Design: The Pursuit of Excellence, Wiley, England (2005)

[11] S. Chandrasekhar, Radiative Transfer, Dover, New York (1960)

[12] C. Kelley, Iterative Methods for Linear and Nonlinear Equations, in Frontiers in Applied Mathematics, SIAM, Philadelphia (1995)

[13] R. Glowinski, H. Keller, L. Reinhart, Continuation-conjugate gradient methods for the least squares solution of nonlinear boundary value problems, SIAM J. Sci. Stat. Comput. **4**, 793–832 (1985)

[14] J. Moré, in Lectures in Applied Mathematics, Eds. E. Allgower, and K. Georg, American Mathematical Society. **26**, 723–762 (1990)

[15] H. C. Edwards, A parallel iterative solver for multiphase flow models in the Implicit Parallel Accurate Reservoir Simulator (IPARS), TICAM Report, The University of Texas at Austin (1998)