City University of New York (CUNY)
CUNY Academic Works

International Conference on Hydroinformatics

2014

Optimization Of Groundwater Remediation With New Efficient Parallel Algorithm For Global Optimization

Min Pang

Christine A. Shoemaker

How does access to this work benefit you? Let us know!

More information about this work at: https://academicworks.cuny.edu/cc_conf_hic/329
Discover additional works at: https://academicworks.cuny.edu

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

OPTIMIZATION OF GROUNDWATER REMEDIATION WITH NEW EFFICIENT PARALLEL

MIN PANG (1), CHRISTINE ANN. SHOEMAKER (2)
(1): Civil and Environmental Engineering, Cornell University, Ithaca, NY14853, U.S.A
(2): Civil and Environmental Engineering and Operations Research and Information Engineering, Cornell University, Ithaca, NY14853, U.S.A

INTRODUCTION

Groundwater Remediation analysis of a complex system requires a computational expensive simulation model. It is hence helpful for optimization to have an effective Parallel optimization algorithm to reduce the time required to get results. This application is applied to two Groundwater superfund remediation sites. One is Umatilla Chemical Depot and the other is Blaine Naval Ammunition Depot. These problems are computationally intensive, as they require solving large-scale system of partial differential equation. Computation takes a large amount of time, especially for Blaine problem, which requires nearly thirty minutes for completing the simulation for a single set of decision variables. In our study, we found out that Parallel Stochastic Radial Basis is an efficient algorithm to use on both the Umatilla and Blaine problems.

Model formulation

The formulation of the remediation problem consists of a minimization function subject to constraints. Both the objective goals of two problems (Umatilla and Blaine [1]) are to minimize the sum of life-cycle cost of the pump and treat system and penalty on constraints violations, while all the variables are bounded in a box constraint.

(1) Umatilla Chemical Depot

The formulation of remediation problem on Umatilla site consists of pumping system of eight pumping wells and two recharge basins. The modeling period is one management period of four years with pumping rate kept the same. The objective is to minimize the operation cost for this 4 years period constrained by the total pumping rate and cleanup level at the end of four years.

(2) Blaine Naval Ammunition Depot

For remediation problem on Blaine site, there is a similar objective goal to the Umatilla problem, which is to minimize the cost for entire project duration. However, the modeling period of Blaine problem consists of 6 management periods of 5 years, and the system contains 15 pumping wells. The objective function is to minimize the management cost including fixed cost related to facility installation once per each management period and maintenance and operation cost for the whole project duration. There is a cleanup level constraint for each of the indicator containments at the end of the project.

Algorithm

Parallel Stochastic Global optimization algorithm using Radial Basis function (RBF) with restart developed by Regis and Shoemaker (2009) [2] is an extended version of Serial Stochastic RBF by Regis and Shoemaker (2007) [3]. Its python code is an open source package implemented by Mueller et al. (2013)[4]. Stochastic RBF is suitable a for global optimization problem with expensive function evaluation. It is a derivative-free method, which can solve a black-box problem efficiently. Each iteration of Parallel Stochastic RBF algorithm consists of building a RBF model to approximate expensive function and using this model to select multiple points for simultaneous expensive function evaluation by multiple processors.

In the serial version of Stochastic RBF, we update the RBF surface using only one newly evaluated point at each iteration, while in the parallel version, we chose to select different number of points evaluated based on number of processors used in order to test the performance of the algorithm. The parallelism scheme of the algorithm is in "master-slave" form. Each time after obtaining the points selected from candidate points using RBF surface and distance criteria, the "master" processor distributes P points to P "slave" processors to generate real function evaluation value. Then after all slave processors finished their job, the "master" processor gathers the results from P "slaves", then uses all real function evaluations values have to update new RBF. The termination criterion is when the maximum evaluation is reached.

Implementation

In this Simulation-Optimization problem, the simulation of groundwater flow is solved by MODFLOW with its latest version MODFLOW2005 [5], while the contaminant transport and fate is simulated by MT3DMS developed by Chunmiao Zheng [6].

MODFLOW is a three-dimension finite difference groundwater model maintained by U.S. Geological Survey. The input of MODFLOW is domain discretizing data, hydro-geological data including transitivity, initial head and hydraulic conductivity in each discretized node, and pumping data containing pumping well locations and pumping rate. The output from MODFLOW (ie the saturated thickness, fluxes across cell interfaces in all directions, and locations and flow rates of various sinks/sources, including transient groundwater storage) is used in MT3DMS to generate concentrations of the contaminants for the whole simulation period. Then, the concentrations are taken as indicators parameters of the optimizations in order to identity best pumping rates that minimize the management cost.

The implementation of simulation-optimization problem is run on the NSF Yellowstone supercomputer, a petascale computing resource in the NCAR-Wyoming Supercomputer Center. It is a 1.5-Petaflops-cluster computer, which contains 9,036 2.6-GHz Intel Xeon E5-2670 8-cores processors. This HPC system provides a large environment to test the performance of the algorithm which we are using.

Result and discussion

We first applied Serial StochasticRBF algorithm on both the Umatilla and Blaine problem using Yellowstone supercomputer described above. For Umatilla, we run 1000 expensive evaluations. Each simulation takes around 1.5 mins, so total wall clock time is 94314s (26.2 hr). Blaine is a more complicated problem; it takes 30 mins for each simulation, with a total of 8 days for 400 iterations.

Comparing the result for Umatilla Problem using Parallel RBF and Serial RBF, we can see that we are able to obtain the best result sooner with more processors used, as shown in Table 1. That means we can achieve better algorithm efficiency having a larger pool of working processors. In terms of parallelism efficiency, we compute the total wall clock time to reach to 1000 function evaluations for all three cases (with one, four or eight processors). Then, the speed up is the ratio of the walkclock time spent on parallel algorithm and the wallclock time for the serial algorithm.. Table 2 shows that we can get high efficiency close to 100% using four and eight processors. However, the efficiency decreases as more processors are used, since communication between processors will take a larger portion of the total wall clock time.

Table 1. Number of evaluations for each number of processors to get to the same average value of the objective function among 5 trials as is obtained by the serial algorithm

Number of processors	Number of evaluations to get to best average of serial algorithm within 1000 evaluations
1	899
4	889
8	759

Table 2. Computation time fraction for 1000 evaluations including total wall clock time (T_{wc}), function evaluation time (T_{cp}), communication time and optimization time (T_{cm} +T_{op}) with different number of processors

No. Processors	Total T_{wc}(s)	T_{cp}(s)	T_{cm} +T_{op}(s)	Speed up	Efficiency
1 proc	94314 (26.2hr)	94036.	278	---	---
4 proc	25345 (7.04hr)	25235	110	3.7	0.975
8 proc	13235 (3.68hr)	13097	138	7.1	0.887

The average result from 5 trials shows that optimal value found converges to the same value for all different multiple processors cases. Generally, it's hard using a large pool of processors to find a good optimal within a limited evaluation. However the advantages of having more processor members in the pool is that we can explore more thoroughly around the best point found so far, which in turn would prevent us from the missing the optimal value if it's located in a narrow valley.

Overall, time saving is very significant with more processors used, even though a large number l of processors takes longer on communication and transfer data to calculate. The disadvantage of using a very large pool of processors is that the performance of the algorithm could deteriorate due to the less frequent updates of the surrogate surface, although this is not seen with 8 processors. With 8 processors the performance is remarkable good since a better average solution is obtained with fewer objective function evaluations with 8 processors than with one or four processors.

ACKNOWLEDGEMENTS
This research was supported in part by NSF CISE grant 1116298 to PI Shoemaker.

REFERENCES

[1] Naval Facilities Engineering Command (NAVFAC) technical report TR-2237-ENV, (2004)

[2] R. G. Regis and C. A. Shoemaker. Parallel Stochastic Global Optimization using Radial Basis Functions. (Summer 2009) "INFORMS Journal on Computing", Vol. 21, No. 3, pp. 411-426

[3] R. G. Regis and C. A. Shoemaker. A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions. "INFORMS Journal on Computing", (Fall 2007) Vol. 19, No. 4, pp. 497-509

[4] J. Mueller, C. Shoemaker and R. Regis, Codes for surrogate model based optimization: Stochastic RBF python code, (2013) https://courses.cit.cornell.edu/jmueller/

[5] Arlen W. Harbaugh, MODFLOW-2005, The U.S. Geological Survey Modular Ground-Water Model—the Ground-Water Flow Process

[6] MT3DMS: A Modular Three-Dimensional Multispecies Transport Model for Simulation of Advection, Dispersion, and Chemical Reactions of Contaminants Groundwater Systems; Documentation and User's Guide

[7] R. G. Regis and C. A. Shoemaker. Combining Radial Basis Function Surrogates and Dynamic Coordinate Search in High-Dimensional Expensive Black-Box Optimization Engineering Optimizatiom Vol. 00, No. 00, (October 2011), 1–28

[8] R. G. Regis and C. A. Shoemaker. A Quasi-Multistart Framework for Global Optimization of Expensive Functions using Response Surface Models

[9] A. Singh, Dissertation on efficient optimization of computationally expensive problems using a new parallel algorithm and response surface based methods