ve-Free Hybrid Optimization Methods for Solving Simulation-Based Problems in Hy

organized by
Owen Eslinger, Kathleen Fowler, and Genetha Gray

Workshop Summary

The focus of this workshop was to explore innovative ways to use derivative-free optimization algorithms for hydrology problems. Derivative-free methods are not traditionally used in the general optimization community, nor, in particular, by hydrologists to tackle the challenges inherent to the geosciences. This class of problems typically has expensive function calls to black-box groundwater flow and transport simulators, strong dependence on the initial data, and small, disconnected feasible regions due to the heterogeneous nature of the subsurface or the constraints on the physical water-bearing region. To reach the workshop goals, we invited researchers in three categories: optimization, sampling techniques, and applied mathematicians and environmental engineers who regularly deal with expensive simulation-based design problems in hydrology. Participants included a diverse mix of professors, post-doctoral fellows, graduate students, and even an undergraduate student as well as researchers from a variety of industrial settings and government labs.

The organizers feel that this workshop exceeded all expectations. We have received extremely positive feedback from the workshop participants. Promising new collaborations formed early in the week and the afternoon discussions typically continued long into the evening. These fruitful conversations alone proved, without a doubt, that the workshop was a success. In this report we will highlight some of the outcomes with an emphasis on emerging collaborations and projects we hope will lead to scientific advancements in this field.

- **Building a suite of simulation-based test problems from hydrology**. The group has collected a set of problems of varying difficulty; these will be posted on the web for algorithmic testing. These sets include the community problems from hydrology, a landfill liner problem, a set of groundwater flow and transport calibration problems, and a group of problems from a Navy waste site which come by way of the Environmental Protection Agency. Moreover, a working group from this workshop also used the Groundwater Modeling Software (GMS) to generate the simulation files for the FEMWATER density driven flow package to model the island of Crete. Once this model is running properly the files will also be available to downloading.
- **Ranking management strategies in the presence of uncertainty through expected loss**. This discussion focused on Karen Ricciardis challenging management problem for seawater intrusion (at Trurro) under uncertainty in the hydraulic model parameters. These problems are typically posed as multi-scenario realizations of the physical system as input to the optimization method. Karen began working with statisticians Max Morris, Bobby Gramacy, and Crystal Linkletter to understand how to better compare management strategies that incorporate uncertainty in input.

They proposed using a loss function to calculate the expected loss across input distribution of each management strategy. Considerations include the choice of loss function, number of realizations sampled from the input distribution, how to deal with simulation failure, and how to compare strategies.

- **Joint prediction of outputs and feasible regions in the context of optimization**. The focus of this work is on how to make use of constraint violations and/or possible simulation failures to help guide the optimization process. This is often debated in the context of derivative-free optimization for computationally expensive, simulation-based problems. Efficiency of these methods can be degraded by spending significant time evaluating infeasible points. However, there may be a trade-off in understanding the design space if these points are considered. At the workshop, these ideas were discussed both at the entire group level and in various smaller working groups. This topic was of interest to nearly every participant. The focus was finally narrowed down to hidden constraintsthat is, when the simulator fails to return a value and typically, a NaN (a value corresponding to Not a Number) is assigned. A group including Max Morris, Bobby Gramacy, Crystal Linkletter, and Herbie Lee have outlined an approach to predict the probability that a design point will return a NaN and incorporate this information into a surrogate framework. They formulated an approach and actually obtained promising numerical results on a toy problem that implied guidance can be provided so that algorithms will spend less time requesting function evaluations in a NaN region. Stefan Wild has since sent them more data to test and Katie Fowler will also be sending them a hydraulic capture application. The hope is that this will lead to a journal paper.
- **Metaheuristics and Pattern Search Hybrids**. Several suggestions were made to enhance the features of the HOPSPACK software, developed at Sandia National Labs. In particular, Josh Griffin led a group discussion on how to include a binary genetic algorithm (GA) into the HOPSPACK framework that would allow for integer or categorical variables. These ideas were then specifically outlined in terms of implementation on the so-called community problems. We outline the three possibilities discussed:
  - Standalone GA Citizen: The first pass would be to incorporate an original existing GA (such as NSGA2) into the existing HOPSPACK framework. This could then be provided essentially as a standalone solver that would be similar in nature to DIRECT in that it could be used by GSS but not visa versa.
  - Apply GA on existing evaluations: In this approach, we would be able to construct a Citizen that is collaborative in that it would select a subset of the existing evaluation cache to form a population. From this population mutations and combinations could be formed to create new points. The hardest part here would be deciding how best to select this subset.
  - Apply GA on existing evaluations and extract most promising candidates for evaluation from existing surrogate models. In this case we would use the surrogate model to rank the next generation of points from best to worst and then keep a percentage of the best for evaluation. To retain the global nature of the GA we may perhaps develop mechanisms that are not purely greedy with respect to the surrogate.

- **Interactive, web-based optimization**. This is something currently be investigate and developed by Thomas Hemker and also Shawn Mattot has a similar interest. Thomas would like to provide capabilities to allow the user to dynamically choose and modify parameter settings, along with algorithm selection, hence steering the optimization process via an interactive-GUI. On top of this, Thomas would like to be able to provide web-based optimization. Thus the optimization software need not be installed on the users machine. Instead the user would receive trial-points that they would then evaluate using their existing simulation code, and the send the complete evaluations back to the algorithm, running remotely. Further the user would not need to worry about giving away proprietary data and simulation software in order to have the problem tested by an experienced optimizer. This may be a feasible way to acquire a large data-base of real-life DFO test problems, while avoiding the confusion of needing to install an equal number of simulators and obtain corresponding licenses.
- **Exploiting the statistical information provided in the Treed Gaussian Process (TGP) surrogate framework**. There was much interest in the use of surrogate models to improve the efficiency of the optimization algorithms while exploiting sampled points. In particular, we spent one morning reviewing the R statistical software package and understanding the TGP approach. Discussions pointed towards extending TGP to handle integer variables and categorical variables with a focus on how to manage the surrogates. Additionally, we feel we have only scratched the surface in discovering how TGP can facilitate the hybrid approach. In particular, statistical analysis through the TGP approximation may help reduce the dimension of the problem, help us uncover feasible and infeasible regions, and also deliver sensitivity information for post-processing.
- **Extension of DFO methods to mixed-integer formulations**. These ideas are mentioned in the above sections as well, but in particular, one working group with Thomas Hemker, Shawn Mattot, Matthew Parno, and Tim Kelley spent time building the implicit filtering algorithm into a branch-and-bound framework. The group will apply this approach to the landfill liner problem as well as the hydraulic capture community problem.

The organizers will maintain a wiki (which has initially been set up by David Farmer) to post problems and results and to facilitate collaborations. We are hopeful that the outcomes from this workshop will lead to a special session at the 2009 SIAM Geosciences Conference in Leipzig, Germany as well as a special issue of the Pacific Journal of Optimization. The organizers would like to thank the entire AIM staff for providing such an enriching experience and productive atmosphere.