BENEFITS OF DEEPER ANALYSIS IN SIMULATION-BASED GROUNDWATER OPTIMIZATION PROBLEMS

Aswin Kannan[*] **and Stefan M. Wild**[*]

[*]Mathematics and Computer Science Division
Argonne National Laboratory, Argonne, IL 60439 USA
email: `akannan` or `wild @mcs.anl.gov`, web page: `http://www.mcs.anl.gov/~wild`

Key words: Black-box optimization, simulation-based optimization, groundwater

Summary.   A common dilemma in many decision, design, and modeling problems involving water resources is determining the level of simulation output to be passed to an optimization solver. A pure black-box approach allows for the easiest interface between the solver and the underlying simulator(s) but often requires a large number of simulation evaluations. On the other hand, full knowledge of the underlying computations enables the use of more specialized solvers but is labor-intensive, increases a solver's overhead, and can be intractable or unrealistic. We explore trade-offs occurring between these two extremes on a groundwater problem based on the Lockwood Solvent Groundwater Plume Site. We make explicit use of knowledge about the problem's nonsmoothness and additional outputs from the underlying simulator. We propose an augmented Lagrangian framework to solve the resulting problem under three different levels of information. Our results show both the benefit of working with richer output from the underlying simulator and the trade-off between computational complexity and accuracy with this increase in information.

1 INTRODUCTION

Many groundwater optimization problems are characterized by objective and constraint functions depending on the evaluation of complex numerical simulations. Examples include the "community problems"[3,10], groundwater bioremediation[15], and the well-field design problem[8,9] considered in this paper. The underlying numerical simulations include solvers for systems of partial or ordinary differential equations, nonlinear or nonsmooth equations, and numerical integration. In each case, the problem was modeled as a *black-box optimization problem*, where it is assumed that an optimization algorithm has access only to the objective/cost function through an evaluation.

Optimization approaches for this class of problems include finite difference-based, model-based, direct search, and heuristic methods. These methods belong to the field of *derivative-free optimization*[1], which concerns the study of problems where the analytic

expressions of derivatives of the objective/constraints with respect to the decision parameters are unavailable. Gradient-based methods rely on finite-difference derivatives, whereas model-based[1] approaches and implicit filtering[6] infer derivative values from generally coarser sampling. Direct search techniques[7] (such as pattern search and the Nelder-Mead method) sample the decision space independently of such approximations. Heuristics such as genetic algorithms tend to be population-based and balance stochastic exploration of the space with refinement. A large study[4] of some of these approaches on a class of black-box groundwater problems provides further insight.

Practitioners often resort to purely black-box formulations because of the ease of use and robustness of black-box solvers. Writing interfaces that pass additional details of the simulation components making up the "black box" is often not straightforward and can be error-prone. Furthermore, providing a richer set of simulation output can in fact make the objective function evaluation more expensive, while still not providing the full derivative information required by more efficient derivative-based optimization methods.

In this paper, we focus on the *Lockwood* problem[8,9], to illustrate the benefit of working with more than just *aggregates* of a set of black-box outputs. As detailed in Sec. 2, the Lockwood problem concerns the minimal cost design of a well field to prevent contaminant plumes expanding from a fixed region. The plume expansion is given by a set of simulation-based, nonsmooth flux constraints. We propose exposing this source of nonsmoothness and reformulating the problem in a more tractable, smooth version at the cost of adding a larger set of black-box inequality constraints.

In Sec. 3 we propose a model-based augmented Lagrangian framework to solve the resulting inequality constrained problem, as well as an inexact version that considers only a subset of the inequality constraints. A `Matlab` implementation of the algorithm is developed, and the problem is analyzed under different levels of information in Sec. 4. We show that providing more information to the optimization algorithm benefits both the number of simulation runs required and the quality of solution obtained. We also compare the trade-off between computational complexity and the level of information provided.

2   ILLUSTRATIVE LOCKWOOD DESIGN PROBLEM

To illustrate our approach, we consider a black-box optimization problem[8,9] based on the Lockwood Solvent Groundwater Plume Site (LSGPS) near Billings, Montana. The site consists of two chlorine plumes (A and B) that threaten to contaminate the neighboring Yellowstone River. The basic version of the problem is to determine pumping rates for six extraction wells in order to minimize the cost of operating the wells, subject to two (steady-state) constraints on the flux that prevent the plumes' migration. The cost objective f is the sum of a fixed installation cost and variable operating costs. In this case, the operating costs are a simple, linear function of the pumping rates, but the flux constraints are evaluated by the simulators `Ostrich` and `Bluebird`[2].

We let r_i refer to the pumping rate of well i and take r to denote the resulting six

decision parameters. Mathematically, the problem can be expressed as

$$\min_r \left\{ f(r) : \Phi_A(r) = 0; \Phi_B(r) = 0; 0 \le r \le u \right\}, \tag{1}$$

where the vector u denotes the maximum pumping rates and Φ_a and Φ_b represent the flux from plumes A and B, respectively. In a purely black-box approach, the problem has also been posed as the penalized, bound-constrained problem[9]

$$\min_r \left\{ f(r) + \mu|\Phi_A(r)| + \mu|\Phi_B(r)| : 0 \le r \le u \right\}. \tag{2}$$

By moving the simulation-based constraints into the objective through the use of a penalty parameter $\mu > 0$, this formulation enables the use of a larger set of black-box solvers. We refer to (1) and (2) as the problems with *some information* and *no information*, respectively.

Richer Simulation Output: Given pumping rates and locations, `Bluebird` uses analytic element methods[5] to discretize the contaminant site and compute a flow field. Plumes A and B are divided into 40 and 38 sections, respectively, across each of which the total flux is computed. For any section η, $\Phi_\eta^{\text{tot}} = \int_0^h q_\eta dl$, where h is the saturated thickness of the domain, q_η is the specific discharge normal to a vertical plane through the aquifer, l is the vertical direction, and Φ_η^{tot} is the total flux[8]. Computationally, this flux is split into two directional components, the in-flux Φ_η^{in} and the out-flux Φ_η^{out},

$$\Phi_\eta^{\text{tot}} = \Phi_\eta^{\text{out}} - \Phi_\eta^{\text{in}}, \qquad \Phi_\eta^{\text{out}} = \int_0^h q_\eta^+ dl, \qquad \Phi_\eta^{\text{in}} = \int_0^h q_\eta^- dl,$$

where $q_\eta^+ = \max(q_\eta, 0)$ and $q_\eta^- = \max(-q_\eta, 0)$.

A deeper inspection of the simulator reveals that the fluxes are numerically obtained by discretizing each section η into 201 elements and that the desired constraint values are given by the out-flux, so that

$$\Phi_A = \sum_{\eta=1}^{40} \sum_{i=1}^{201} q_{\eta,i}^+, \qquad \Phi_B = \sum_{\eta=41}^{78} \sum_{i=1}^{201} q_{\eta,i}^+.$$

Therefore, the original problem (1) can be equivalently formulated as

$$\min_r \left\{ f(r) : q_{\eta,i}(r) \le 0, \eta = 1, \ldots, 78, i \in \mathcal{A}; 0 \le r \le u \right\}, \tag{3}$$

where $\mathcal{A} = \{1, \cdots, 201\}$. The major advantage with the formulation (3) is that it eliminates the source of nonsmoothness arising from q_η^+ in Φ_A and Φ_B. This approach requires only that the necessary (new black-box) output, $\{q_{\eta,i}\}$, be provided instead of the aggregates Φ_a and Φ_b used by the problem with *some information*.

However, the 15,678 black-box inequality constraints in (3) can pose significant challenges in terms of computational overhead and susceptibility to roundoff error. If $q_{\eta,i}$ is linear in i, then the problem can be equivalently modeled such that $q_{\eta,1}, q_{\eta,201} \leq 0, \forall \eta \in \mathcal{N}$, thereby replacing 15,678 inequality constraints by 176. With this motivation, we propose an inexact version of the above problem where we replace the index set $\mathcal{A}$ by $\mathcal{A}_{\text{ie}}$, a set containing $|\mathcal{A}_{\text{ie}}| < 201$ equally spaced indices. For example, for $|\mathcal{A}_{\text{ie}}| = 3$ and $|\mathcal{A}_{\text{ie}}| = 5$, we would consider $\mathcal{A}_{\text{ie}} = \{1, 101, 201\}$ and $\mathcal{A}_{\text{ie}} = \{1, 51, 101, 151, 201\}$, respectively, and solve the inexact problem

$$\min_r \left\{ f(r) : q_{\eta,i}(r) \leq 0, \eta = 1, \ldots, 78, i \in \mathcal{A}_{\text{ie}}; 0 \leq r \leq u \right\}. \tag{4}$$

We refer to the inequality constrained problem, with both exact (3) and inexact (4) sets, as the problem with *more information.*

3 FAMILY OF BLACK-BOX-BASED OPTIMIZATION ALGORITHMS

We now propose a model-based derivative-free approach to solve the black-box optimization problems of interest. We begin by considering the inequality-constrained problem (4) where, for ease of notation, we denote the constraints by $\{c_j : j \in \mathcal{J}\}$, with $\mathcal{J} = \{1, \cdots, 78|\mathcal{A}_{\text{ie}}|\}$. By adding slack variables s, (4) can be reformulated as

$$\min_{r,s} \left\{ f(r) : c_j(r) + s_j = 0, j \in \mathcal{J}; s \geq 0; 0 \leq r \leq u \right\},$$

and then solved with a bound-constrained augmented Lagrangian scheme. At every outer iteration, for given estimates of the Lagrange multipliers and penalty parameter, (λ, μ), the bound-constrained problem

$$\min_{r,s} \left\{ h(r,s) = f(r) - \sum_{j \in \mathcal{J}} \lambda_j (c_j(r) + s_j) + \frac{\mu}{2} \sum_{j \in \mathcal{J}} (c_j(r) + s_j)^2 : s \geq 0; 0 \leq r \leq u \right\}$$

is solved. A standard technique[13] for solving such problems is to approximate the objective function by a quadratic model based on gradient and Hessian information and then solve a sequence of subproblems

$$\min_d \left\{ d^T \nabla h(r,s) + \frac{1}{2} d^T \nabla^2 h(r,s) d : s + d_s \geq 0, \ 0 \leq r + d_r \leq u, \ \|d\|_\infty \leq \Delta \right\}. \tag{5}$$

In our case, however, the derivatives $\nabla h(r,s)$ and $\nabla^2 h(r,s)$ are not available because h depends on the black-box quantities c_j. Instead of using finite-difference approximations, we propose to use models that interpolate known values of the c_j to determine the coarse approximations for the unknown gradient and Hessian terms. This approach has been shown to be efficient for minimizing the number of simulation runs required[1,11].

Thus, at every inner iteration, quadratic surrogate models of $f(r)$ and $c_j(r)$,

$$
\begin{aligned}
m^f(r^{\text{old}} + d_r) &= f(r^{\text{old}}) + d_r^T g^f + \frac{1}{2} d_r^T H^f d_r, \\
m^{c_j}(r^{\text{old}} + d_r) &= c_j(r^{\text{old}}) + d_r^T g^{c_j} + \frac{1}{2} d_r^T H^{c_j} d_r, \qquad j \in \mathcal{J},
\end{aligned}
$$

are formed based on the output from prior simulation evaluations obtained in a neighborhood of the current point r^{old}. The model parameters (g^f, H^f) and $\{(g^{c_j}, H^{c_j})\}_j$ are obtained by solving a multivariate interpolation problem[14]. We therefore use g^f, H^f, g^{c_j}, and H^{c_j} in place of the unavailable derivatives $\nabla f(r), \nabla^2 f(r), \nabla c_j(r)$, and $\nabla^2 c_j(r)$, respectively. The bound constrained quadratic problem (5) thus becomes

$$
\min_d \left\{ m(d) = d^T g + \frac{1}{2} d^T H d : 0 \leq r^{\text{old}} + d_r \leq u; \ s^{\text{old}} + d_s \geq 0; \|d\|_\infty \leq \Delta \right\},
$$

where $d = \begin{pmatrix} d_r \\ d_s \end{pmatrix}$, $H = \begin{pmatrix} H_{rr} & H_{rs} \\ H_{rs}^T & \mu I \end{pmatrix}$, $H_{rs} = \mu g^c$, and

$$
g = \begin{pmatrix} g^f + \sum_{j \in \mathcal{J}} (\mu c_j(r^{\text{old}}) - \lambda_j) g^{c_j} + H_{rs} s^{\text{old}} \\ \mu c(r^{\text{old}}) - \lambda + \mu s^{\text{old}} \end{pmatrix}, \quad H_{rr} = H^f + \sum_{j \in \mathcal{J}} \left((-\lambda_j + \mu c_j(r^{\text{old}}) + \mu s^{\text{old}}) H^{c_j} + \mu g^{c_j} (g^{c_j})^T \right).
$$

We note that the derivatives with respect to the slack variables s are employed explicitly to ensure that the effective dimension of the black-box problem is not unnecessarily increased. While beyond the scope of this paper, we note that it can be shown that the model m can provide a sufficiently good local approximation of h under mathematical assumptions on $f(r)$ and $c(r)$[1]. Once the augmented Lagrangian subproblem (3) is approximately solved, the multipliers λ and penalty parameter μ are updated[13]. We then proceed iteratively until the constraint violation drops below a user-set threshold.

A major advantage of this framework of separately modeling constraints and the objective is that evaluations from one iteration can be used in forming the interpolation models for subsequent iterations. Furthermore, developing separate models for the objective and constraint terms allows us to benefit from analytic derivatives where they are available (such as in the simple cost function f in the Lockwood problem).

For the cases with *some information* and *no information*, the same scheme is employed except without slack variables and two and no equality constraints, respectively.

4   NUMERICAL RESULTS

We developed a `Matlab` extension, using the model-based augmented Lagrangian framework described in the previous section, of the `POUNDERS` solver in `TAO`[12]. For the purposes of discussion, the call to `Bluebird` was modified so that double-precision (instead of single-precision) black-box output was obtained. For cases with *some information* and
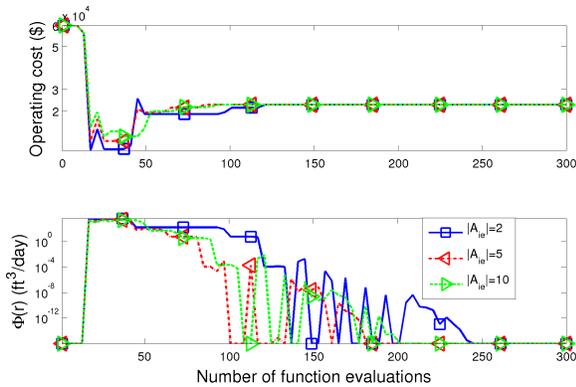
Figure 1: Evaluation trajectories of operating cost (above) and flux (below) for different levels of inexactness (*more information*).
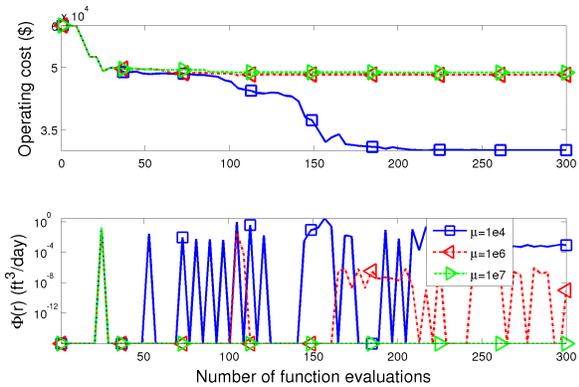
Figure 2: Evaluation trajectories of operating cost (above) and flux (below) for different penalty levels (*no information*).

more information, the structure of the objective being linear was explicitly used, and analytical derivatives were supplied. In all the numerical results presented, the constraint violation plotted was calculated by using all 201 elements specified in the original problem (and not just the possible subset in $\mathcal{A}_{ie}$ modeled by the optimization algorithm). The six pumping rates were set to 10,000 (ft^3/day) to form the standard starting point r^0.

We first focus on the problem with *more information* for different sizes of $|\mathcal{A}_{ie}|$. The values of the operating cost, $f(r)$ less a constant installation cost term, and the total flux constraint violation, $\Phi(r) = |\Phi_A(r)| + |\Phi_B(r)|$, obtained are shown as a function of the number of simulation evaluations in Fig. 1. We see that for higher levels of exactness, a solution is obtained in fewer evaluations. We note that the optimal operating cost obtained in this framework is \$22,739.67.

Figure 2 shows the trajectory for the problem with *no information* for different values of the penalty parameter, μ. For sufficiently large μ, the constraints are not violated; however, the objective function remains suboptimal because the solver is stuck at a point of nonsmoothness. For $\mu = 10^4$, the progress made by the algorithm is significantly slower than the case with *more information* shown in Fig. 1.

For all subsequent plots, we report results with $|\mathcal{A}_{ie}| = 2$ as the level of inexactness for the case with *more information* and $\mu = 10^4$ for the case with *no information*. Figures 3 and 4 compare the approaches with different levels of information for two different starting points, r^0 and r^1, where r^1 is a randomly generated point. Nonsmoothness arising from the flux constraint makes it difficult for the algorithm with *no information* to make progress beyond a certain point. The algorithm with *some information* performs better but can still be adversely affected by the nonsmoothness so that the flux appears to not drop beyond a particular value. On the other hand, the flux vanishes in the case of the algorithm with *more information* in fewer than 150 function evaluations.

Table 1 shows the overhead times for different levels of inexactness averaged across
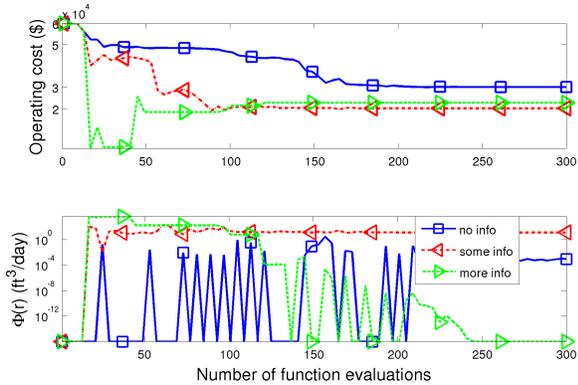
Figure 3: Trajectories of operating cost (above) and flux (below) for different levels of information for starting point r^0.
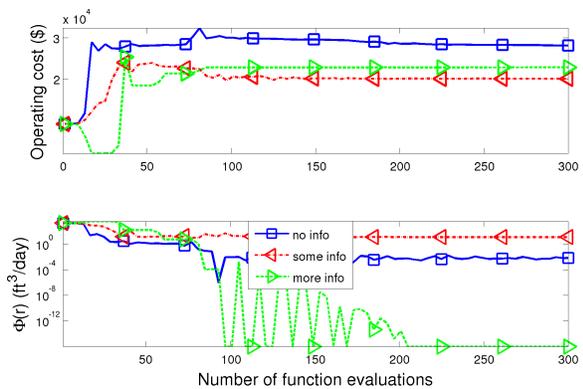

Figure 4: Trajectories of operating cost (above) and flux (below) for different levels of information for starting point r^1.

different starting points. The reported times correspond to all optimization algorithm overhead occurring between evaluations and is seen to grow faster than linearly with increasing levels of exactness. Although this overhead could be improved with a more efficient, production version of our code, the results illustrate that using a high level of inexactness for the Lockwood problem (where evaluation times ranged from 3 to 5 seconds) may not reduce the total time to solution. On the other hand, for more expensive simulations, the additional cost of overhead may be negligible relative to the savings in simulation time.

Some information	$|\mathcal{A}_{\mathrm{ie}}| = 2$	$|\mathcal{A}_{\mathrm{ie}}| = 3$	$|\mathcal{A}_{\mathrm{ie}}| = 5$	$|\mathcal{A}_{\mathrm{ie}}| = 8$	$|\mathcal{A}_{\mathrm{ie}}| = 10$
0.17s	0.57s	0.72s	1.10s	2.27s	3.48s

Table 1: Average computational overhead time per evaluation versus levels of information/inexactness.

5   CONCLUSIONS

This paper has illustrated how model-based derivative-free optimization algorithms can solve groundwater problems in fewer simulation evaluations if the algorithm has access to a greater level of information. On the Lockwood problem, our results show that this is the case even though full derivative information remains unavailable, with performance gains resulting simply by granting an algorithm access to both objective *and* constraint values. We have shown that reformulations with smooth, black-box constraints can overcome hurdles to obtaining high-quality solutions often found with penalized problems without any explicit nonsmooth structure. The cost of reducing the number of simulations is greater computational overhead. Hence, the level of inexactness and information employed should be chosen based on the expense of the underlying simulation and the degree of constraint violation permitted and the solution quality demanded.

ACKNOWLEDGMENTS

References

[1] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. SIAM, Philadelphia, PA, 2009.

[2] J. R. Craig. Bluebird developers manual. Technical report, Department of Civil, Structural, and Environmental Engineering, University at Buffalo, Buffalo, NY, 2002.

[3] K. R. Fowler, C. T. Kelley, C. T. Miller, C. E. Kees, R. W. Darwin, J. P. Reese, M. W. Farthing, and M. S. C. Reed. Solution of a well-field design problem with implicit filtering. *Optim. Eng.*, 5 (2):207–234, 2004.

[4] K. Fowler et al. Comparison of derivative-free optimization methods for groundwater supply and hydraulic capture community problems. *Adv. Water Resour.*, 31(5):743–757, 2008.

[5] I. Janković and R. Barnes. High-order line elements in modeling two-dimensional groundwater flow. *J. Hydrol.*, 226:211–223, 1999.

[6] C. T. Kelley. *Implicit Filtering*. SIAM, Philadelphia, PA, 2011.

[7] T. G. Kolda, R. M. Lewis, and V. J. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Rev.*, 45(3):385–482, 2003.

[8] L. S. Matott, A. J. Rabideau, and J. R. Craig. Pump-and-treat optimization using analytic element method flow models. *Adv. Water Resour.*, 29(5):760–775, 2006.

[9] L. S. Matott, K. Leung, and J. Sim. Application of MATLAB and Python optimizers to two case studies involving groundwater flow and contaminant transport modeling. *Comput. Geosci.*, 37(11): 1894–1899, 2011.

[10] A. S. Mayer, C. Kelley, and C. T. Miller. Optimal design for problems involving flow and transport phenomena in saturated subsurface systems. *Adv. Water Resour.*, 25(8-12):1233–1256, 2002.

[11] J. J. Moré and S. M. Wild. Benchmarking derivative-free optimization algorithms. *SIAM J. Optimiz.*, 20(1):172–191, 2009.

[12] T. Munson, J. Sarich, S. M. Wild, S. Benson, and L. C. McInnes. TAO 2.0 users manual. Technical Memorandum ANL/MCS-TM-322, MCS Division, Argonne National Laboratory, 2012.

[13] J. Nocedal and S. J. Wright. *Numerical optimization*. SIAM, Philadephia, PA, second edition, 2006.

[14] S. M. Wild. MNH: A derivative-free optimization algorithm using minimal norm Hessians. In *Tenth Copper Mountain Conf. On Iterative Methods*, April 2008.

[15] S. M. Wild, R. G. Regis, and C. A. Shoemaker. ORBIT: Optimization by radial basis function interpolation in trust-regions. *SIAM J. Sci. Comput.*, 30(6):3197–3219, 2008.