Sequential Design for Constraint Satisfaction in Optimization

Robert B. Gramacy
The University of Chicago Booth School of Business
faculty.chicagobooth.edu/robert.gramacy

with H.K.H. Lee, S. Wild, G.A. Gray, S. Le Digabel

DAE Athens, GA — October 2012

General constrained optimization

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad \text{s.t.,} \;\; c(\mathbf{x}) \leq 0,$$

where f and c may be blackboxes.

Except in a few special cases, we don't have statistical tools to solve this problem.

- $c(\mathbf{x})$ is linear, you can use EI (Jones, et al. 1998)
- $c(\mathbf{x}) \in \{0, 1\}$, you can try IECI (G. & Lee, 2011)

That's a shame because stats methods have lots to offer:

- global solutions/robustess/UQ
- a monopoly on methods for noisy simulator evaluations

In fact, in many real constrained optimization problems are easier, but we still don't have solutions:

- the objective f may be known and linear
- the hard part is the constraint function

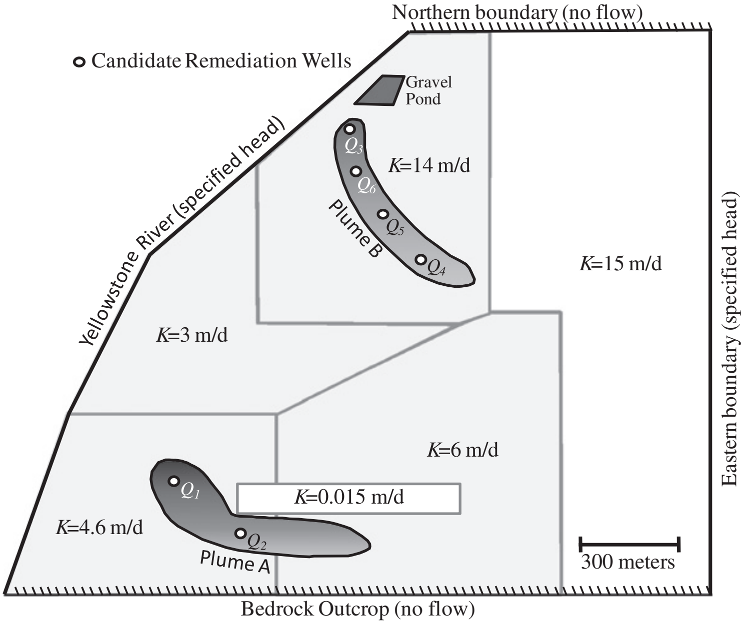
Examples include any problem where resource costs can be summed, but one cannot know whether the allocation is sufficient without expensive simulation/experimentation.

- Goal: tackle problems like these while retaining benefits of statistical optimization.

Motivation: pump-and-treat

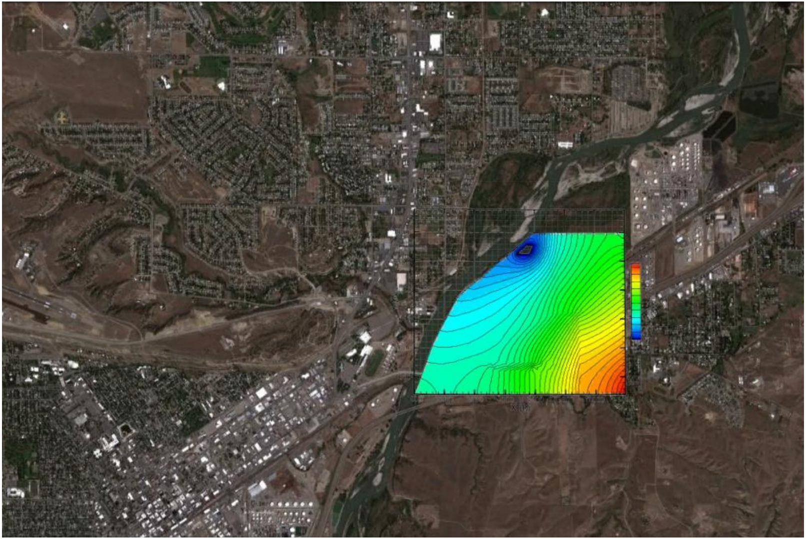
A “hypothetical” groundwater contamination scenario based on the Lockwood Solvent Groundwater Plume Site located near Billings, Montana (Tetra Tech Inc., 2003)

- Industrial practices have resulted in the development of two separate plumes containing chlorinated solvents that threaten the Yellowstone river.
- Six pump-and-treat wells have been proposed to prevent further expansion.
- The optimization objective is to contain both plumes using the minimum amount of pumping.



Northern boundary (no flow)
Candidate Remediation Wells
Gravel Pond
Yellowstone River (specified head)
Q_3
Q_6
Q_5
Q_4
Plume B
K=14 m/d
K=15 m/d
K=3 m/d
Eastern boundary (specified head)
K=6 m/d
Q_1
K=0.015 m/d
K=4.6 m/d
Q_2
Plume A
300 meters
Bedrock Outcrop (no flow)

Pressure gradients



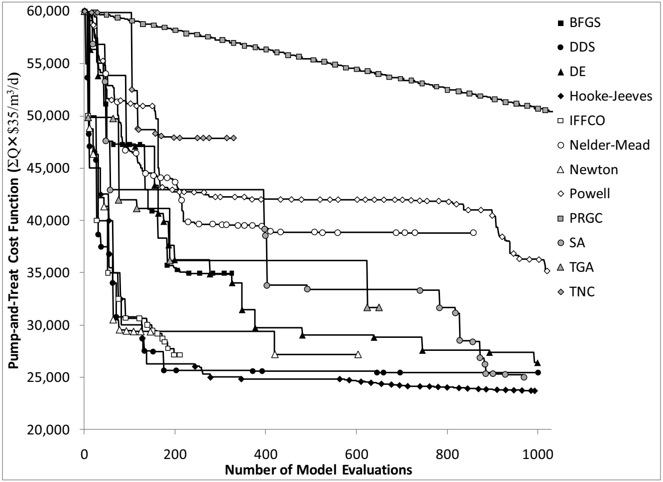
An analytic element method (AEM) groundwater model is used to simulate the amount of contaminant exiting both boundaries under pumping scenarios (Craig & Matott, 2005).

So the constrained optimization problem is:

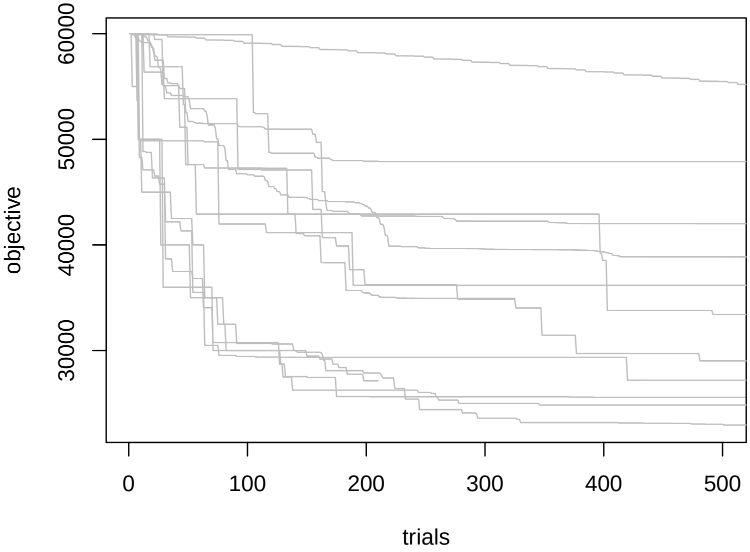
$$\min_{0 \leq x_j \leq 2 \times 10^5} \sum_{j=1}^{6} x_j, \qquad \text{s.t.} \;\; c_{1,2}(\mathbf{x}) \leq 0,$$

- a linear objective (pumping rates), and
- two expensive (to evaluate) quantified constraints on the amount of contaminant existing the system.
  - a large highly non-convex satisfaction region; very narrow in the neiighborhood of the optimum

MATLAB and Python optimizers usign an additive penalty method (APM; Matott, et al, 2011; Hilton and Culver, 2000)



New best is MADS, from S. Le Digabel, et al.



A baseline

Here is a strategy that leverages the simplicity of the underlying objective:

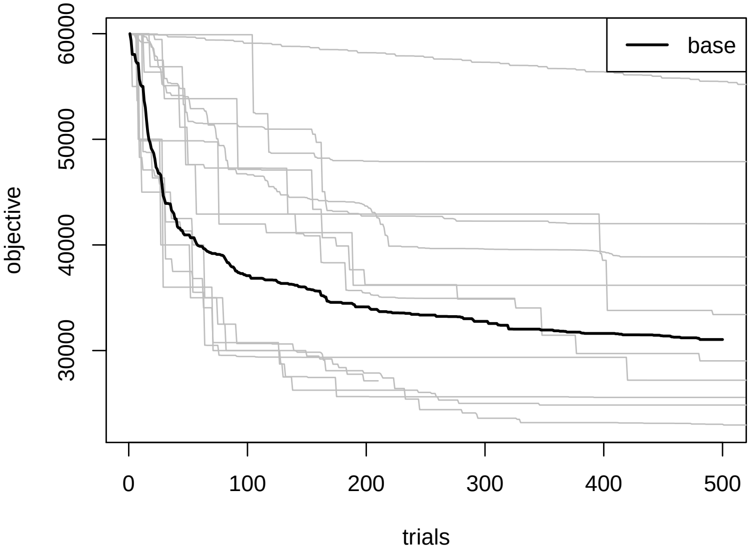
For $n = n_{\text{start}}, \ldots$ (i.e., at each trial), do

- Let $y_n^* = \min_{i=1,\ldots,n}\{\sum_j x_{ij} : c_{1,2}(\mathbf{x}_i) = 0\}$
- Choose $\mathbf{x}_{n+1} \sim \mathrm{Unif}([0, 2 \times 10^5]^6)$ s.t. $\sum_j x_{n+1,j} < y_n^*$

A good starting point is $\mathbf{x} = (1 \times 10^5)^6$, which is valid.

- No modeling or other calculation required,
- except maybe a rejection sampling routine.

Sampling only from the objective improvement region.



Finding the constraint boundary

Due to the linearity of the objective, the solution must lie on the constraint boundary.

At each trial, fit a classification model to data pairs comprised of $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ where

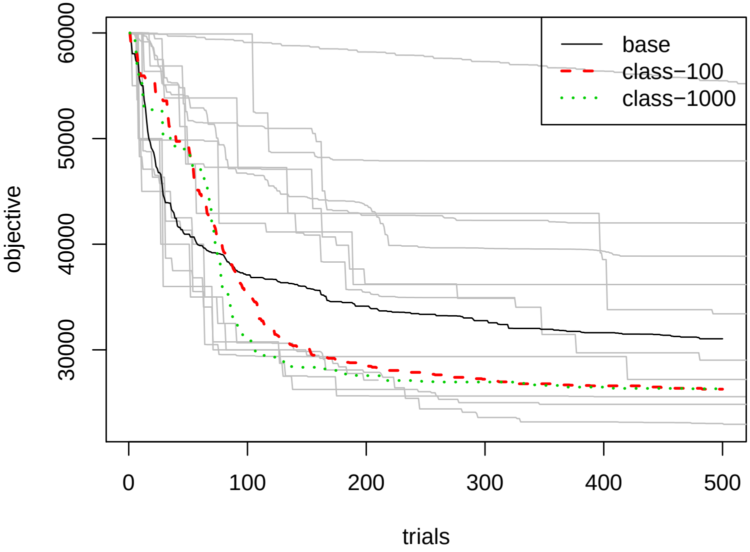
$$y_i = \mathbb{I}(c_1(\mathbf{x}_i) = 0) \times \mathbb{I}(c_2(\mathbf{x}_i) = 0).$$

- we use a classification GP (CGP) via `plgp` package for R

Among candidates improving on the objective, choose $\mathbf{x}_{n+1}$ via an active learning heuristic

- we use predictive class-entropy

CGP/entropy active learning.



Pros:

- Global.
- Uses important analytic information (about the objective)
- Make a use of rich constraint information (by forecasting)

Cons:

- Does not use constraint quantification information.
- Entropy is a poor active learning heuristic—too myopic, although choosing from the improvement region helps.

Augmented Lagrangian

The *augmented Lagrangian* is a penalty function often used in optimization contexts due to favorable asymptotic properties.

It offers an amenable framework for balancing a scalar objective f and vector-valued constraints c:

$$L_A^{\lambda,\mu}(\mathbf{x}) = f(\mathbf{x}) + \lambda^\top c(\mathbf{x}) + \mu^\top c(\mathbf{x})^2$$

see, e.g., Kannan & Wild (2012).

Optimizing $L_A^{\lambda,\mu}(\mathbf{x})$ by unconstrained methods leads to good *intermediate* solutions.

When the tuning parameters λ, μ follow particular updating rules, e.g.,

$$\lambda' = \lambda + \mu^\top c(x^*), \qquad \mu' = \rho\mu, \text{ for } \rho > 1$$

implementing heavier penalization of constraint violations in the composite objective as the algorithm progresses

- convergence is guaranteed to a local, *valid*, solution under very mild conditions.

- A little like simulated annealing but not guarenteeing a global solution.

Expected Improvement (EI) offers global scope and the ability to accomodate noise.

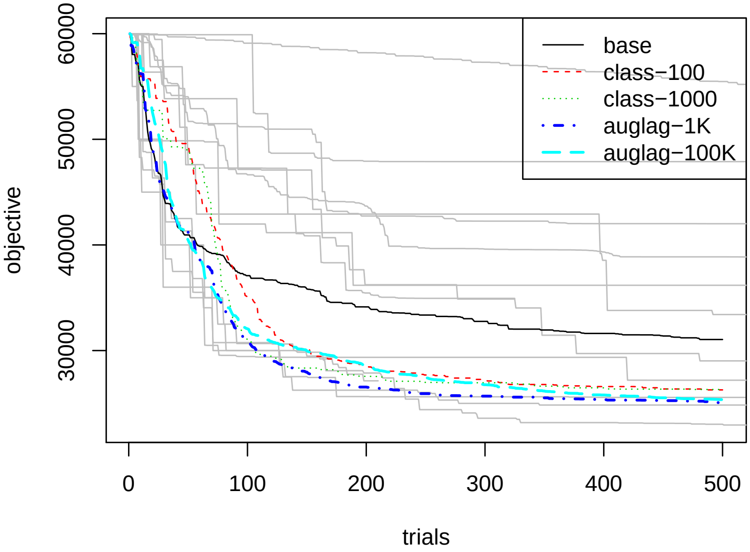
In each trial:

- Let $y(\mathbf{x}) = L_A^{\lambda,\mu}(\mathbf{x})$, and fit a surrogate model to $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$.
- Let $y^* = \min\{y_1, \ldots, y_n\}$ and define

$$I(\mathbf{x}) = \max\{y^* - Y(\mathbf{x}), 0\}, \quad \text{for surrogate} \quad Y(\mathbf{x}).$$

- Choose $\mathbf{x}_{n+1}$ to maximize $\mathbb{E}\{I(\mathbf{x})\}$ using candidates $\mathbf{x}$, possibly only those improving on the objective;
- update λ and μ at convergence and restart with new y.

EI on augmented Lagrangian objective with GP surrogate.



Pros:

- Uses quantified constraint information.
- Easily extended to non-trivial objective functions.*

Cons:

- Does not leverage that the solution is on the constraint boundary.
- EI behaves strangely: global scope is hurting.
- Composite objective is pathologically non-stationary, which makes choosing surrogate modeling hard.
- Unnecessarily models a known objective.*

One thought is to use a non-stationary surrogate model, like `tgp`.

- that helps a little, but it only addresses one of the cons.

A better idea may be to separately model each of the components of $L_A^{\lambda,\mu}(\mathbf{x})$: f, $c_1, \ldots, c_m$.

- In many cases, surrogate $Y_f(\mathbf{x})$ may not be needed for f.
- GPs can be used for each c_i, producing surrogates $Y_{c_i}(\mathbf{x})$.
- Then $Y(\mathbf{x}) = Y_f(\mathbf{x}) + \lambda^\top Y_c(\mathbf{x}) + \mu^\top Y_c(\mathbf{x})^2$ is a surrogate for $L_A^{\lambda,\mu}(\mathbf{x})$.

Unfortunately, the composite $Y(\mathbf{x})$ does not easily emit an analytic EI statistic.

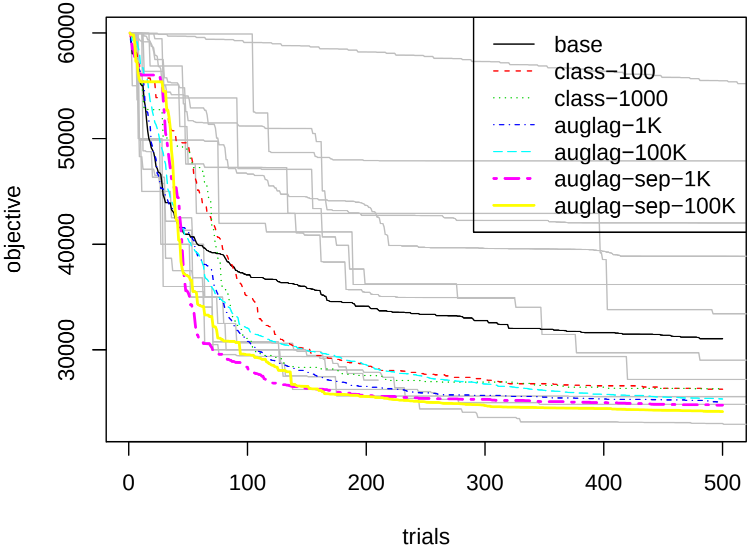
- We're working on it.
- Numerical evaluation is harder than you might think.

But it can still be useful, since $\mathbb{E}\{Y(\mathbf{x})\}$ is analytic.

For each trial

- choose $\mathbf{x}_{n+1}$ to maximize $\mathbb{E}\{Y(\mathbf{x})\}$ using candidates $\mathbf{x}$, possibly only those improving on the objective, f
- update λ and μ at convergence and restart with new Y_f, and Y_{c_i}

Separated augmented Lagrangian surrogate model(s).



Pros:

- Uses quantified constraint information.
- Easily extended to non-trivial objective functions.

Cons:

- Without EI, no loger global in scope.
- Does not leverage that the solution is on the constraint boundary.
- GP surrogate models $Y_{c_i}(\mathbf{x})$ may find the limiting “kink” at zero difficult to emulate.

Wrapping up

EI and related methods offer very compelling solutions to hard (global) optimization problems.

But not ones that are often encountered in practice.

- We are lacking good ideas for dealing with constraints.

This talk has shown that the integration of several small ideas can add up in a big way.

More synergies are need:

- new surrogate modeling ideas
- new improvement heuristics combining quantifiable constraints and analytic structure