THE STATISTICAL FILTER APPROACH TO CONSTRAINED OPTIMIZATION

Tony Pourmohamad
Herbie Lee
Genentech and U.C. Santa Cruz

2021 FTC Webinar Series

Winner of the 2021 Wilcoxon Award

Taylor & Francis
Taylor & Francis Group

Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

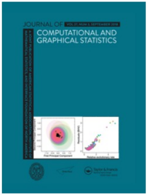
The Statistical Filter Approach to Constrained Optimization

Tony Pourmohamad & Herbert K. H. Lee

To cite this article: Tony Pourmohamad & Herbert K. H. Lee (2020) The Statistical Filter Approach to Constrained Optimization, Technometrics, 62:3, 303-312, DOI: 10.1080/00401706.2019.1638304

To link to this article: https://doi.org/10.1080/00401706.2019.1638304

More on Optimization



Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

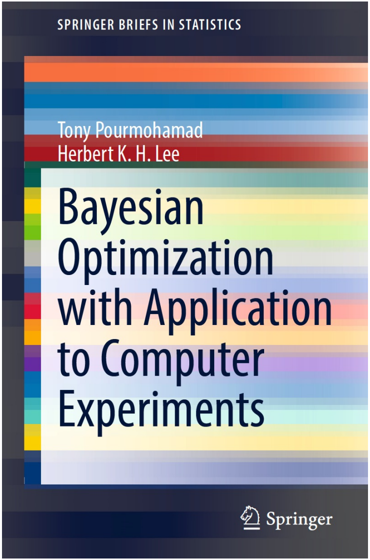
Bayesian Optimization Via Barrier Functions

Tony Pourmohamad & Herbert K. H. Lee

To cite this article: Tony Pourmohamad & Herbert K. H. Lee (2021): Bayesian Optimization Via Barrier Functions, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2021.1935270

To link to this article: https://doi.org/10.1080/10618600.2021.1935270

More on Optimization



MOTIVATION

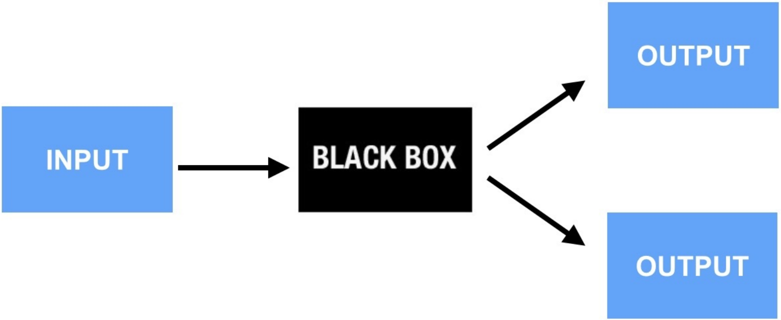
Motivation

- Computer model
- Black box system

- Expensive evaluations

Motivation

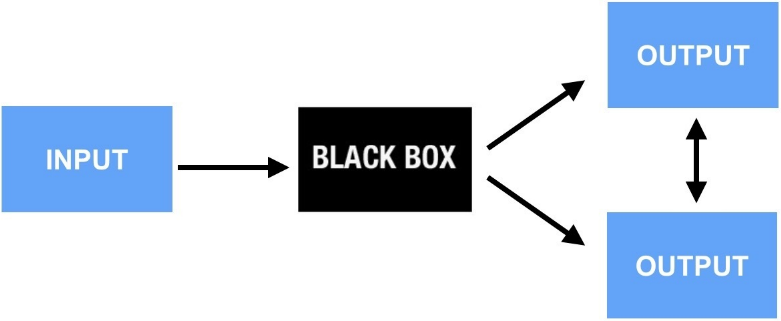
- Computer model
- Black box system



- Expensive evaluations

Motivation

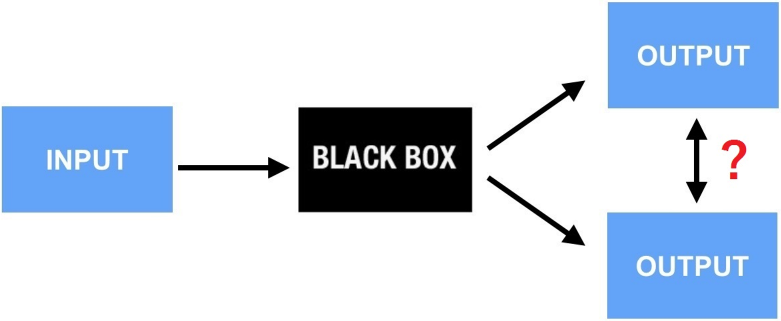
- Computer model
- Black box system



- Expensive evaluations

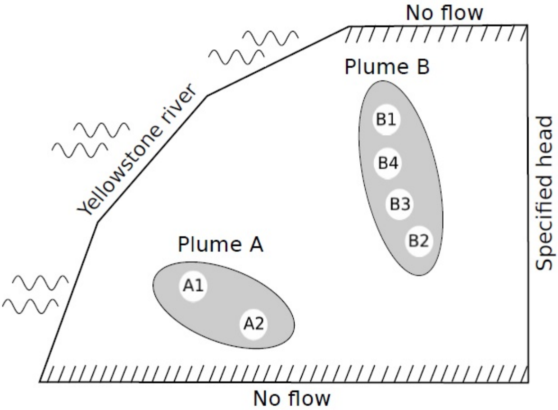
Motivation

- Computer model
- Black box system

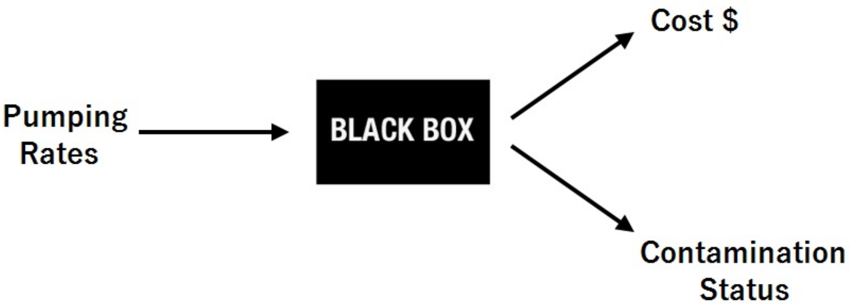


- Expensive evaluations

Example: The Lockwood Problem



Example: The Lockwood Problem



Optimization of Computer Models

We want to

$$\begin{aligned} \min_{x} \quad & f(x) \\ s.t. \quad & c(x) \leq 0 \\ & x \in \mathcal{X} \end{aligned}$$

- $\mathcal{X} \subset \mathbb{R}^d$ is a known and bounded region
- $f : \mathcal{X} \to \mathbb{R}$ denotes a scalar-valued objective function
- $c : \mathcal{X} \to \mathbb{R}^m$ denotes a vector of constraint functions

- f and c are the outputs from running the computer model
- x is the input to the computer model

Motivation

Problems

- Need to be able to model $f(x)$ and $c(x)$ cheaply
- Want to be able to model correlation between $f(x)$ and $c(x)$
- Need an efficient optimization algorithm that can handle the case of black-box derivative free optimization

Solutions

- We build a multivariate Gaussian process model for $f(x)$ and $c(x)$
- Sequential Monte Carlo will be used to speed up computations
- We combine filter methods with statistical surrogate modeling to leverage optimal properties of both

SURROGATE MODEL

Gaussian Process

Definition

For any index set $\mathcal{X}$, the real-valued stochastic process $\{Y(\mathbf{x}), \mathbf{x} \in \mathcal{X}\}$, is a Gaussian process if all the finite-dimensional distributions, say, $F(\mathbf{x}_1, ..., \mathbf{x}_n)$, are multivariate normal distributions, for any choice of $n \geq 1$ and $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathcal{X}$.

- A Gaussian process thus requires a specification of a mean function, $m(\mathbf{x})$, and a covariance function $C(\mathbf{x}, \mathbf{x}')$
- Gaussian processes are distributions over functions, i.e.,

$$Y(\mathbf{x}) \sim GP(m(\mathbf{x}),\, C(\mathbf{x}, \mathbf{x}'))$$

Why use a Gaussian Process?

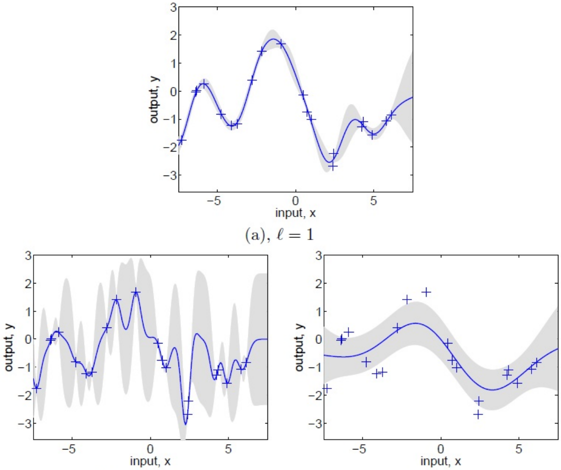
- The canonical choice for modeling computer experiments
- GP's are much cheaper/faster to evaluate than actual computer model code
- GP's allow for uncertainty quantification of outputs at untried (or unobserved) time points, i.e.,

$$Y(\mathbf{x}^*)|Y(\mathbf{x}) \sim GP(\tilde{m}, \tilde{C})$$

where

$$\tilde{m} = m(\mathbf{x}^*) + C(\mathbf{x}, \mathbf{x}^*)^T C(\mathbf{x}, \mathbf{x})^{-1}(Y(\mathbf{x}) - m(\mathbf{x}))$$
$$\tilde{C} = C(\mathbf{x}^*, \mathbf{x}^*) - C(\mathbf{x}, \mathbf{x}^*)^T C(\mathbf{x}, \mathbf{x})^{-1} C(\mathbf{x}, \mathbf{x}^*)$$

Gaussian Process Prediction

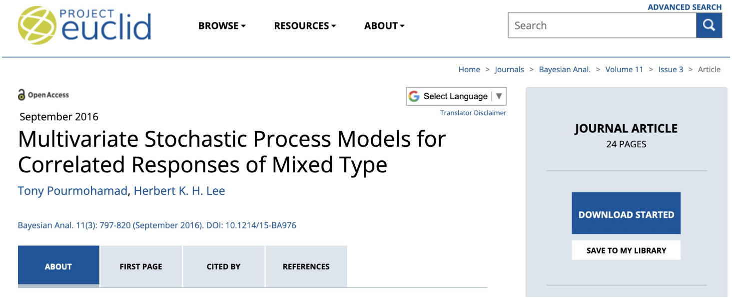


(a), $\ell = 1$

(b), $\ell = 0.3$

(c), $\ell = 3$

Multivariate Gaussian Process



- Capable of modeling correlated outputs
- Fast sequential Monte Carlo inference

STATISTICAL FILTER

Solvers

Mathematical programming has efficient algorithms for non-linear (black-box) optimization (under constraints) with

- provable local convergence properties
- lots of polished open source software

Statistical approaches, e.g., expected improvement (Jones et al., 1998)

- enjoy global convergence properties
- excel when simulation is expensive, noisy, non-convex

... but offer limited support for constraints

Filter Methods

Introduced by Fletcher and Leyffer (2002) to solve nonlinear programming problems without the use of a penalty function

A filter is like a Pareto front

Let $h(x) = ||\max\{0, c(x)\}||_1 = \sum \max\{0, c_i(x)\}$. Then the filter $\mathcal{F}$ is defined as the set of all points $x \in \mathcal{X}$ such that there does not exist a point $x' \in \mathcal{X}$ satisfying all of:

1. $f(x') \leq f(x)$
2. $h(x') \leq h(x)$
3. $(h(x'), f(x')) \neq (h(x), f(x))$

Reformulated Optimization Problem

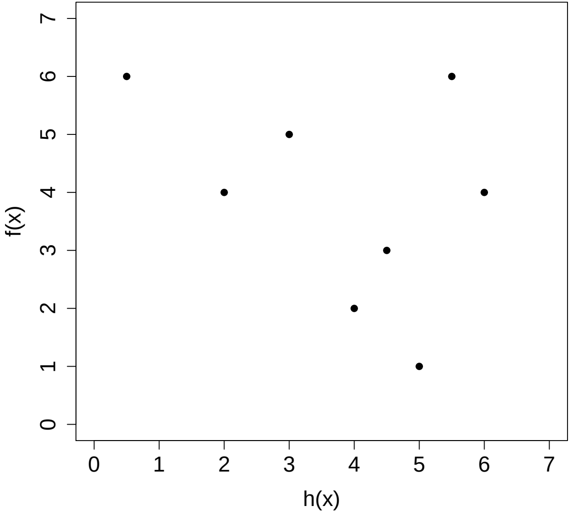
We want to

$$\begin{aligned} \min_x \quad & f(x) \\ s.t. \quad & h(x) = 0 \\ & x \in \mathcal{X} \end{aligned}$$

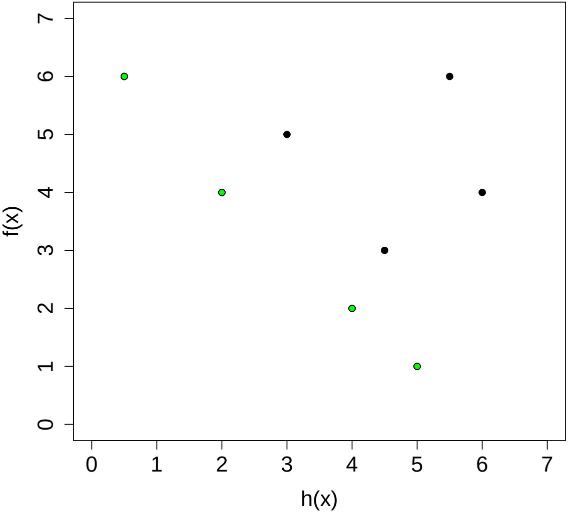
- $\mathcal{X} \subset \mathbb{R}^d$ is a known and bounded region
- $f : \mathcal{X} \to \mathbb{R}$ denotes a scalar-valued objective function
- $h : \mathcal{X} \to \mathbb{R}$ denotes a scalar-valued feasibility fuction

- We have replaced the constraint function $c(x)$ now with the feasibility function $h(x)$

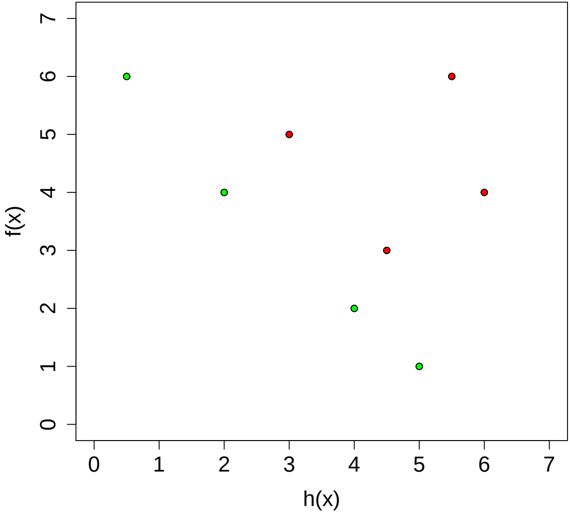
The Filter – Visually



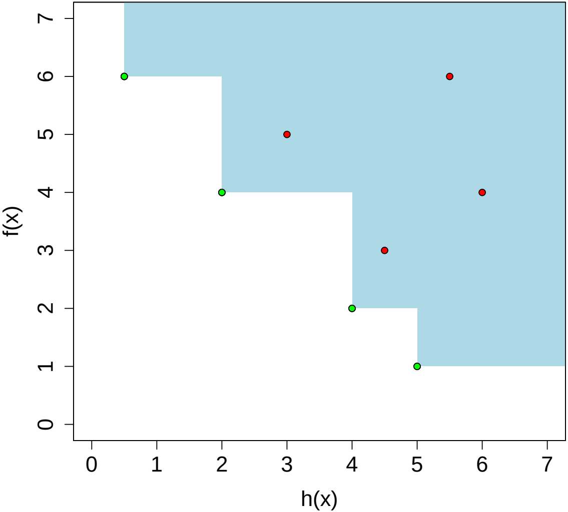
The Filter – Visually



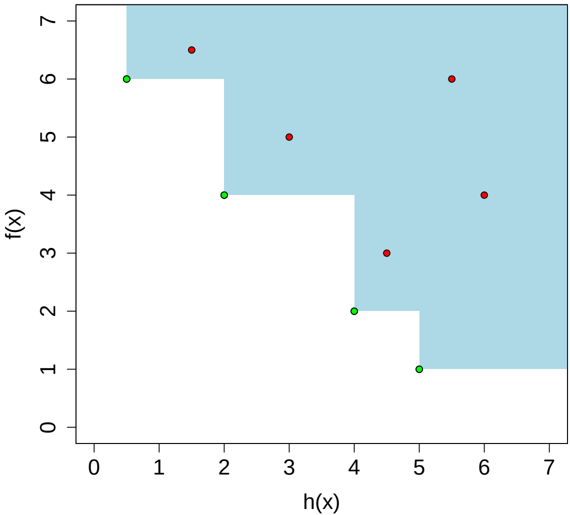
The Filter – Visually



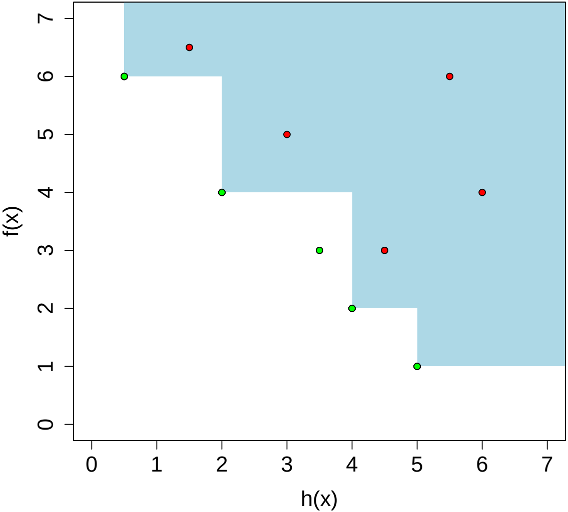
The Filter – Visually



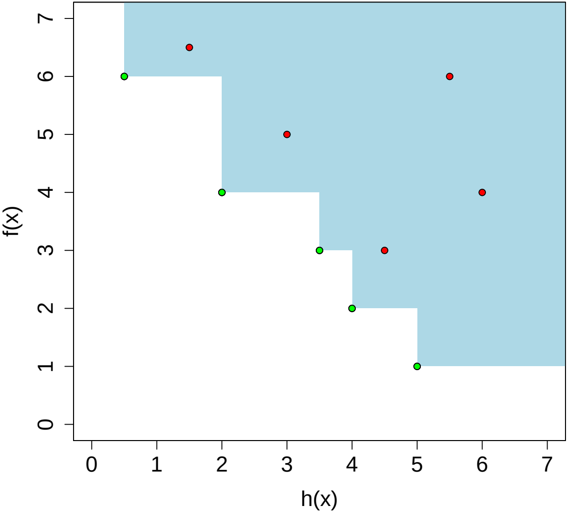
The Filter – Visually



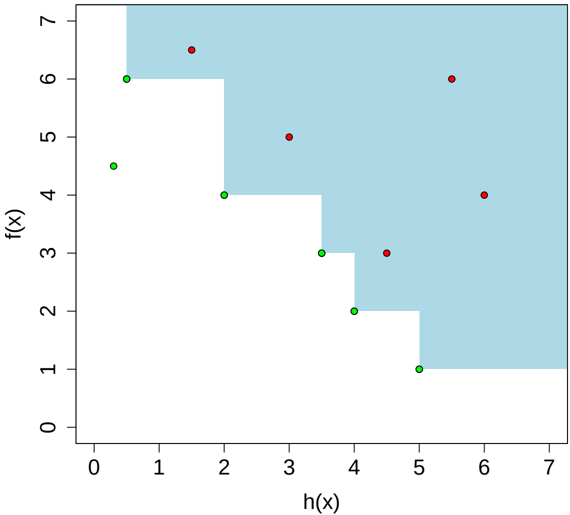
The Filter – Visually



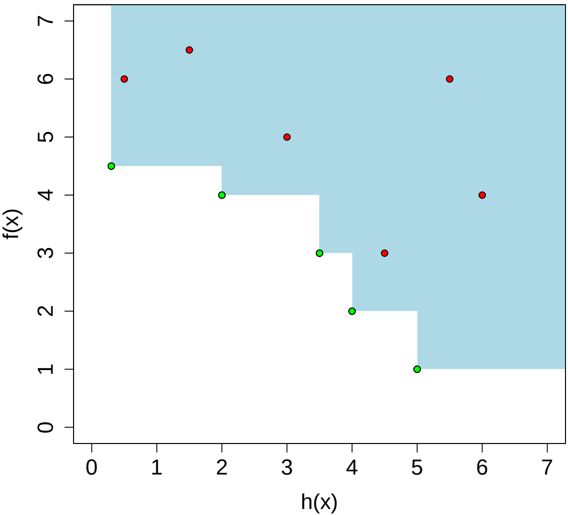
The Filter – Visually



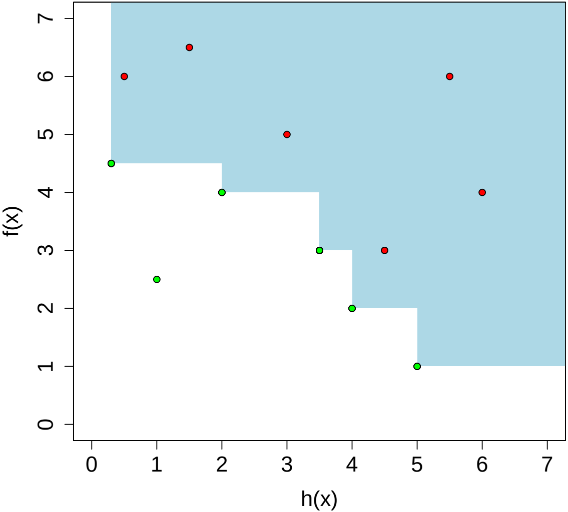
The Filter – Visually



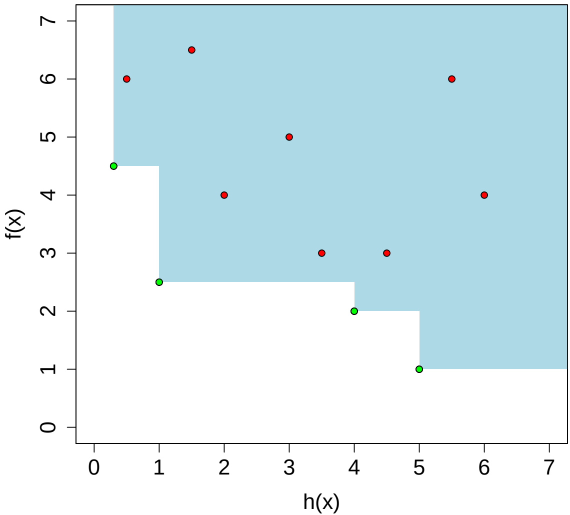
The Filter – Visually



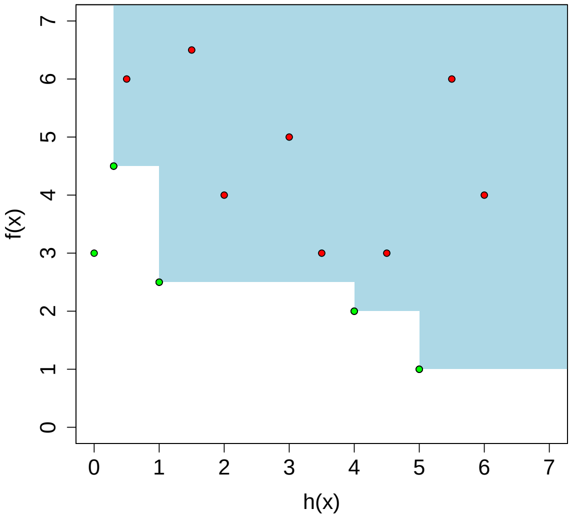
The Filter – Visually



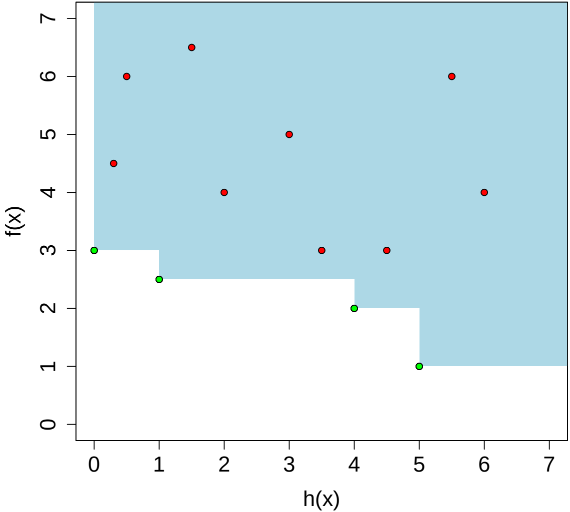
The Filter – Visually



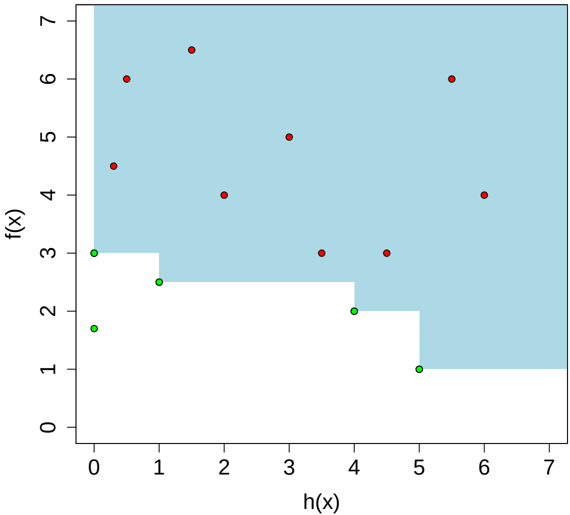
The Filter – Visually



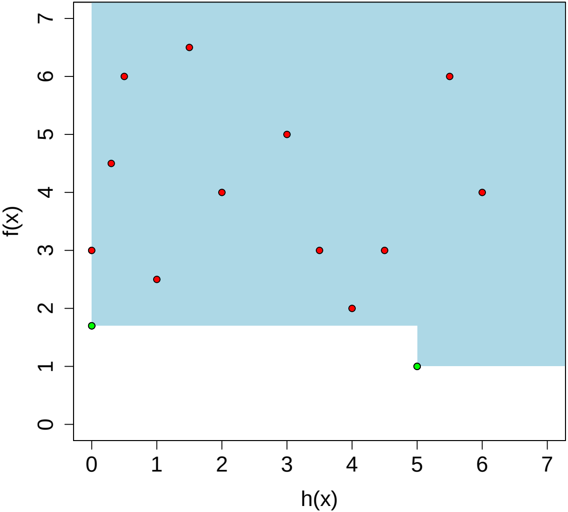
The Filter – Visually



The Filter – Visually



The Filter – Visually



The Filter Algorithm

1. Initialize the filter $\mathcal{F}$
2. While not terminated do
   a Obtain a candidate point x_*
   b Evaluate $f(x_*)$ and $c(x_*)$
   c If $(h(x_*), f(x_*))$ is acceptable to $\mathcal{F}$, then
      i Add $(h(x_*), f(x_*))$ to $\mathcal{F}$
      ii Remove any points in $\mathcal{F}$ dominated by $(h(x_*), f(x_*))$
   d Check for termination

Extra Conditions

This approach is provably convergent to a local mode. However, care must be taken to avoid convergence to an infeasible point $(h(x) > 0)$ or a local minimum.

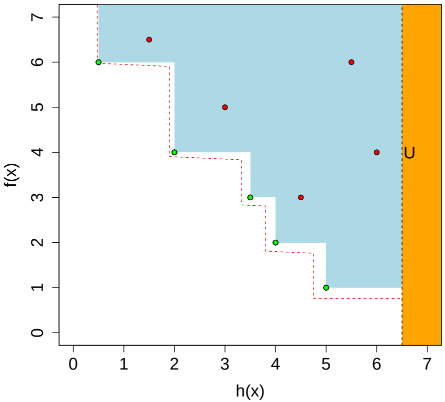
Two enhancements to improve convergence to a global minimum:

- Use an envelope

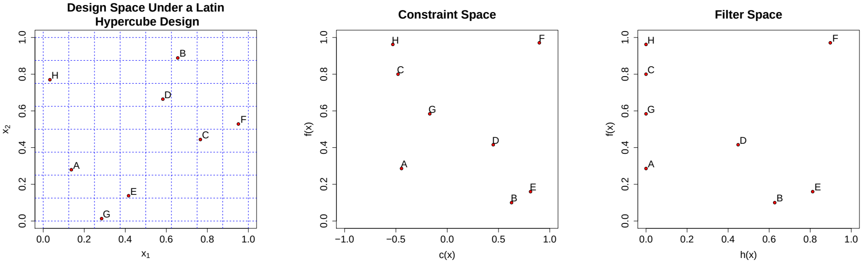
$$h(x_*) \leq \beta h(x_i) \text{ or } f(x_*) \leq f(x_i) - \gamma h(x_*)$$
$$\forall (h(x_i), f(x_i)) \in \mathcal{F}$$

  for $\beta, \gamma \in (0, 1)$
- Upper bound U on the acceptable constraint violation

Extra Conditions



Spaces of Interest



The Statistical Filter

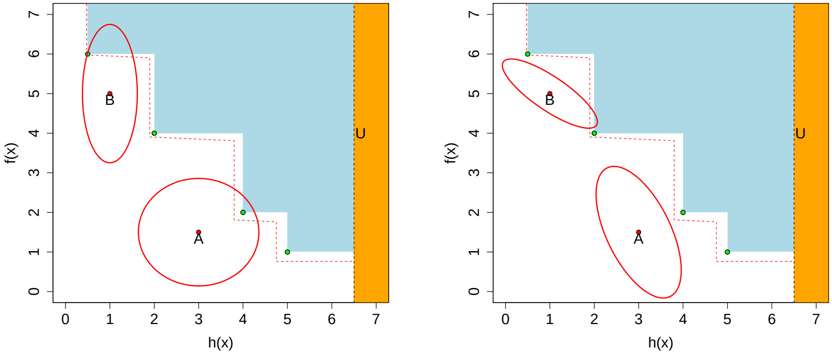
Combine filter method with Gaussian process surrogate modeling

Surrogate models operate in the constraint space

Choose the next candidate point as the one that maximizes the probability it will be acceptable to the filter, i.e.,

$$x_* = \max_{x \in \mathcal{X}} \text{Pr}\{(h(x), f(x)) \text{ is acceptable to the filter } \mathcal{F}\}$$

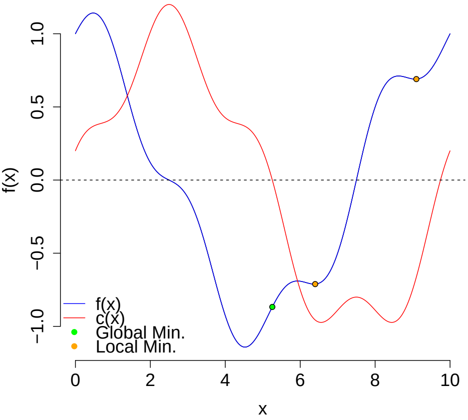
Probability Acceptable to the Filter



The Statistical Filter Algorithm

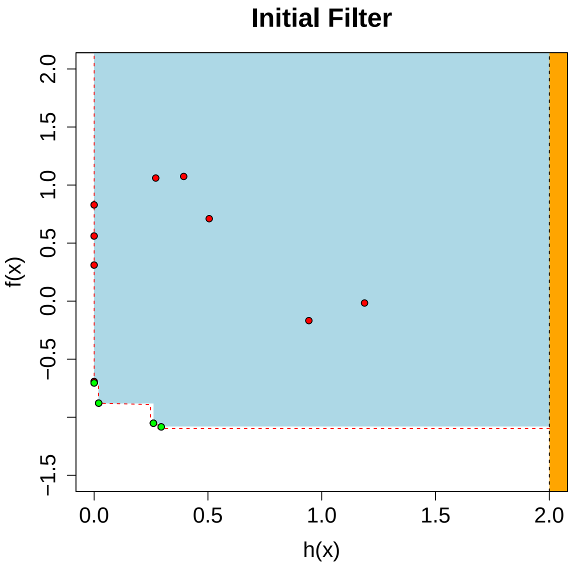
1. Sample initial inputs from a LHD
2. Initialize the filter $\mathcal{F}$
3. While not terminated do
   a Fit a surrogate model for $f(x)$ and $c(x)$
   b Map the surrogate model in the constraint space to the filter space
   c Obtain a candidate point x_* that maximizes the probability acceptable to the filter
   d Evaluate $f(x_*)$ and $c(x_*)$
   e If $(h(x_*), f(x_*))$ is acceptable to $\mathcal{F}$, then
      i Add $(h(x_*), f(x_*))$ to $\mathcal{F}$
      ii Remove any points in $\mathcal{F}$ dominated by $(h(x_*), f(x_*))$
   f Check for termination
   g Sample new candidate inputs from a LHD

Illustrating Example

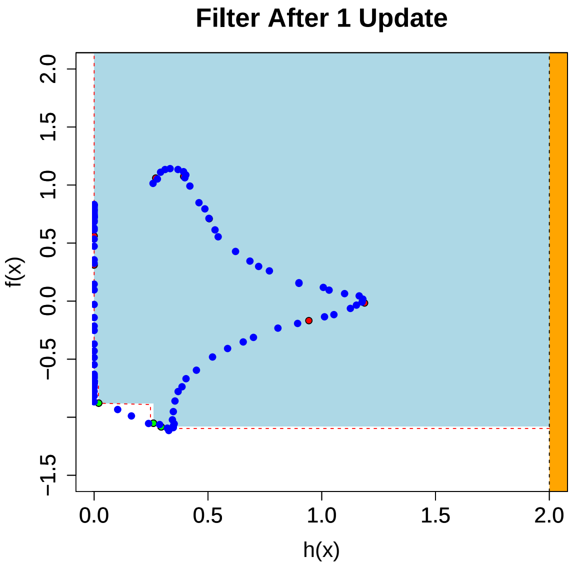


- $f(x) = \cos\left(\frac{\pi x}{5}\right) + 0.2\sin\left(\frac{4\pi x}{5}\right)$
- $c(x) = \sin\left(\frac{\pi x}{5}\right) + 0.2\cos\left(\frac{4\pi x}{5}\right)$
- Global minimum on the constraint boundary
- Two local minima in the feasible region

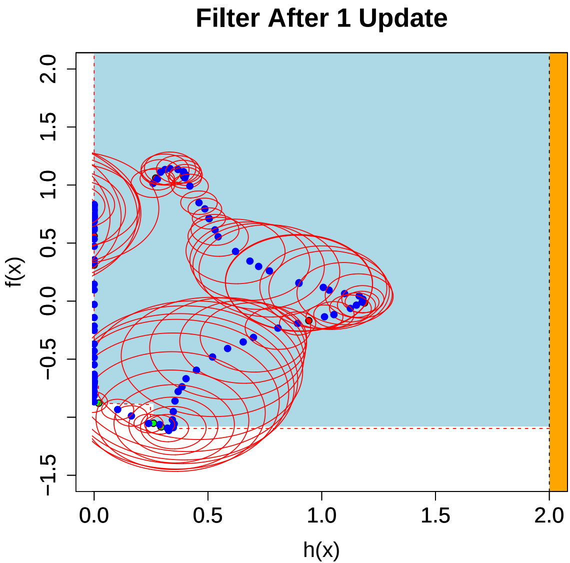
Statistical Filter Solution



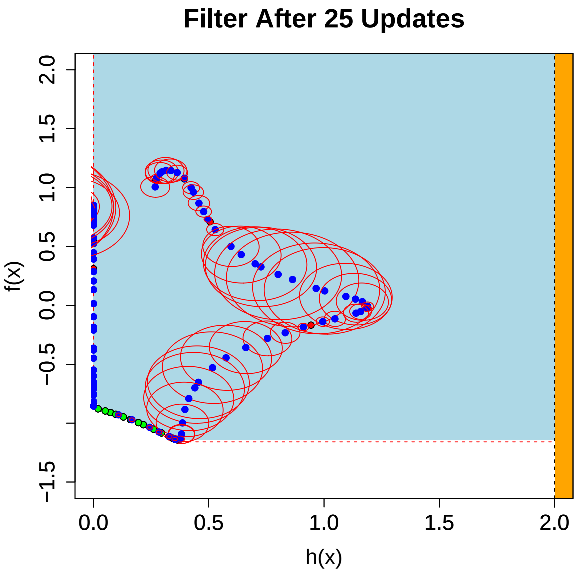
Statistical Filter Solution



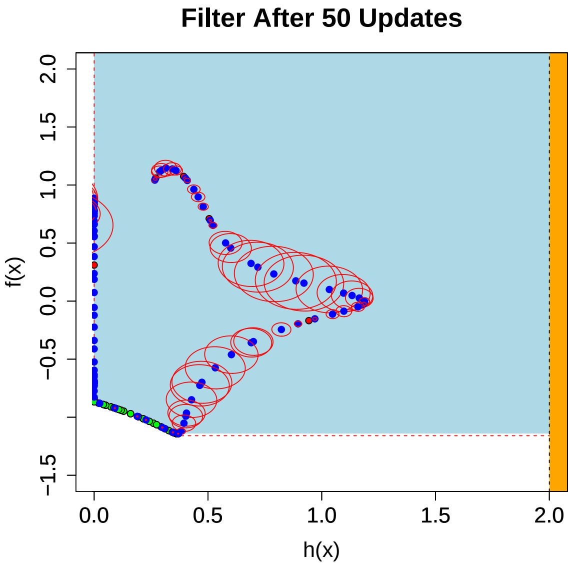
Statistical Filter Solution



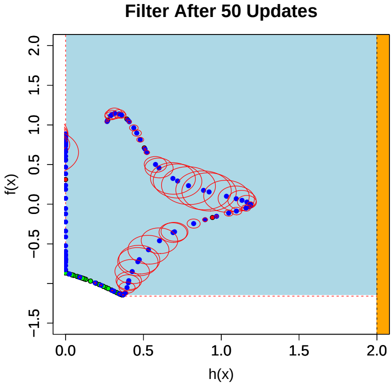
Statistical Filter Solution

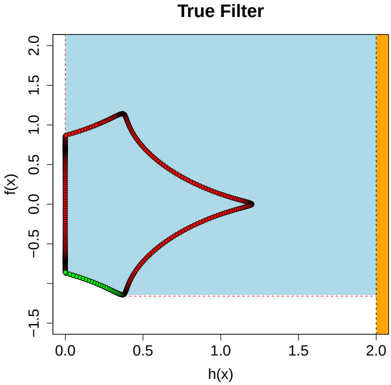


Statistical Filter Solution



Statistical Filter Solution





- $f(x^*) = -0.86718$
- $f(x) = -0.86718$

RESULTS

The Toy Problem

Toy problem from Gramacy et al. 2016

- A linear objective in two variables

$$\min_x \{x_1 + x_2 : c_1(x) \leq 0, c_2(x) \leq 0, x \in [0,1]^2\}$$

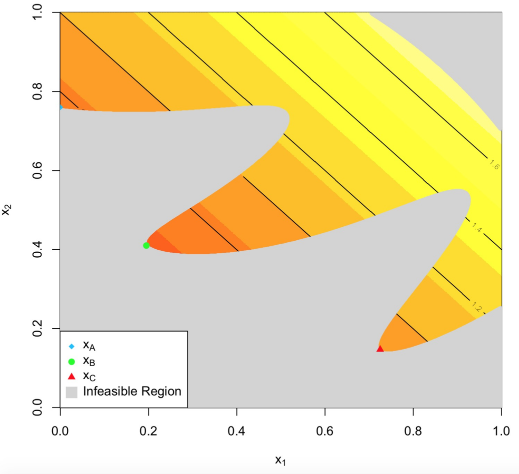
- where two non-linear constraints are given by

$$c_1(x) = \frac{3}{2} - x_1 - 2x_2 - \frac{1}{2}\sin(2\pi(x_1^2 - 2x_2))$$
$$c_2(x) = x_1^2 + x_2^2 - \frac{3}{2}$$

Even when treating $f(x) = x_1 + x_2$ as known, this is a hard problem when $c(x)$ is treated as a black-box

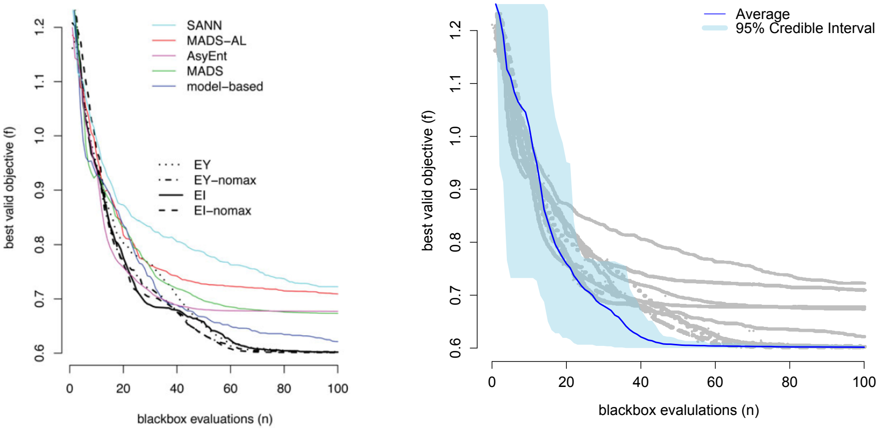
The Toy Problem

$$
\begin{aligned}
x_A &\approx [0.1954,\ 0.4044],\\
f(x_A) &\approx 0.5998,\\
x_B &\approx [0.7197,\ 0.1411],\\
f(x_B) &\approx 0.8609,\\
x_C &= [0,\ 0.75],\\
f(x_C) &= 0.75,
\end{aligned}
$$

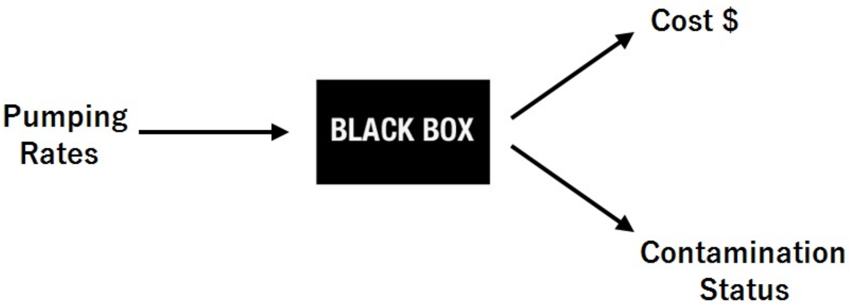


- $c_2(x)$ may seem uninteresting, but it reminds us that solutions may not exist on every boundary

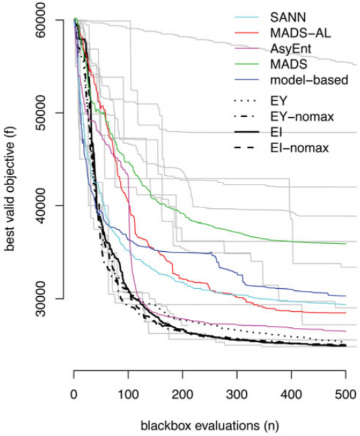
Results on Toy Problem

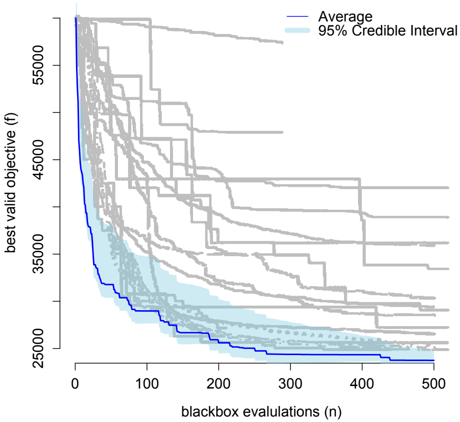


Pump-and-Treat Hydrology Problem



Pump-and-Treat Hydrology Problem





CONCLUSIONS

Conclusions

- Filter approach provides provable convergence
- Statistical surrogate modeling improves efficiency of the filter approach
- Multivariate modeling can improve surrogate modeling
- Combined for efficient constrained black-box optimization

THANK YOU!

Questions?