Optimization of Noisy Functions: Application to Simulations

Geng Deng Michael C. Ferris

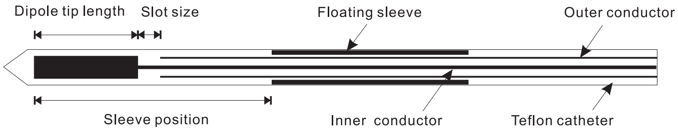
University of Wisconsin-Madison

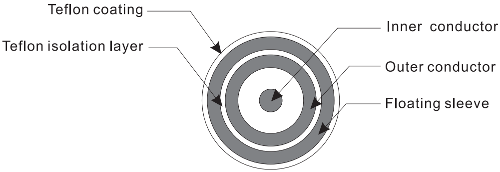
8th US-Mexico Workshop on Optimization and its Applications
Huatulco, Mexico
January 8, 2007

Simulation-based optimization problems

- Computer simulations are used as substitutes to evaluate complex real systems.
- Simulations are widely applied in epidemiology, engineering design, manufacturing, supply chain management, medical treatment and many other fields.
- The goal: Optimization finds the best values of the decision variables (design parameters or controls) that minimize some performance measure of the simulation.
- Other applications: calibration, SVM parameter tuning, inverse optimization, two-stage stochastic integer programming

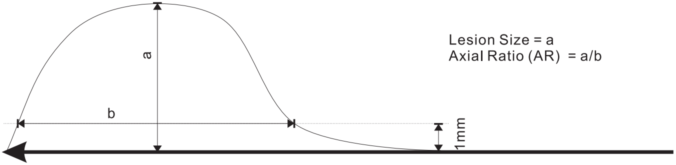
Design a coaxial antenna for hepatic tumor ablation





Simulation of the electromagnetic radiation profile

Finite element models (MultiPhysics v3.2) are used to generate the electromagnetic (EM) radiation fields in liver given a particular design



Metric	Measure of	Goal
Lesion radius	Size of lesion in radial direction	Maximize
Axial ratio	Proximity of lesion shape to a sphere	Fit to 0.5
S_{11}	Tail reflection of antenna	Minimize

A general problem formulation

- We formulate the simulation-based optimization problem as

$$\min_{x \in \mathcal{S}} F(x) = \mathbb{E}_{\omega}[f(x, \omega(x))],$$

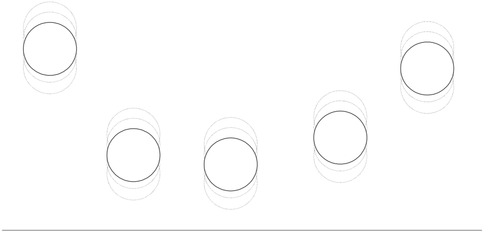
$\omega(x)$ is a random factor arising in the simulation process.
The sample response function $f(x, \omega)$

  - typically does not have a closed form, thus cannot provide gradient or Hessian information
  - is normally computationally expensive
  - is affected by uncertain factors in simulation

The underlying objective function $F(x)$ has to be estimated.

A simple discrete optimization case

- For example, test elasticity of a set of balls. Here $\mathcal{S} = \{1, 2, 3, 4, 5\}$ represents a set of 5 (beach) balls.



- Objective: Choose the ball with the largest expected bounce height $F(x_i)$. $f(x_i, \omega_j)$ corresponds to a single measurement in an experiment.

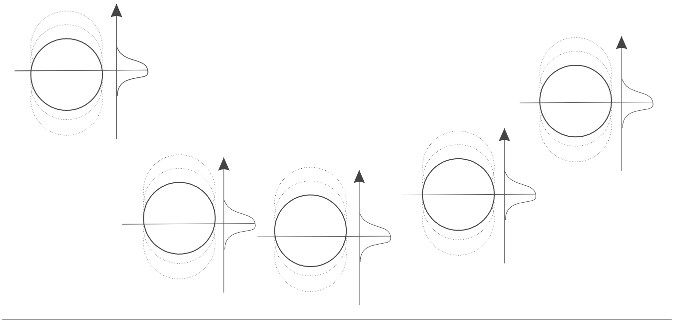
Bayesian approach

- Denote the mean of the simulation output for each system as $\mu_i = F(x_i) = \mathbb{E}_\omega[f(x_i, \omega)]$.
- In a Bayesian perspective, the means are considered as Gaussian random variables whose posterior distributions can be estimated as

$$\mu_i | X \sim N(\bar{\mu}_i, \hat{\sigma}_i^2 / N_i),$$

where $\bar{\mu}_i$ is sample mean and $\hat{\sigma}_i^2$ is sample variance. The above formulation is one type of posterior distribution.

Posterior distributions facilitate comparison



Easy to compute the probability of correct selection (PCS).

Basic framework and tools

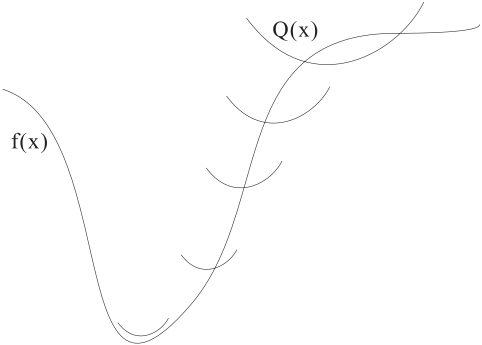
- Small scale x controls/design variables
- Simulation is refinable (replications, more samples in DES, finer discretization)

$$F(x) \simeq \frac{1}{N} \sum_{j=1}^{N} f(x, \omega_j)$$

- WISOPT: Linked two-phase approach
  - Phase I: global issues / exploration: rough
  - Phase II: local issues / exploitation: refined

WISOPT Phase II: noisy UOBYQA (Powell)

The base derivative free optimization algorithm: The UOBYQA (Unconstrained Optimization BY Quadratic Approximation) algorithm is based on a trust region method. It constructs a series of local quadratic approximation models of the underlying function.



Quadratic model construction and trust region subproblem solution

For iteration $k = 1, 2, \ldots,$

- $\cdots$
- Construct a quadratic model via interpolation

$$Q(x,\omega) = f(x_k,\omega) + g_Q^T(\omega)(x - x_k) + \frac{1}{2}(x - x_k)^T G_Q(\omega)(x - x_k)$$

  The model is unstable since interpolating noisy data

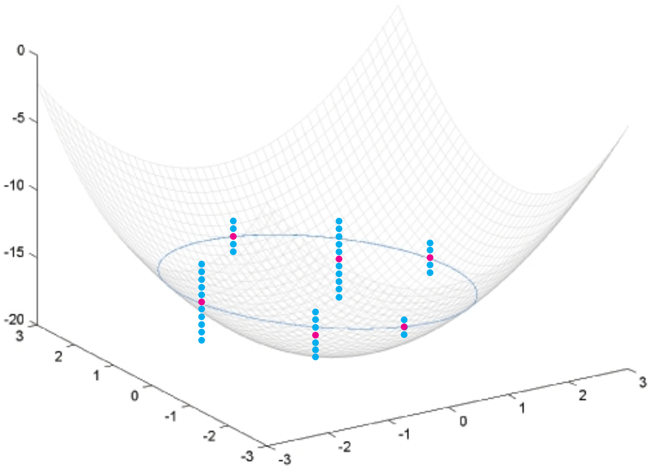
- Solve the trust region subproblem

$$\begin{aligned} s_k(\omega) = \arg\min_s \quad & Q(x_k + s, \omega) \\ s.t. \quad & \|s\|_2 \leq \Delta_k \end{aligned}$$

  The solution is thus unstable

- $\cdots$

Why is the quadratic model unstable?



How to stabilize the quadratic model?

Let $\{y^1, y^2, \ldots, y^L\}$ be the interpolation set.

- Quadratic interpolation model is a linear combination of Lagrange functions:

$$Q(x, \omega) = \sum_{j=1}^{L} f(y^j, \omega) l_j(x).$$

- Each piece $l_j(x)$ is a quadratic polynomial, satisfying

$$l_j(y^i) = \delta_{ij}, i = 1, 2, \cdots, L.$$

- The coefficients of l_j are uniquely determined, independent of the random objective function.

Bayesian estimation of coefficients c_Q, g_Q, G_Q

In Bayesian approach, the mean of function output $\mu(y^j) := \mathbb{E}_\omega f(y^j, \omega)$ is considered as a random variable:
Normal posterior distributions:

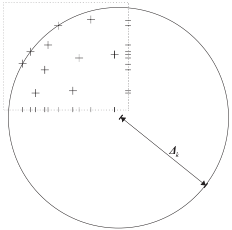
$$\mu(y^j)|X \sim N(\bar{\mu}(y^j), \hat{\sigma}^2(y^j)/N_j).$$

Thus the coefficients of the quadratic model are estimated as:

$$\begin{aligned} g_Q|X &= \textstyle\sum_{j=1}^{L}(\mu(y^j)|X)g_j, \\ G_Q|X &= \textstyle\sum_{j=1}^{L}(\mu(y^j)|X)G_j. \end{aligned}$$

- g_j, G_j are coefficients of Lagrange functions l_j.
- g_j, G_j are deterministic and determined by points y^j.

Constraining the variance of coefficients



- Generate samples of function values from these (estimated) distributions.
- Trial solutions are generated within a trust region. The standard deviation of the solutions is constrained.

$$\max_{i=1}^{n} std([s^{*(1)}(i), s^{*(2)}(i), \cdots, s^{*(M)}(i)]) \leq \beta\Delta_k.$$

Noisy UOBYQA for Rosenbrock, $n = 2$ and $\sigma^2 = 0.01$

Iteration (k)	FN	$F(x_k)$	Δ_k
1	1	404	2
20	78	3.56	9.8×10^{-1}
40	140	0.75	1.2×10^{-1}
60	580	0.10	4.5×10^{-2}
80	786	0.0017	5.2×10^{-3}
✓ Stops with the new termination criterion			
100	1254	0.0019	2.8×10^{-4}
120	2003	0.0016	1.1×10^{-4}
✓ Stops with the termination criterion $\Delta_k \leq 10^{-4}$			

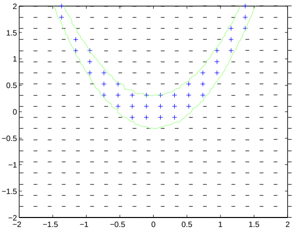
WISOPT Phase I: Classifier

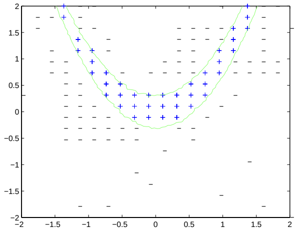
- Global search process
- Classifier: surrogate for indicator function of the level set

$$L(c) = \{x \mid F(x) \le c\} \simeq \left\{ x \;\middle|\; \frac{1}{N}\sum_{j=1}^{N} f(x,\omega_j) \le c \right\}$$

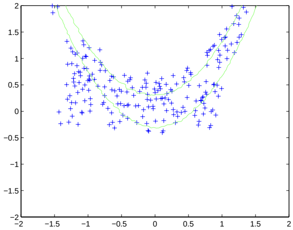
- c is a quantile point of the responses
- Training set: space filling samples (points) from the whole domain (e.g. mesh grid; Latin Hypercube Sampling)

Banana example



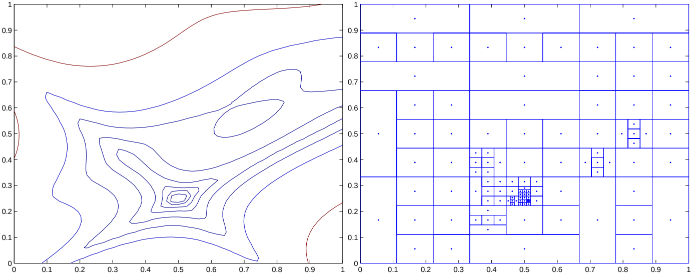


Original Predicted Training



WISOPT Phase I: noisy Direct (Jones)

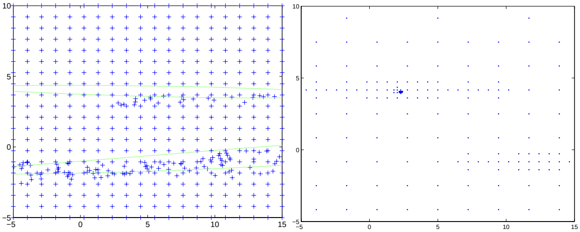
- At each iteration, trisect a collection of promising boxes (large box or small F)
- Evaluate F at center of newly generated boxes



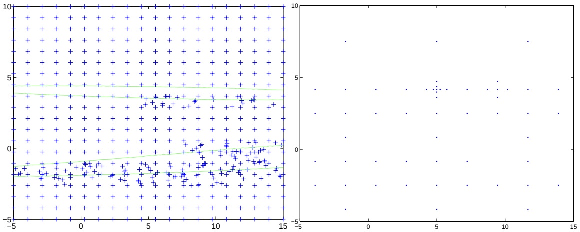
Noisy extension

- Bayesian methods determine posterior distribution of “box center” F values
- Monte Carlo methods to generate “sampled” values for F; then use DIRECT to generate “trial” potential boxes
- Compare error rates against ‘boxes generated from sample means
- When error rate large (sets of boxes chosen differ greatly), increase replications on those boxes that produce errors

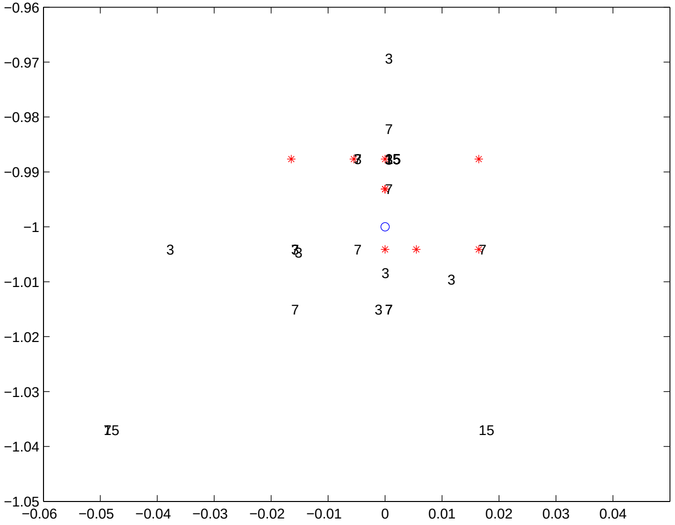
Classifier vs Direct (example FR)



increasing noise



Compare noisy DIRECT vs fixed accuracy



Link: Determine x_0 and TR radius Δ

The idea is to determine the best 'window size' for non-parametric local regression, and then use the 'window size' as the initial trust region radius Δ.

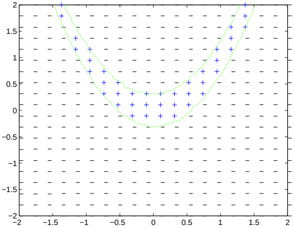
1. $\Delta \in \arg\min_h sse(h)$
2. $sse(h)$ is the sum of squares error of knock-one out prediction. Given a window-size h and a point x_0, the knock-one out predicted value is $Q(x_0)$, where $Q(x)$ is constructed using the data points within the ball $\{x| \, \|x - x_0\| \leq h\}$.

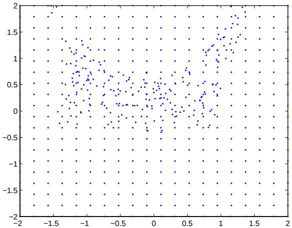
$$Q(x) = c + g^T(x - x_0) + \frac{1}{2}(x - x_0)^T H(x - x_0)$$

.

3. Sort x_0 by F and ensureh points separated by Δ

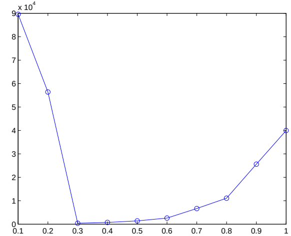
The non-parametric “linking” idea

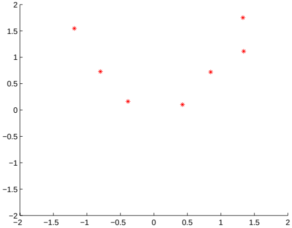




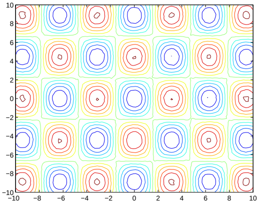
Original / $sse(h)$

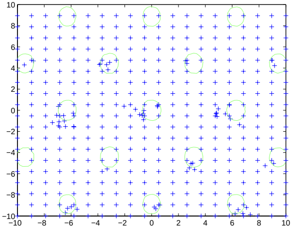
Data / Result

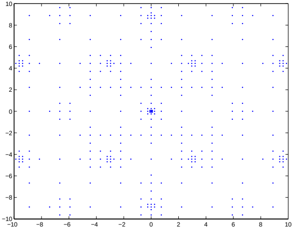




Classifier vs Direct (example Griewank)



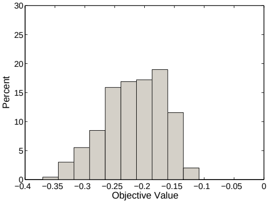




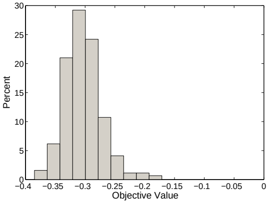
Two-phase approach to optimize antenna design metrics

- Uniform LHS to generate 2,000 design samples to evaluate with the FE simulation model (range [-0.3705, 3597])
- Histogram of objective values over interval [-0.3705, 0]
- $c = -0.2765$ the 10% quantile. $L(c)$ has 199 positive samples (1801 negative)
- Balancing procedure: 398 positive vs. 388 negative samples
- 5 (of 6 tested) classifiers in ensemble
- Refined data: 15,000 designs, 522 predicted by classifiers as positive, 74% correctly
- The best Phase I design has value -0.3850.

Coaxial antenna design



(a) First stage evaluations (training data)



(b) Our new antenna design

- (Modified) UOBYQA started from best point: (13.6 2.7 19.0 0.3 0.1) mm, value -0.3850.
- UOBYQA returned an optimal solution: (15.9 2.4 19.0 0.3 0.1) mm, value -0.4117.

Sample path extension: changing liver properties

- Common random numbers allow variance reduction, correlated noise.
- Extension of ideas to Variable-Number Sample-Path Optimization method.
- Application: Dielectric tissue properties varied within $\pm 10\%$ of average properties to simulate the individual variation.
- Bayesian VNSP algorithm yields an optimal design that is a 27.3% improvement over the original design and is more robust in terms of lesion shape and efficiency.

Conclusions and future work

- Coupling statistical and optimization techniques can effectively process noisy function optimizations
- Significant gains in system performance and robustness are possible using function value distributions
- WISOPT framework allows multiple methods to be "hooked" up
- How to reuse function evaluations from Phase I in Phase II?
- Application to more engineering problems, including two stage stochastic integer programs