Optimization of Noisy Functions

Geng Deng   Michael C. Ferris

University of Wisconsin-Madison

INFORMS, October 15, 2008

A general problem formulation

- We formulate a noisy optimization problem as

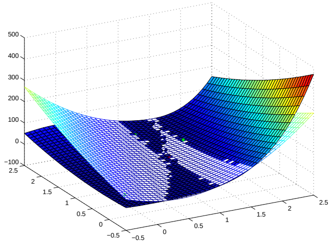
$$\min_{x \in \mathcal{S}} f(x) = \mathbb{E}[F(x, \xi(\omega))],$$

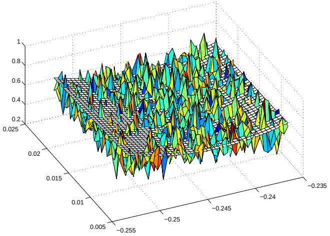
  $\xi(\omega)$ is a random component arising in some simulation process.

- The sample response function $F(x, \xi(\omega))$
  - 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.

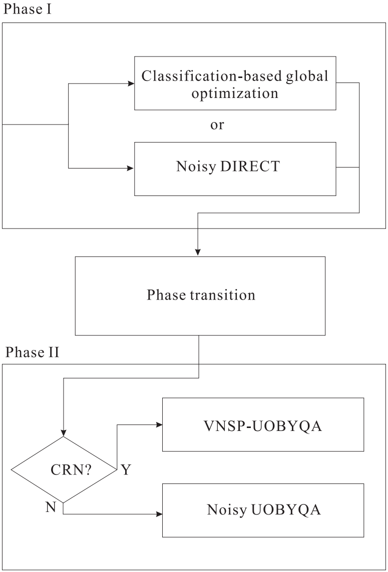
WISOPT two-phase optimization framework

1. **Phase I is a global exploration step.** The algorithm explores the entire domain and proceeds to determine potentially good subregions for future investigation.
2. **Phase II is a local exploitation step.** Local optimization algorithms are applied to determine the final solution.





The flow chart of WISOPT



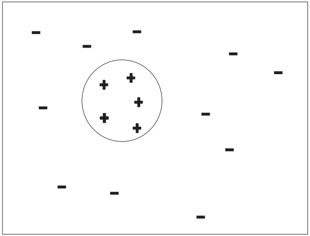
WISOPT Phase I: a classification based global search

- Classifier: surrogate for indicator function of the level set

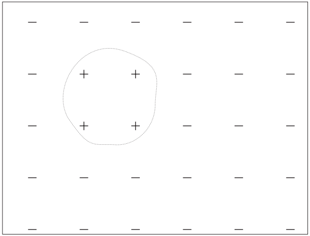
$$L(c) = \{x \mid f(x) \leq c\} \simeq \left\{ x \;\middle|\; \bar{f}(x) = \frac{1}{N} \sum_{j=1}^{N} F(x, \xi_j) \leq c \right\}$$

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

Classifiers predict new refined samples as promising



(a) Training samples in $L(c)$ are classified as positive and others are negative. The solid circle represents estimated $L(c)$.



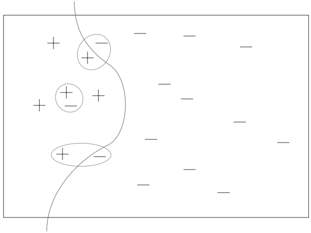
(b) Classify a set of more refined space-filling samples. Four points are predicted as positive and rest are negative. The classifier is refined.

Validate the subset of the identified promising points by performing additional simulations

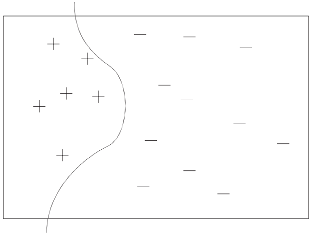
Imbalanced data

- To identify the top promising regions, the best 10% of the training samples are labeled as '+', and the rest are '-'
- The imbalance of the training set causes low classification accuracy, especially for positive members
- Balance the training data set
  - Under-sample of the negative class using one-sided selection
    - Use 1-NN and retain only those negative samples needed to predict training set
    - Clean the dataset with Tomek links
  - Over-sample of the positive class by duplicating positive samples
- Adjust the misclassification penalty

Cleaning the dataset with Tomek links

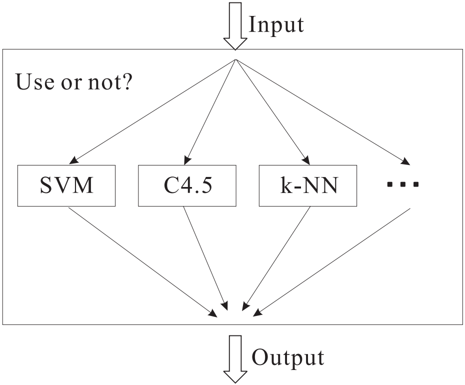


(c) Determine the pairs of Tomek links



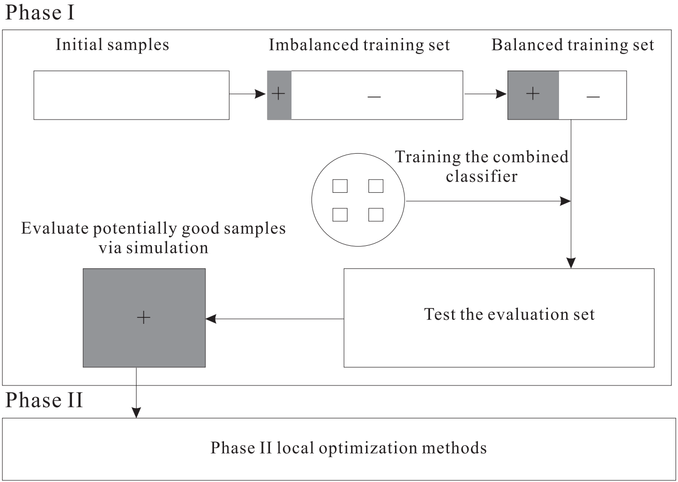
(d) Remove the negative samples participating as Tomek links

Assemble classifiers using a voting scheme

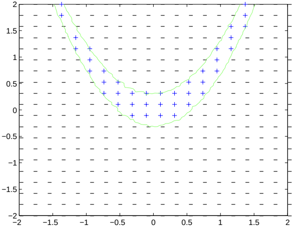


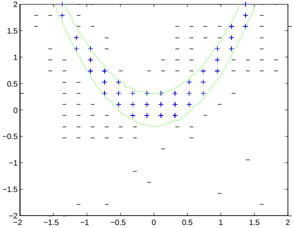
Use g-means on training set to determine which to use

Classifier Phase I approach

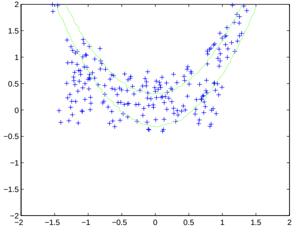


Banana example

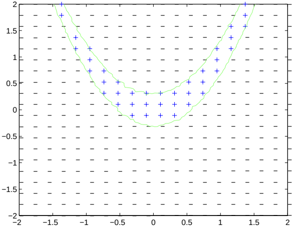


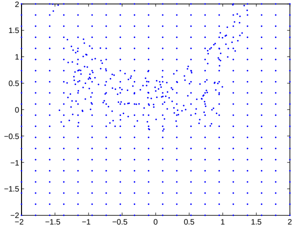


Original Predicted Training

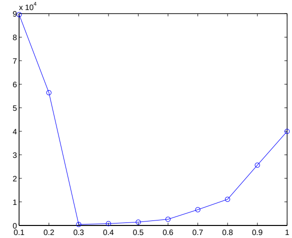


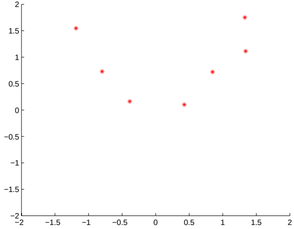
The non-parametric “linking” idea





Original / $sse(h)$ Data / Result





Determine subregion radius by non-parametric regression

The idea is to determine the best 'window size' for non-parametric local quadratic regression

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

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

.

Phase II: refine solution

- Local optimization methods to handle noise
- Derivative-free methods
- Basic approach: reduce function uncertainty by averaging multiple samples per point.
- Potential difficulty:
  efficiency of algorithm vs number of simulation runs
- We apply Bayesian approach to determine appropriate number of samples per point, while simultaneously enhancing the algorithm efficiency

Quadratic model construction and trust region subproblem solution

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

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

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

  The model is unstable since interpolating noisy data

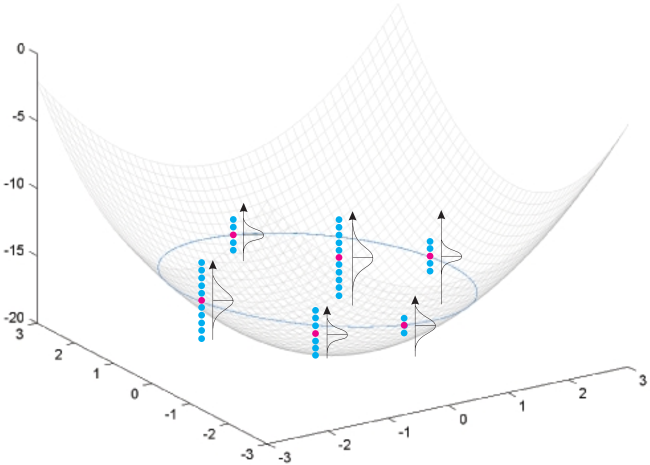
- Solve the trust region subproblem

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

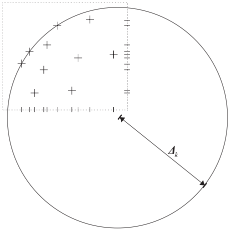
  The solution is thus unstable

- $\cdots$

Bayesian posterior distributions



Constraining the variance of solutions (Monte Carlo validation)

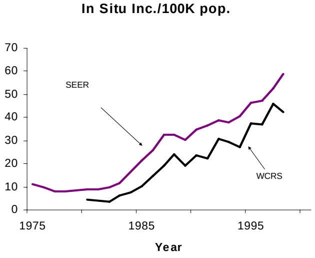


- Generate 'sample quadratic functions' that could arise given current function evaluations.
- Trial solutions are generated within a trust region. The standard deviation of the solutions are constrained.

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

Simulation calibration

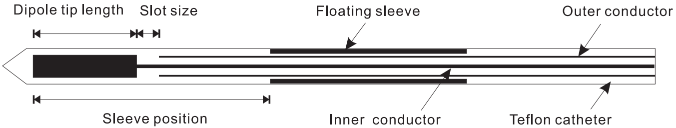
- Detailed individual-woman level discrete event simulation of Wisconsin Breast Cancer Incidence (using 4 processes):
  - Breast cancer natural history
  - Breast cancer detection
  - Breast cancer treatment
  - Non-breast cancer mortality among US women
- Replicate breast cancer surveillance data: 1975-2000

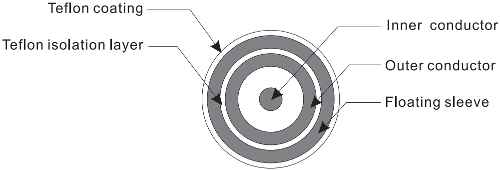


Application to WBCE

- 500,000 points x generated uniformly at random
- Using CONDOR (120 machines) can evaluate approximately 1000 per day
  - $F(x, \xi)$ involves simulation of 3 million women
  - 363 are in $L(10)$: "simulated points out of data envelope"
- Using Phase I: 10,000 points evaluated, 220 points suggested, 195 are in L(10)
- Phase I results in new points (all are good), but 2 of which seem better than the "experts" best solution
- Phase II: application was not necessary

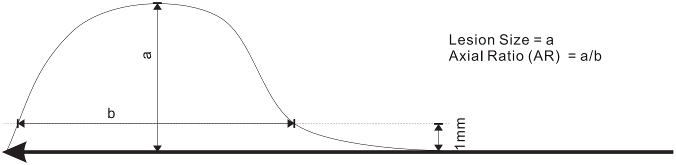
Design a coaxial antenna for hepatic tumor ablation





Simulation of the electromagnetic radiation profile

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

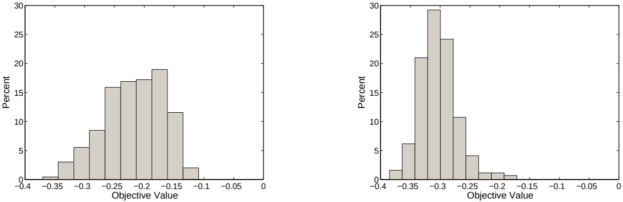


Metric	Measure of	Goal
Lesion radius	Size of lesion in radial direction	Maximize
Axial ratio	Proximity of lesion shape to a sphere	Fit to 0.5
S_{11}	Tail reflection of antenna	Minimize

Two-phase approach to optimize antenna design parameters

- Uniform LHS to generate 1,000 design samples to evaluate with the FE simulation model (range [-0.1409, 0.2903])
- $c = -0.0354$ the 10% quantile. $L(c)$ has 100 positive samples (900 negative)
- Balancing procedure: 200 positive vs. 269 negative samples
- 3 (of 6 tested using g-means) classifiers in ensemble
- Refined data: 20,000 designs, 1914 predicted by classifiers as positive, 74.5% correctly
- The best Phase I design has value -0.2238

Coaxial antenna design



(e) First stage initial designs

(f) Designs predicted by classifiers

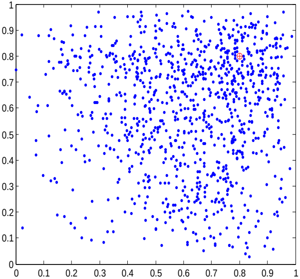
- Phase II started from 5 points predicted: value -0.2238
- Phase II returned an optimal solution: value -0.2501
- Total simulations used = 1000 + 1914 + 750
- DIRECT (4000): -0.2064; SnobFit (4000): -0.1955

Noisy extension: changing liver properties

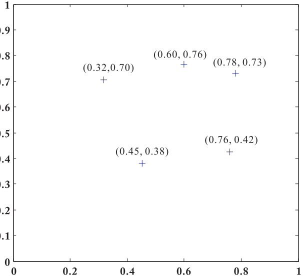
- Dielectric tissue properties varied within $\pm 10\%$ of average properties to simulate the individual variation.
- WISOPT yields an optimal design that is a 27.3% improvement over the original design and is more robust in terms of lesion shape and efficiency.

Ambulance simulation

An ambulance is called when an emergency call occurs. Determine the locations of the ambulance bases such that the expected response time to emergency calls is minimized.



(g) The distribution of emergency calls



(h) The locations of ambulance bases

Conclusions and future work

- Coupling statistical and optimization techniques can effectively process noisy function optimizations
- Significant gains in system performance and robustness are possible
- WISOPT framework allows multiple methods to be "hooked" up

Future work:

- Problems with general constraints
- More optimization algorithms in both phases
- A phase transition module with variable radii