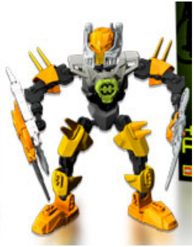
Why model? A view from optimization



Michael C. Ferris

University of Wisconsin, Madison

Supported partly by AFOSR, DOE and NSF

Mayo Clinic, Rochester, MN
March 9, 2012

Quote from Wikipedia: modeling

- A mathematical model is a description of a system using mathematical concepts and language.
- Mathematical models are used in:
    - the natural sciences (such as physics, biology, earth science, meteorology)
    - engineering disciplines (e.g. computer science, artificial intelligence)
    - in the social sciences (such as economics, psychology, sociology and political science)
- Physicists, engineers, statisticians, operations research analysts and economists use mathematical models extensively
- Lack of agreement between theoretical mathematical models and experimental measurements often leads to important advances as better theories are developed

Building mathematical models

- How to model: pencil and paper, excel, Matlab, R, python, ...

  - Linear vs nonlinear
  - Deterministic vs probabilistic
  - Static vs dynamic (differential or difference equations)
  - Discrete vs continuous

- Other issues: Large scale, stochasticity, data (rich and sparse)
- Abstract/simplify:
  - Variables: input/output, state, decision, exogenous, random...
  - Exogenous = data/parameters
  - Objective/constraints
  - Black box/white box
  - Subjective information, complexity, training, evaluation
- Just solving a single problem isn't the real value of modeling: optimization finds "holes" in the model

Why model?

- to understand (descriptive process, validate principles and/or explore underlying mechanisms)
- to predict (and/or discover new system features)
- to combine (engaging groups in a decision, make decisions, operate/control a system of interacting parts)
- to design (strategic planning, investigate new designs, can they be economical given price of raw materials, production process, etc)

To do these, we must be able to capture the problem easily/naturally

Sudoku

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

Understand: Sudoku Model

The aim of this puzzle is to enter a numerical digit from 1 through 9 in each cell of a 9x9 grid made up of 3x3 subgrids (called "regions"), starting with various digits given in some cells (the "givens"). Each row, column, and region must contain only one instance of each numeral.

- r, c, v, k (rows, cols, vals, regions) range from 1 to 9
- binary variables $x_{r,c,v}$

row entries unique: $$\sum_c x_{r,c,v} = 1, \quad \forall r, v$$

col entries unique: $$\sum_r x_{r,c,v} = 1, \quad \forall c, v$$

one val per cell: $$\sum_v x_{r,c,v} = 1, \quad \forall r, c$$

one val per region: $$\sum_{(r,c)\in\mathcal{R}_k} x_{r,c,v} = 1, \quad \forall k, v$$

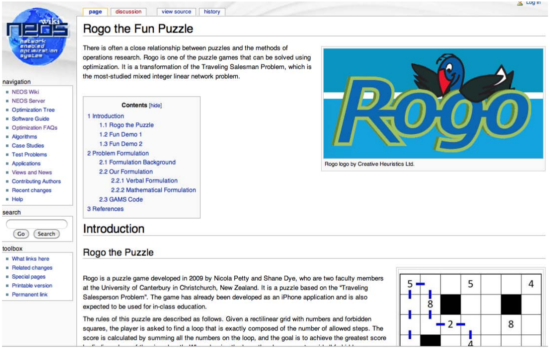
Here $\mathcal{R}_k$ runs over all the k "regions"

Show me on a problem like mine

- Repeated solutions of multiple (different) problems enables "understanding" of the solution space (or sensitivity)
- NEOS wiki (www.neos-guide.org) or try out NEOS solvers (www.neos-solvers.org) for extensive examples

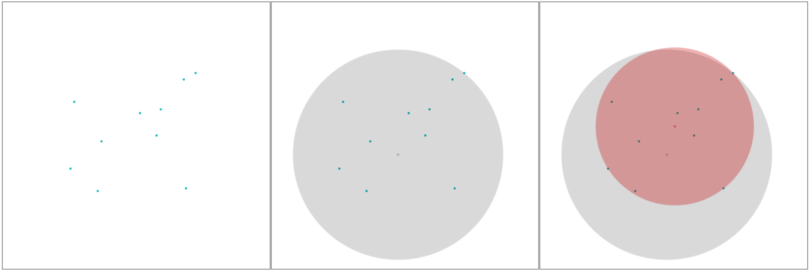
Building a class of case studies:

- Web description of problem
- Solution on NEOS
- Ability to modify and resolve
- Comparison of results
- e.g. Sudoku, Rogo



Abstraction: circle cover problem

Given a set of points find the location (x, y) of the center of the circle with minimum radius that covers all points (coverage problem)

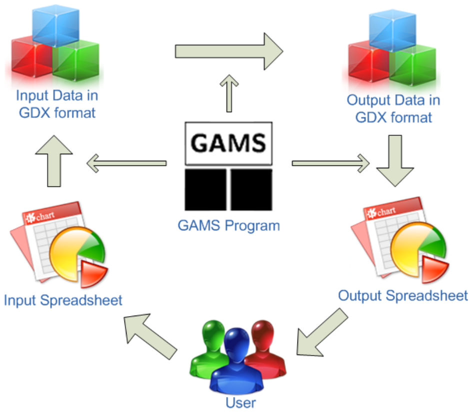


- This is an example of a nonlinear program (second order cone program).
- What if the points are only known by distribution?
- Notion of robust optimization (all points in enclosing circle) or a stochastic programming formulation or chance constraints

Resident rotation scheduling

- Supervised on-the-job training (called residency) in teaching hospitals or academic medical centers
- Residents have to undergo a series of clinically-based trainings (rotations)
- Duration of a rotation usually spans a block of consecutive weeks – depends on the post-graduate year (PGY) of the resident
- The rotation schedules are made once a year
- A schedule must satisfy various training and staffing requirements and certain regulatory restrictions
- A schedule should preferably contain multiple views and enable search, sort and filter functions for easy information retrieval
- In the UW surgery department, this scheduling task used to be performed in Microsoft Excel, by hand

The modeling process



- User specifies input data and parameters in Excel spreadsheets
- Resulting schedule is written back to a spreadsheet in “user” format

	A	B	C	D	E
1	PG	Type	Name	Rotation	Week
2	PG1	GS	Tevis	Vascular	22
3	PG1	GS	Tevis	Vascular	23
4	PG1	GS	Tevis	Vascular	24
5	PG1	GS	Tevis	Vascular	25
6	PG1	GS	Tevis	Vascular	26
7	PG1	GS	Tevis	Peds	37
8	PG1	GS	Tevis	Peds	38
9	PG1	GS	Tevis	Peds	39
10	PG1	GS	Tevis	Peds	40
11	PG1	GS	Tevis	Peds	41
12	PG1	GS	Tevis	Peds	42
13	PG1	GS	Tevis	Night Float	12
14	PG1	GS	Tevis	Night Float	13
15	PG1	GS	Tevis	Night Float	14
16	PG1	GS	Tevis	Night Float	15
17	PG1	GS	Tevis	Night Float	16
18	PG1	GS	Tevis	Gold	48
19	PG1	GS	Tevis	Gold	49
20	PG1	GS	Tevis	Gold	50
21	PG1	GS	Tevis	Gold	51
22	PG1	GS	Tevis	Gold	52
23	PG1	GS	Tevis	Gold	53
24	PG1	GS	Tevis	Thoracic	32
25	PG1	GS	Tevis	Thoracic	33
26	PG1	GS	Tevis	Thoracic	34
27	PG1	GS	Tevis	Thoracic	35

Formulation

x_{ijk} assign resident i to rotation j in block k

$$\min \sum_{j \in R_o, p \in P, w \in W} h_{jpw}$$

$$\sum_{k \in B(i)} x_{ijk} \geq TN_{ij}, \forall i \in R_e, j \in R_o(i)$$

$$\sum_{k \in B(i)} x_{ijk} \leq TA_{ij}, \forall i \in R_e, j \in R_o(i)$$

$$\sum_{j \in R_o(i)} x_{ijk} \leq 1, \forall i \in R_e, k \in B(i)$$

$$\sum_{i \in R_e(p)} \sum_{k \in B(w) \cap B(i)} x_{ijk} + h_{jpw} \geq SN_{jpw}, \forall j \in R_o, p \in P, w \in W$$

Extend also for "preferences"

Deployment and abstraction

- Example of an assignment model
- Compared to the surgery department's 2010-2011 rotation schedule:
  - Result matches very closely with realized schedule
  - Model running time is short, within 10 minutes
  - Moreover, model finds the "best" schedule, whereas the manual procedure only identifies one "satisfactory" schedule
- Hope to deploy the model in a wider user community

Challenges:

- Abstraction/simplification/key drivers
- Size: (spatial/temporal/decision hierarchical)
- Nature of data: sparse, rich, uncertain

Opportunities: facilitates prediction, improved operation, strategic behavior and design

Predict: tradeoff accuracy and simple structure

Many models from statistics: e.g. regression:

$$\min_x \|Ax - y\|^2$$

Additional structure: Compressed sensing: sparse signal to account for y

$$\min_x \|Ax - y\|_2^2 \text{ s.t. } \|x\|_0 \leq c$$

Regularized regression:

$$\min_x \|Ax - y\|_2^2 + \alpha \|x\|_1$$

Predict: tradeoff accuracy and simple structure

Many models from statistics: e.g. regression:

$$\min_x \|Ax - y\|^2$$

Additional structure: Compressed sensing: sparse signal to account for y

$$\min_x \|Ax - y\|_2^2 \text{ s.t. } \|x\|_0 \le c$$

Regularized regression:

$$\min_x \|Ax - y\|_2^2 + \alpha \|x\|_1$$

Machine learning: SVM for classification

$$\min_{w,\xi,\gamma} \sum_i \xi_i + \frac{\alpha}{2} \|w\|^2 \text{ s.t. } D(Aw - \gamma 1) \ge 1 - \xi$$

General model:

$$\min_{x \in X} E(x) + \alpha S(x)$$

X are constraints, E measures "error" and S penalizes bad structure

Image denoising

Rudin-Osher-Fatemi (ROF) model ($\ell_2 - \text{TV}$). Given a domain $\Omega \subset \mathbb{R}^2$ and an observed image $f : \Omega \to \mathbb{R}$, seek a restored image $u : \Omega \to \mathbb{R}$ that preserves edges while removing noise. The regularized image u can typically be stored more economically. Seek to "minimize" both

- $\|u - f\|_2$ and
- the total-variation (TV) norm $\int_\Omega |\nabla u| \, dx$

Use constrained formulations, or a weighting of the two objectives:

$$\min_u P(u) := \|u - f\|_2^2 + \alpha \int_\Omega |\nabla u| \, dx$$

The minimizing u tends to have regions in which u is constant ($\nabla u = 0$). More "cartoon-like" when α is large.

Original, noisy, denoised (tol $= 10^{-2}$, 10^{-4})

Data driven models: challenges and opportunities

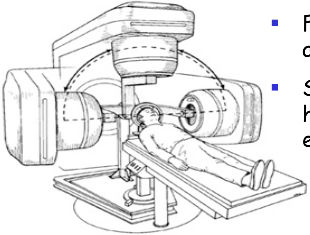
- Matrix completion (e.g. Netflix prize, covariance estimation)
- Machine learning: supervised, unsupervised, semi-supervised, reinforcement, and representation learning
- Probabilistic graphical modeling
- Stochastic processes, statistics, uncertainty quantification

Challenges:

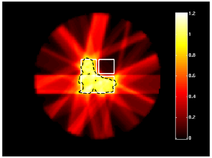
- Terminology issues: active learning = optimal experimental design, reinforcement learning = approximate dynamic programming
- Incorporating domain knowledge into models
- Size and speed for realistic application settings (data sparse and rich environments)
- Online settings, stochastics

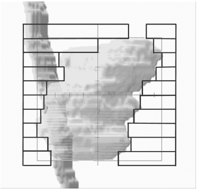
Opportunities: to exploit theory and structure with "big data" via effective algorithms, generalizability, learning behavior

Conformal Radiotherapy

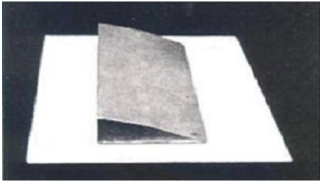


- Fire from multiple angles
- Superposition allows high dose in target, low elsewhere



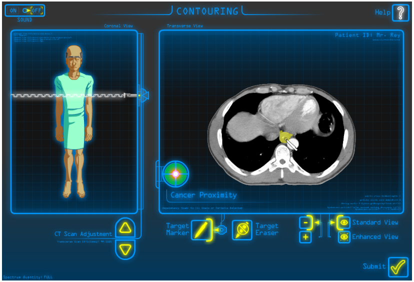


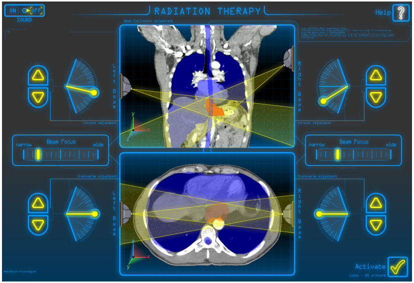
- Beam shaping via collimator
- Gradient across beam via wedges



Combine: the planning process

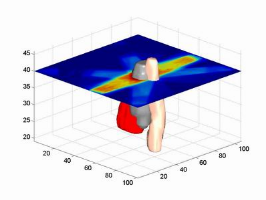
- first contour tumor
- then determine beam angles
- avoid critical structures
- but do it in 3d using only 2d image slices





- Grey – prostate
- Pink – rectum
- Red - bladder

Patient Example



Treatment Modalities

- External beam therapy (photons, electrons or protons)
  - Conformal radiation therapy (CRT)
  - Intensity modulated radiation therapy (IMRT)
  - Intensity modulated arc therapy (IMAT)
  - Tomotherapy
  - Proton therapy (Bragg peak)
- Stereotactic radiosurgery (precise localization)
- Brachytherapy (radioactive seeds)
- Systemic radioactive isotopes (e.g. iodine)

and alternative treatments of surgery and chemotherapy or some combination of all three modalities

The mathematical problem

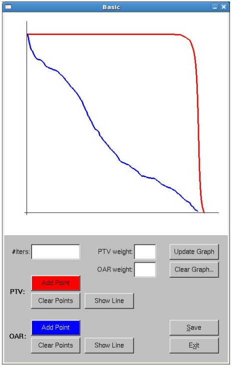
$$\min F(d) \text{ s.t. } d = Px, x \in X, d \in D$$

- P is the fluence map from a given angle in 3dCRT, x are the angle weights
- X represents constraints on the device (typically $x \geq 0$, or cardinality restrictions)
- D represents constraints on the dose distribution (bound constraints, DVH-constraints)

- P could be the pencil beam matrix in IMRT, x are then the bixel weights
- P could represent shots of radiation in Gamma Knife radiosurgery

Many forms for F, X and D

Implementation: a graphical tool



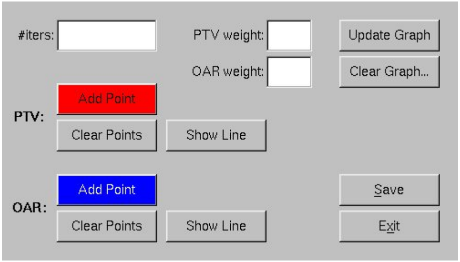
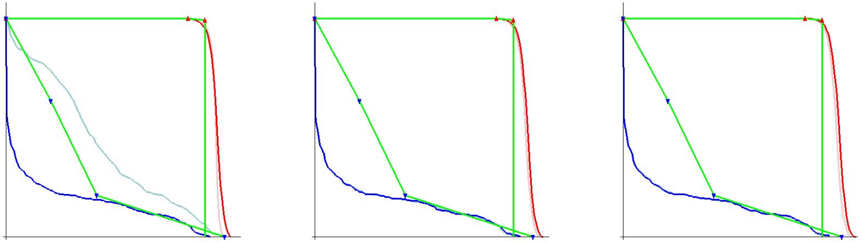


Figure: The user interface presented in our tool, including controls for constraining the PTV and OAR, limiting the number of iterations, weighting the volumes, running solves, clearing new solves and saving images.

The cutting plane approach



(a) The solution generated after one iteration of the algorithm.

(b) The improvement on the solution in (e) after one more iteration.

(c) The improvement on the solution in (f) after 10 iterations.

Figure: A comparison of the progress made by the tool after various numbers of iterations using the same constraints. In each figure, the previous iteration's solution is displayed as the lighter lines.

And then there is uncertainty...

- Parameteric uncertainty (least squares fit of pencil beam/EUD parameters)
- Input data uncertainty (tumor extent/patient characteristics: GTV/CTV/PTV)
- Multi-period models (fractionation/dynamics: positioning/setups)
- Outcome uncertainty (one treatment precludes another follow up treatment/patient variability)
- Uncertainty resolution dependent on action (measurements affect dosage/interactions between treatments)
- Model structural uncertainty (biological response)

Optimization of a model under uncertainty

Modeler: assumes knowledge of distribution

Often formulated mathematically as

$$\min_{x \in X} f(x) = \mathbb{E}[F(x,\xi)] = \int_{\xi} F(x,\xi)p(\xi)d\xi$$

(p is probability distribution).

- Can think of this as optimization with noisy function evaluations
- Traditional Stochastic Optimization approaches: (Robbins/Munro, Keifer/Wolfowitz)
- Often requires estimating gradients: IPA, finite differences
- Compare to stochastic neighborhood search

Sampling methods

- Take sample $\xi_1, \ldots, \xi_N$ of N realizations of random vector ξ
  - viewed as historical data of N observations of ξ, or
  - generated via Monte Carlo sampling
- for any $x \in X$ estimate $f(x)$ by averaging values $F(x, \xi_j)$

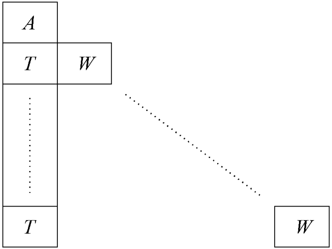
$$\text{(SAA): } \min_{x \in X} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{j=1}^{N} F(x, \xi_j) \right\}$$

- Nice theoretical asymptotic properties
- Can use standard optimization tools to solve the SAA problem
- Implementation uses common random numbers, distributed computation

Stochastic recourse

- Two stage stochastic programming, x is here-and-now decisi on, recourse decisions y depend on realization of a random variab le
- $\mathbb{R}$ is a risk measure (e.g. expectation, CVaR)

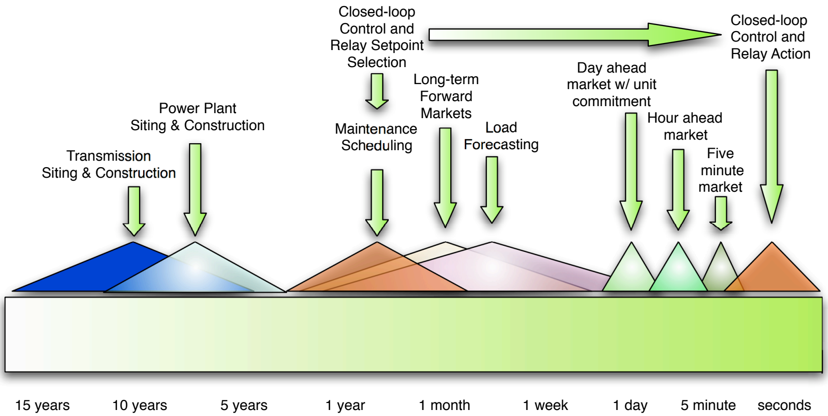
$$
\begin{aligned}
\text{SP:}\ \min \quad & c^\top x + \mathbb{R}[q^\top y] \\
\text{s.t.} \quad & Ax = b, \quad x \geq 0, \\
\forall \omega \in \Omega : \quad & T(\omega)x + W(\omega)y(\omega) \leq d(\omega), \\
& y(\omega) \geq 0.
\end{aligned}
$$



Modeling extensions for Stochastic Programming

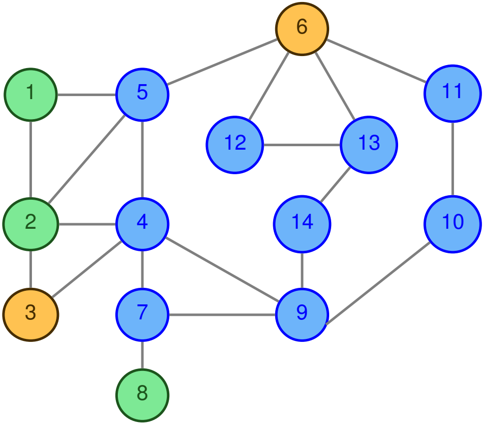
- Robust or stochastic programming
- Can model random variables via distributions
- Have a collection of customizable algorithms available within the modeling system
- Continuous distributions, sampling functions, density estimation
- Chance constraints: $Prob(T_i x + W_i y_i \geq h_i) \geq 1 - \alpha$ - can reformulate as MIP and adapt cuts (Luedtke) empinfo: chance E1 E2 0.95
- Conic or semidefinite programs - alternative reformulations that capture features in a manner amenable to global computation

Combine: Representative decision-making timescales in electric power systems



Combine: Transmission Line Expansion Model

$$\min_{x \in X} \sum_{\omega} \pi_\omega \sum_{i \in N} d_i^\omega p_i^\omega(x)$$



- Nonlinear system to describe power flows over (large) network
- Multiple time scales
- Dynamics (bidding, failures, ramping, etc)
- Uncertainty (demand, weather, expansion, etc)
- $p_i^\omega(x)$: Price (LMP) at i in scenario ω as a function of x
- Use other models to construct approximation of $p_i^\omega(x)$

Many others ... challenges and opportunities

Model predictive control, PDE constrained optimization,...

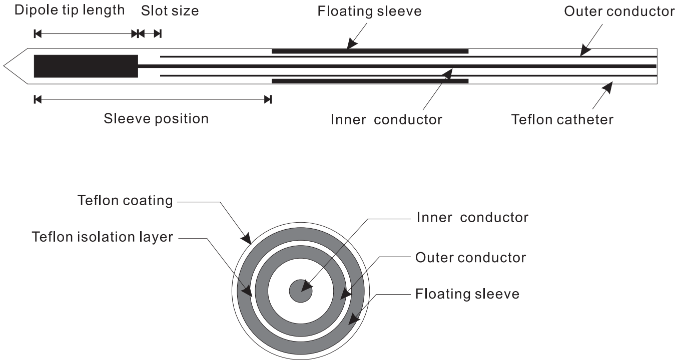
Challenges:

- Size: monster model unable to exploit underlying structure and provide solution quality guarantees
- Stochasticity: How to deal with noisy, sparse, incomplete or inconsistent data and models
- How to coupling collections of (sub)-models: design of interfaces

Opportunities:

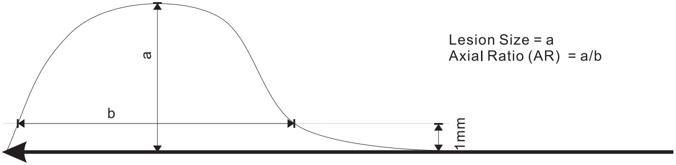
- appropriate detail and consistency of sub-model formulation
- ability for individual subproblem solution verification and engagement of decision makers
- ability to treat uncertainty by stochastic and robust optimization at submodel level and with evolving resolution
- ability to solve submodels to global optimality

Design: 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

Simulation Optimization

- Computer simulations are used as substitutes to understand or predict the behavior of a complex system when exposed to a variety of realistic, stochastic input scenarios
- Widely used in epidemiology, engineering design, manufacturing, supply chain management, medical treatment and many other fields (calibration, parameter tuning, inverse optimization)

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

- The sample response function $F(x, \xi)$
  - 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
- Use of derivative free methods

Issues

- Our approach only valid for small scale (≤ 30) design variables (but the simulation may be very complex -black box)
- Evaluations may be noisy:
  - 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.
- Computational time: variance reduction, correlated noise.

Challenges and opportunities

Challenges:

- Engaging the designer, collecting appropriate data
- Incorporating domain design tools into general (optimization) framework
- Modeling human behavior
- Determining appropriate model: Linear vs nonlinear, deterministic vs probabilistic, static vs dynamic, discrete vs continuous (smooth or nonsmooth)

Opportunities:

- Enormous: medical device design, drug design, radiation therapy machine and planning, bio-engineering
- economic instrument and policy design, smart grid, electric batteries, environmental remediation, offshore drilling and wind farms
- recommender systems, fabrication, election district gerrymandering

Conclusions

- Optimization helps understand what drives a system
- Operational (tactical) and strategic models used in decision processes
- Understand, predict, combine, design
- Uncertainty is present everywhere (the world is not "normal")
- We need not only to quantify it, but we need to hedge/control/ameliorate it
- Modeling, optimization, and computation embedded within the application domain is critical
- Wisconsin Institutes for Discovery is doing this (http://www.discovery.wisc.edu)