Why use a modeling language: a view from optimization

Michael C. Ferris

Computer Sciences and Wisconsin Institutes for Discovery
University of Wisconsin, Madison

Mathematical tools for evolutionary systems biology
May 30, 2013

Why use optimization

Michael C. Ferris

Computer Sciences and Wisconsin Institutes for Discovery
University of Wisconsin, Madison

Mathematical tools for evolutionary systems biology
May 30, 2013

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:

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

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 (Wright)

Rudin-Osher-Fatemi (ROF) model (ℓ_2−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})

Parameter estimation

Example (Crombach):

$$\min_p J(x(p) - \bar{x}) \text{ s.t. } \frac{\partial x}{\partial t} = D\Delta x + f(x,p), p \in P$$

Key points:

- Constraints on parameter choice $p \in P$
- Can solve using PDE constrained optimization. Huge literature in applied mathematics. Key computational idea for optimization is that of the adjoint operator

Parameter estimation

Example (Crombach):

$$\min_p J(x(p) - \bar{x}) + \alpha \|p\|_1 \text{ s.t. } \frac{\partial x}{\partial t} = D\Delta x + f(x,p), p \in P$$

Key points:

- Constraints on parameter choice $p \in P$
- Can solve using PDE constrained optimization. Huge literature in applied mathematics. Key computational idea for optimization is that of the adjoint operator
- Can discretize/optimize, and then add L_1 penalization to get "sparse" (parameter) solution via nonlinear optimization
- Extension to nonsmooth f - DVI, and MPEC, allows for switching

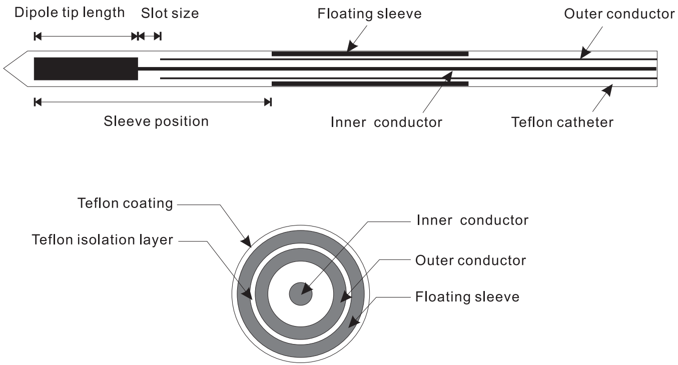
Simulation Optimization

- 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)

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

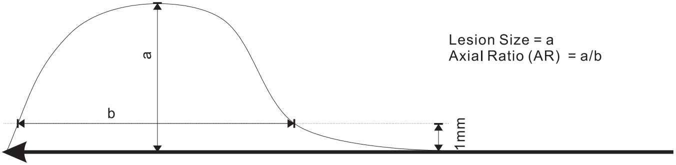
- The sample response function $F(p, \xi)$
  - typically does not have a closed form, thus cannot provide gradient or Hessian information
  - is normally computationally expensive
  - is affected by uncertain factors in simulation

Design a coaxial antenna for hepatic tumor ablation



Simulation of the electromagnetic radiation profile

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

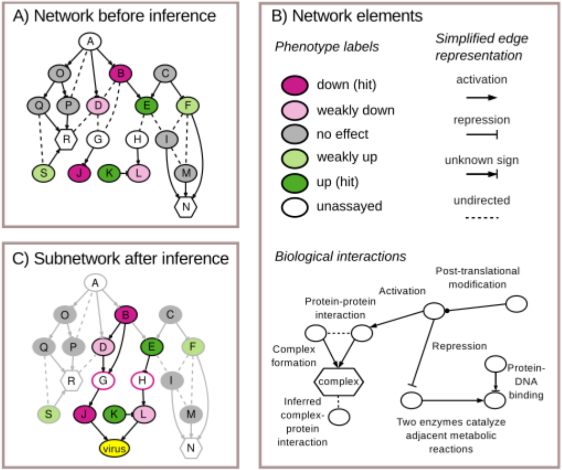


Metric	Measure of	Goal
Lesion radius	Size of lesion in radial direction	Maximize
Axial ratio	Proximity of lesion shape to a sphere	Fit to 0.5
S_{11}	Tail reflection of antenna	Minimize

Computational results

- Use of derivative free (surrogate) methods
- 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 (variable number sample path) 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.

Network inference



- Given prior knowledge, select paths, color nodes and sign arcs to explain as many hits as possible
- e.g. sign of a relevant edge is consistent with the phenotypes of nodes it connects

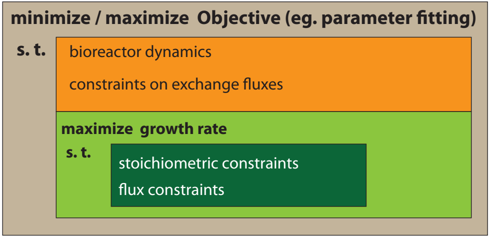
- Can model (propositional) logic constraints in a mixed integer program
- Key issue is to determine objective

Biological Hierarchical Models

I: Opt knock (a bilevel program)

max bioengineering objective (through gene knockouts)
s.t. max cellular objective (over fluxes)
    s.t. fixed substrate uptake
        network stoichiometry
        blocked reactions (from outer problem)
    number of knockouts $\leq$ limit

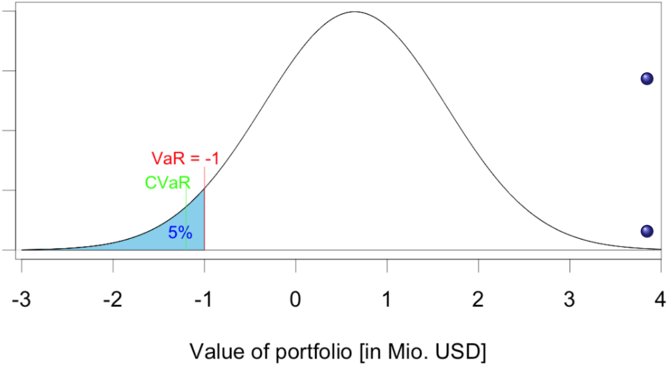
II: Bio-reactor dynamics:



Different mathematical programming techniques are used to transform the problem to a nonlinear program. The differential equations are transformed into nonlinear constraints using collocation methods.

Optimization of risk measures

- Determine portfolio weights w_j for each of a collection of assets
- Asset returns v are random, but jointly distributed
- Portfolio return $r(w, v)$



- Value at Risk (VaR) can be viewed as a chance constraint (hard):
- CVaR gives rise to a convex optimization problem (easy)

- Chance constraints (implemented using mixed integer programming):

$$\min_x c^T x \text{ s.t. } Pr(Ax \leq b) \geq \pi$$

Example: Portfolio Model

- Maximize the mean of the lower tail (mean tail loss):

$$\begin{array}{ll} \max & \underline{CVaR}_{\alpha}(r) \\ \text{s.t.} & r = \sum_j v_j * w_j \\ & \sum_j w_j = 1, \;\; w \geq 0 \end{array}$$

- Jointly distributed random variables v, realized at stage 2
- Variables: portfolio weights w in stage 1, returns r in stage 2
- Coherent risk measures $\mathbb{E}$ and $\underline{CVaR}$ (or convex combination)

Example: Portfolio Model

- Maximize the mean of the lower tail (mean tail loss):

$$\begin{array}{ll} \max & \underline{CVaR}_\alpha(r) \\ \text{s.t.} & r = \sum_j v_j * w_j \\ & \sum_j w_j = 1, \; w \geq 0 \end{array}$$

- Jointly distributed random variables v, realized at stage 2
- Variables: portfolio weights w in stage 1, returns r in stage 2
- Coherent risk measures $\mathbb{E}$ and $\underline{CVaR}$ (or convex combination)
- Optimization modeling systems have new tools for sampling, risk measures and solution of stochastic programs (ref: M. Loewe)
- Classical: mean-variance model (Markowitz)

$$\begin{array}{ll} \min & w^T \Sigma w - q \sum_j v_j * w_j \\ & \sum_j w_j = 1, \; w \geq 0 \end{array}$$

Conclusions

- Optimization helps understand what drives a system
- Constraints are a crucial design/modeling tool
- Uncertainty is present everywhere: we need to hedge/control/ameliorate it
- Collections of, and interactions between, models are critical
- Modern computational optimization tools can be very fast, deal with large amounts of data and variables, address non-convex and discrete issues, interact with dynamics