Radiation Treatment Planning: A View from Optimization

Michael C. Ferris

University of Wisconsin-Madison

AAPM, Houston, July 27, 2008
The (deterministic) mathematical problem

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

- $P$ is a pencil beam matrix, $x$ are bixel weights
- $X$ represents constraints on the bixel weights (typically $x \geq 0$, or cardinality restrictions)
- $D$ represents constraints on the dose distribution (bound constraints, DVH-constraints)
- Alternatively: Columns of $P$ are the dose deposited to voxels from an aperture.
- $x$ is now the aperture weight.
- In practice, too many apertures, generate them on the fly.

Details hidden in definitions of $F$, $X$ and $D$
Issues

\[ \min F(d) \text{ s.t. } d = P x, x \in X, d \in D \]

- \( P \) is extremely large and dense
- Must solve problem relatively quickly
- Ability to modify solution quickly

In many cases, objective is made up of a weighted sum of objectives:

\[ F(d) = \sum_s w_s F^s(d) \]

\( s \) may range, for example, over structures

Different machines, different deliveries, common goals:

- conformity/avoidance
- homogeneity (old), dose shaping (new)
- no streaking
Standard Optimization Approaches

- $F_S$ is weighted least squares - leading to bound constrained quadratic programs
- Key step for algorithms (gradient projection, two-metric projection, conjugate gradients):
  - calculate $P_S^T P_S \cdot v$ for any $v$ and any $s$
  - or $P_I^T P_I \cdot v$ where $I \subseteq S$
- Alternative: EUD (convex optimization), TCP, etc
- Alternative: Linear programming, piecewise linear approximation
  - How to restart? $P \rightarrow P + \Delta P$
  - Barrier method - hard
  - Simplex method - can generate a feasible solution easily
- Extension: discrete variables - MIP approaches (CPLEX, XPRESS are commercial methods)
Column Generation

Master Problem: Only use a subset $J$ of the apertures/pencils:

$$\min F(d) \text{ s.t. } d = P_J x_J, x_J \geq 0$$

Optimality conditions:

$$\pi = \nabla F(d), \; d = P_J x_J, \; 0 \leq P_J^T \pi \perp x_J \geq 0$$

So solving Master Problem gives solution $x$ and $d$ and $\pi$.

Pricing problem: generate a new column $a$ of $P$ that violates optimality conditions: $a^T \pi < 0$

- Decomposes over beams (i.e. $P = [P^1 \; P^2 \; \cdots \; P^n]$)
- Specialized subproblems (network flows) for
  - Interdigitation
  - Disconnected apertures
- DAO approach is related, generates columns of $P$ on the fly via simulated annealing, limits number of apertures from each angle.
Extra constraints on leaf movement along arc

One idea: adapt DAO (simulated annealing) - stochastic guided local search
  - “Randomly” force change in one aperture
  - Cheap update to objective function
  - Problem: resulting sweep may not be deliverable: reject change

Given forced change of shape on one sweep, minimize total leaf distance, or maximum leaf distance on sweep to get feasible.

Each subproblem is an easy network flow problem.

Update weights (using restart procedures) based on new shapes.
Dose Shaping

At least fraction $\alpha$ of volume $Y$ should receive doses exceeding $L_Y$:

$$G(L_Y) = P(D_Y \leq L_Y) \leq 1 - \alpha$$

Embed this in a broader class of problems:

$$G(t) \leq \Psi(t), 0 \leq t \leq L_Y$$

where $\Psi()$ is a postulated profile of radiation doses in $Y$.
Solution via cutting plane methodology (Dentcheva et al)
Key ideas
Example: Chance Constrained Problems

$$\min_{x \in X} f(x) \text{ s.t. } \Pr(C(x, \xi) > 0) \leq \alpha$$

$\alpha$ is some threshold parameter, $C$ is vector valued

- joint probabilistic constraint: all constraints satisfied simultaneously - possible dependence between random variables in different rows
- extensive literature
- linear programs with probabilistic constraints are still largely intractable (except for a few very special cases)
  - for a given $x \in X$, the quantity $\Pr(C(x, \xi) > 0)$ requires multi-dimensional integration
  - the feasible region defined by a probabilistic constraint is not convex
- Recent work by Ahmed, Leudtke, Nemhauser and Shapiro
Types of Uncertainty

- Parameteric uncertainty (least squares fit of pencil beam)
- Input data uncertainty (tumor extent/patient characteristics)
- Multi-period models (fractionation/dynamics)
- Outcome uncertainty (one treatment precludes another follow up treatment)
- Uncertainty resolution dependent on action (measurements affect dosage)
- Model structural uncertainty (biological response)
Extension: 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: (Robinson/Munro, Keifer/Wolfowitz)
- Often require estimating gradients: IPA, finite differences
- Stochastic neighborhood search
Example: Two stage stochastic LP with recourse

\[
\min_{x \in \mathbb{R}^n} c^T x + \mathbb{E}[Q(x, \xi)] \quad \text{s.t. } Ax = b, \quad x \geq 0
\]

\[
Q(x, \xi) = \min_y q^T y \quad \text{s.t. } Tx + Wy = h, \quad y \geq 0
\]

\(\xi = (q, h, T, W)\) (some are random). Expectation wrt \(\xi\).

\(x\) are first stage vars, \(y\) are second stage vars.

Special case: discrete distribution \(\Omega = \{\xi_i : i = 1, 2, \ldots, K\}\)

Diagram:

```
  A
 /|
/ T|
| W|
```

Dotted line:

```
  \ldots  \ldots  \ldots
  T       \ldots
  \ldots  \ldots  \ldots
    W
```
Key-idea: Non-anticipativity constraints

- Replace $x$ with $x_1, x_2, \ldots, x_K$
- Non-anticipativity: 
  \[(x_1, x_2, \ldots, x_K) \in L (a \text{ subspace}) - \text{the } H \text{ constraints} \]

Computational methods exploit the separability of these constraints, essentially by dualization of the non-anticipativity constraints.

- Primal and dual decompositions (Lagrangian relaxation, progressive hedging, etc)
- $L$ shaped method (Benders decomposition applied to det. equiv.)
- Trust region methods and/or regularized decomposition
Sampling methods

But what if the number of scenarios is too big (or the probability distribution is not discrete)? use sample average approximation (SAA)

- Take sample \( \xi_1, \ldots, \xi_N \) of \( N \) realizations of random vector \( \xi \)
  - viewed as historical data of \( N \) observations of \( \xi \), 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
- Monte Carlo Sampling (Quasi-Monte Carlo Sampling)
Example: Robust Linear Programming

Data in LP not known with certainty:

$$\min c^T x \text{ s.t. } a_i^T x \leq b_i, \ i = 1, 2, \ldots, m$$

Suppose the vectors $a_i$ are known to be lie in the ellipsoids (no distribution)

$$a_i \in \varepsilon_i := \{\bar{a}_i + P_i u : \|u\|_2 \leq 1\}$$

where $P_i \in \mathbb{R}^{n \times n}$ (and could be singular, or even 0).
Conservative approach: robust linear program

$$\min c^T x \text{ s.t. } a_i^T x \leq b_i, \ \text{for all } a_i \in \varepsilon_i, i = 1, 2, \ldots, m$$
Robust Linear Programming as SOCP

The constraints can be rewritten as:

\[ b_i \geq \sup \{ a_i^T x : a_i \in \varepsilon_i \} \]
\[ = \bar{a}_i^T x + \sup \{ u^T P_i^T x : \|u\|_2 \leq 1 \} = \bar{a}_i^T x + \|P_i^T x\|_2 \]

Thus the robust linear program can be written as

\[ \min c^T x \text{ s.t. } \bar{a}_i^T x + \|P_i^T x\|_2 \leq b_i, \ i = 1, 2, \ldots, m \]

\[ \min c^T x \text{ s.t. } (b_i - \bar{a}_i^T x, P_i^T x) \in C \]

where \( C \) represents the second-order cone. Solution (as SOCP) by Mosek or Sedumi, CVX, etc.
Example: 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.
- Simulations are widely applied in epidemiology, engineering design, manufacturing, supply chain management, medical treatment and many other fields.
- Optimization applications: calibration, parameter tuning, inverse optimization, pde-constrained 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.
Bayesian approach

- The underlying objective function $f(x)$ still has to be estimated.
- Denote the mean of the simulation output for each system as $\mu_i = f(x_i) = \mathbb{E}[F(x_i, \xi)]$
- 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.

- Instrument existing optimization codes to use this derived distribution information
  - Derivative free optimization, surrogate optimization
  - Response surface methodology
  - Evolutionary methods
Example: Risk Measures

- Classical: utility/disutility function $u(\cdot)$:

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

- Modern approach to modeling risk aversion uses concept of risk measures
  - mean-risk
  - semi-deviations
  - mean deviations from quantiles, VaR, CVaR
  - Römish, Schultz, Rockafellar, Urasyev (in Math Prog literature)
  - Much more in mathematical economics and finance literature
  - Optimization approaches still valid, different objectives
CVaR_\alpha(d):

CVaR_\alpha: mean of upper tail at level \alpha the average dose received by the subset of relative volume \((1 - \alpha)\) receiving the highest dose. (Think of \(\alpha = 0.95\) and this is then the mean of the upper tail, ie those values beyond the 95th percentile).
Example: Model Predictive Control

- Models predict outputs of dynamical system due to changes in inputs
- Used heavily in chemical engineering (also DP and extensions)

@ time = $t_k$

Take process measurements

Process model:
- Current & future
  - Control actions
  - Disturbances
- Future process outputs

Objectives
Constraints

Solve above optimization problem

Best current and future control actions

Implement best current control action

time = $t_{k+1}$

courtesy: Nikolaou
Michael Ferris (University of Wisconsin)
Recap points

Solving a problem with “averaged” data does not work (1/2 time in A, B: never at average location)
How to quantify/measure: tumor/organs might not be volume preserving
- Time available for solution
- Recourse actions available
- Knowledge of uncertainty distribution

Error vs uncertainty: patient positioning
- Overdose today - cannot remove dose
- Stochastic integer programming
- Nonlinear (convex or otherwise) recourse models
So what’s my point?

- Modeling and optimization model building is key!
- Many different optimization approaches to treat (model) uncertainties
- How much do I know about distribution of data?
- Specific models needed for these applications
- Stochastic model implementation and interfaces to these tools are needed
- Specialized implementations to allow “dense” data, fast updates, nonlinear approaches and approximations