New Stochastic Programming Approaches for Radiation Therapy Planning

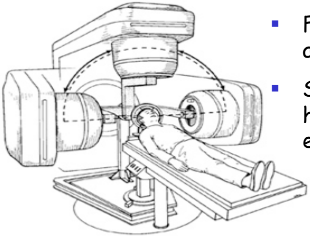
Darinka Dentcheva   Michael C. Ferris   Andrzej Ruszczynski

Laura Legault

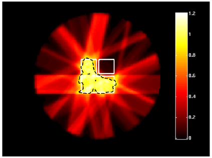
Computer Sciences, University of Wisconsin-Madison

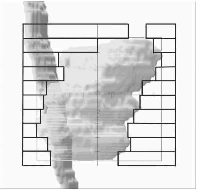
BIRS Workshop, Banff, March 12, 2011

Conformal Radiotherapy

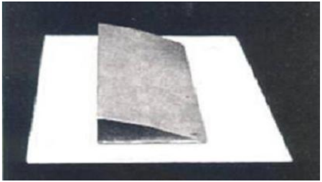


- Fire from multiple angles
- Superposition allows high dose in target, low elsewhere



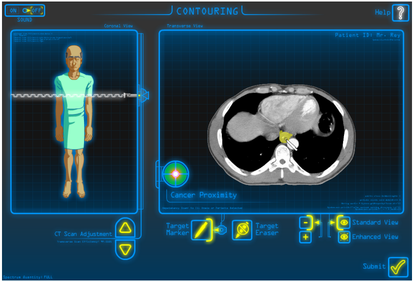


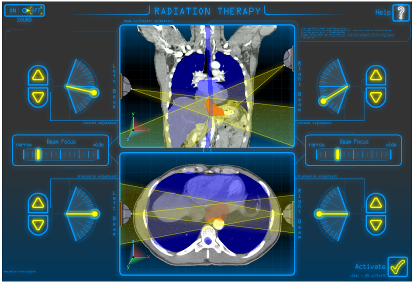
- Beam shaping via collimator
- Gradient across beam via wedges



The planning process

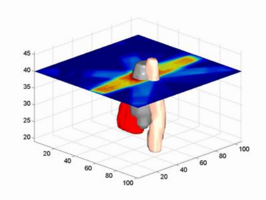
- first contour tumor
- then determine beam angles
- avoid critical structures
- but do it in 3d using only 2d image slices



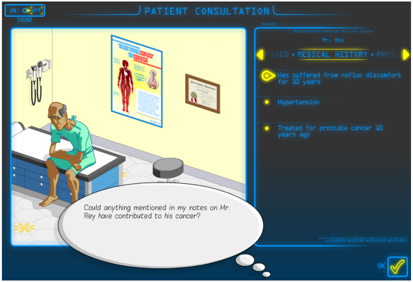


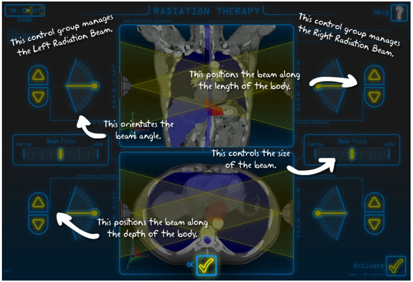
- Grey – prostate
- Pink – rectum
- Red - bladder

Patient Example

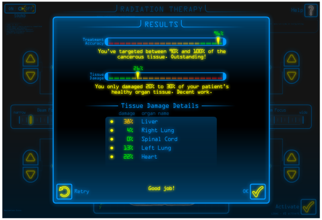


An Oncology Game





- Game developed by the Educational Research Challenge Area (ERCA) of the Wisconsin Institutes of Discovery
- Coming soon to a web browser near you!



The classical 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

Types of 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

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

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

Chance Constrained Problems

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

α 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 $Prob(C(x, \xi) > 0)$ requires multi-dimensional integration
  - the feasible region defined by a probabilistic constraint is not convex
- Recent work by Ahmed, Luedtke, Nemhauser and Shapiro

Other approaches for treating uncertainty

- Robust optimization (worst case analysis)
- Recourse problems (multi-stage decisions)
- Model predictive control
- Approximate Bayesian computation (ABC)
- Importance sampling and variance reduction
- Risk measures, etc

Financial Modeling: Risk Measures (Shaping Distributions)

Key idea is to control features of the probability distribution

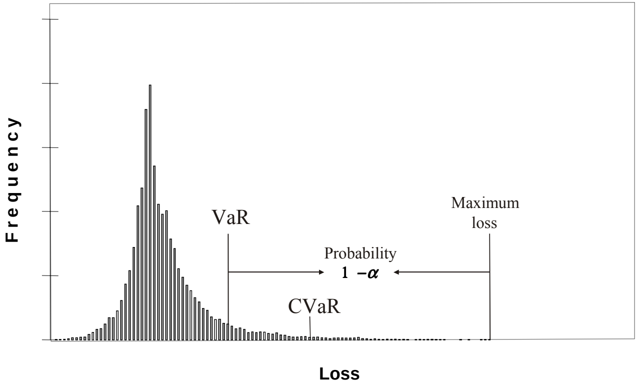
- Classical: utility/disutility function $u(\cdot)$:

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

- 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 optimization literature)
  - Much more in mathematical economics and finance literature
  - Optimization approaches still valid, different objectives

Could use for personalized “risk” preferences

$\overline{CVaR}_\alpha(d)$



$\overline{CVaR}_\alpha$: mean of upper tail at level α - the average dose received by the subset of relative volume $(1-\alpha)$ receiving the highest dose.

Key observation

Clear that this is equal to the average dose of the $(1-\alpha)N$ voxels (point volumes) receiving highest dose.
Rewriting this in symbols:

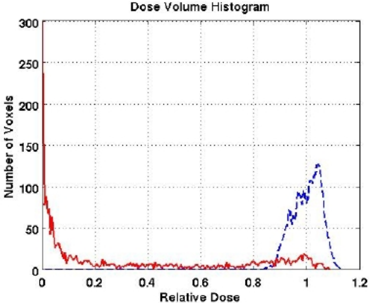
$$\overline{CVaR}_\alpha(d) = \overline{VaR}_\alpha(d) + \frac{1}{(1-\alpha)N}\sum_{j=1}^{N}\left(d - \overline{VaR}_\alpha(d)\right)_+$$

Thus $\overline{CVaR}$ is just $\overline{VaR}$ moved to the right by the average of the tail.
The next step is a clever theorem due to Ogryczak and Tamir (2003) that states this expression can be written as:

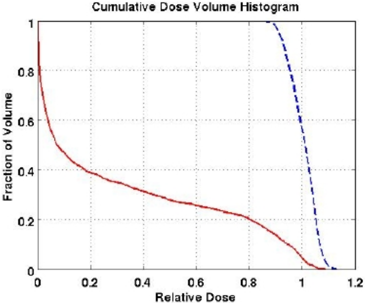
$$\overline{CVaR}_\alpha(d) = \min_{a\in\mathbb{R}}\left\{a + \frac{1}{(1-\alpha)N}\sum_{j=1}^{N}(d-a)_+\right\}$$

Thus can impose linear constraints to get $\overline{CVaR}_\alpha(d) \le U$

Dose volume and cumulative dose volume histograms



(a) The dose volume histogram corresponding to a particular solve .



(b) The cumulative dose volume histogram of the same solve.

Dose shaping problems

- Standard constraints (of the DVH form) are:
  No more than the fraction α of volume X should receive doses exceeding U_X:

$$F(U_X) = P\{D_X \leq U_X\} \geq 1 - \alpha. \tag{1}$$

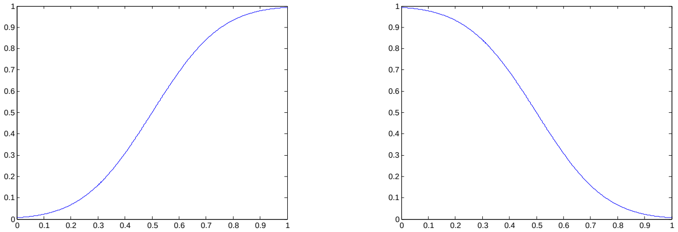
- These are probabilistic constraints.
- Existing approaches use counting and MIP constraints for example

A probabilistic formulation (for OAR)

The proportion of voxels $v \in X$ receiving dose not exceeding t:

$$F(t) = \frac{\mathrm{card}(v \in X : D_X(v) \leq t)}{\mathrm{card}(X)} \tag{2}$$

Cumulative distribution function of random variable D_X representing dose received by randomly selected voxel in X



$F(t)$ related to Cumulative Dose Volume Histogram (cDVH) of OAR:

$$F(t) = 1 - \mathrm{cDVH}_X(t)$$

PTV is similar

Proportion of voxels $v \in Y$ receiving doses not exceeding t as a function of t is a distribution function of a random variable D_Y representing the dose received by a randomly selected voxel in Y:

$$G(t) = \frac{\mathrm{card}(v \in Y : D_Y(v) \leq t)}{\mathrm{card}(Y)}. \tag{3}$$

$G(t)$ is related to cDVH of PTV as follows:

$$G(t) = 1 - \mathrm{cDVH}_Y(t).$$

The requirement that at least fraction β of volume Y should receive doses exceeding L_Y can be written as

$$G(L_Y) = P\{D_Y \leq L_Y\} \leq 1 - \beta. \tag{4}$$

Stochastic dominance

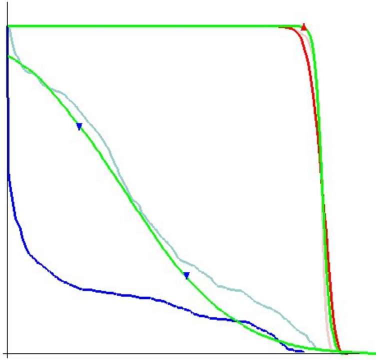
Problems involving such constraints are well-studied in the literature. They are, in general, very difficult to analyze and to solve, because of the non-convexity of the feasible region defined by (1) and (4).
It appears natural to control the shape of these histograms, not just their values at U_X and L_Y, respectively. Therefore, we require

$$\begin{aligned} F(t) &\geq \Phi(t), \quad U_X \leq t < \infty, \\ G(t) &\leq \Psi(t), \quad 0 \leq t \leq L_Y. \end{aligned} \tag{5}$$

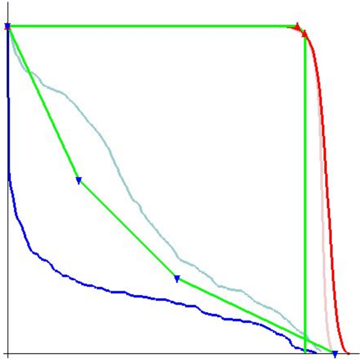
Here $\Phi(\cdot)$ and $\Psi(\cdot)$ are fixed benchmark distribution functions satisfying the conditions

$$\Phi(U_X) = 1 - \alpha, \qquad \Psi(L_Y) = 1 - \beta.$$

Benchmark distributions



(c) A solution using a cumulative normal constraint curve.



(d) A solution using piecewise linear constraint curves constructed from similar points.

The model

Denoting by $f(w,\xi)$ the cost function, we obtain the formulation

$$
\begin{aligned}
\min\ & f(w,\xi) \\
\text{s.t. } & F(t) \geq \Phi(t), \quad U_X \leq t < \infty, && (6) \\
& G(t) \leq \Psi(t), \quad 0 \leq t \leq L_Y, && (7) \\
& (w,\xi) \in S.
\end{aligned}
$$

Here S represents the feasible set for the decision vector (w,ξ), involving various technical restrictions, while the inequality constraints (6)–(7) are the dose limitations. Observe that $F(t)$ and $G(t)$ in these constraints depend on our decisions (w,ξ).

The key approximation

- Constraints on distribution functions are known in the literature as the first order stochastic dominance constraints (Dentcheva and Ruszczyński)
- Our constraints (6)–(7) are left and right tail versions of such conditions.
- In general, they define a non-convex feasible region, even though the random variables D_X and D_Y depend linearly on the decision variables.
- A very tight convex approximation of the feasible region is obtained by using the following family of constraints:

$$\int_T^\infty (1 - F(t))\, dt \le \int_T^\infty (1 - \Phi(t))\, dt, \quad U_X \le T < \infty,$$

$$\int_0^T G(t)\, dt \le \int_0^T \Psi(t)\, dt, \quad 0 \le T \le L_Y.$$

The key approximation

- Constraints on distribution functions are known in the literature as the first order stochastic dominance constraints (Dentcheva and Ruszczyński)
- Our constraints (6)–(7) are left and right tail versions of such conditions.
- In general, they define a non-convex feasible region, even though the random variables D_X and D_Y depend linearly on the decision variables.
- A very tight convex approximation of the feasible region is obtained by using the following family of constraints:

$$\int_T^{\infty} (1 - F(t))\, dt \le u(T), \quad U_X \le T < \infty,$$

$$\int_0^{T} G(t)\, dt \le v(T), \quad 0 \le T \le L_Y.$$

An equivalent formulation of the last relation can be obtained by employing the following identities:

$$\int_T^\infty (1 - F(t))\, dt = \frac{1}{\mathrm{card}(X)} \sum_{v \in X} \max\left(0, D_X(v) - T\right), \quad U_X \leq T < \infty,$$

$$\int_0^T G(t)\, dt = \frac{1}{\mathrm{card}(Y)} \sum_{v \in Y} \max\left(0, T - D_Y(v)\right), \quad 0 \leq T \leq L_Y.$$

Observe that the right hand sides are convex with respect to $D_X(v)$ and $D_Y(v)$, respectively.

Convex approximation

We obtain a convex approximation of our problem:

$$\min \; f(w,\xi) \text{ subject to}$$

$$\frac{1}{\mathrm{card}(X)} \sum_{v \in X} \max\left(0, D_X(v) - T\right) \le u(T), \quad U_X \le T < \infty, \tag{8}$$

$$\frac{1}{\mathrm{card}(Y)} \sum_{v \in Y} \max\left(0, T - D_Y(v)\right) \le v(T), \quad 0 \le T \le L_Y, \tag{9}$$

$$(w,\xi) \in S.$$

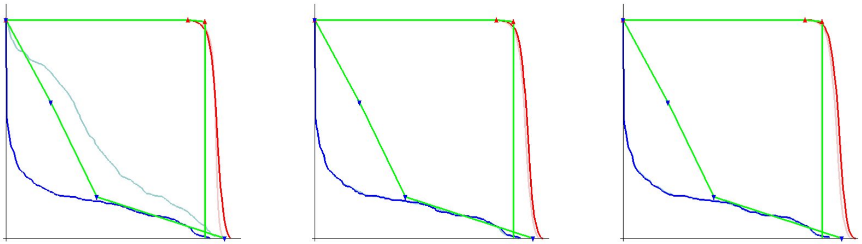
One can have several OAR's $X_1, \ldots, X_m$ with their corresponding benchmark functions $\Phi_i(t)$, $i = 1, \ldots, m$ (i.e. several u_i functions).

The algorithmic idea

- Problem appears to be a rather difficult optimization problem, because it involves a continuum of constraints, for all possible values of the target dose T.
- The idea of the method is to select a small finite set of inequalities (8) and (9), solve the resulting approximation, and add new constraints, if needed.
- The inequality at $T = U_X$

$$\frac{1}{\mathrm{card}(X)} \sum_{v \in X} \left(D_X(v) - U_X \right) \leq u(U_X)$$

  limits the *average* dose in Organ at Risk X and is certainly not sufficient to guarantee the relations (8) for all T.
- We select sets $A^l \subset X$, $l = 1, 2, \ldots,$ and we restrict from above average doses in each of the sets A^l



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

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

(g) 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.

Implementation: a graphical tool

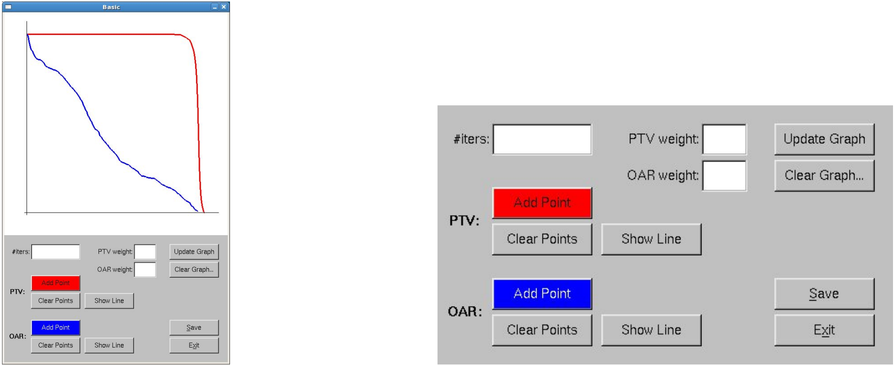


Figure: The user interface presented in our tool, including controls for constraining the PTV and OAR, limiting the number of iterations, weighting the volumes, running solves, clearing new solves and saving images.

Conclusions

- Many different optimization approaches to treat (model) uncertainties
- How much do we know about distribution of data?
- Specific models needed for these applications
- Stochastic model implementation and tool interfaces are needed

- New approach to deal with full distributions (cDVH constraints)
- Efficient implementation via cutting plane approach
- Tool available that allows piecewise linear shaping
- Can be used in conjunction with any underlying planning model

Many extensions to this work and other medical applications needed