Chapter 26

Robust optimization in radiation therapy

Albin Fredriksson

This is the author-version of a chapter published in Advances and Trends in Optimization with Engineering Applications. Please cite as:

Fredriksson, A. "Robust optimization in radiation therapy," *Advances and Trends in Optimization with Engineering Applications*, T. Terlaky, M. F. Anjos, and S. Ahmed (editors), MOS-SIAM Book Series on Optimization, SIAM, Philadelphia, 2017.

26.1 ▪ Introduction

Radiation therapy is the medical use of ionizing radiation. It is used to treat nearly two-thirds of the cancer patients in the US [1], either alone or in combination with surgery or chemotherapy. Radiation kills cells by damaging the cellular DNA. A curative treatment requires administration of a sufficiently high dose to the tumor to eradicate the clonogenic cancer cells to an extent that results in permanent tumor control. The amount of dose delivered to the surrounding healthy tissues must at the same time be restricted for the treatment not to result in adverse effects. Radiation therapy treatment planning aims to strike the right balance between the probability of tumor control and the probability of complications due to the treatment.

In intensity-modulated radiation therapy (IMRT), the dose is delivered by external beams that are incident to the patient from multiple directions. The fluences over the cross-sections of the beams are modulated. This enables the superposition of the beam doses to conform close to the target while avoiding the nearby healthy organs. The most common form of IMRT is delivered in the form of high-energy photon beams. An illustration of a photon-mediated IMRT plan is shown in Figure 26.1. For a review of IMRT treatment planning, see Bortfeld [4]. An alternate treatment modality using external beams is intensity-modulated proton therapy (IMPT). A key difference between proton and photon beams is that proton beams have a finite and controllable range with a sharp increase in the dose deposition at the end of the range—the Bragg peak. The possibility to control the range of the protons provides an additional degree of freedom compared to photon beams. Disadvantages are that the proton equipment is more costly, and that protons are more sensitive to errors. The sensitivity to errors makes robustness an especially important topic for IMPT. For a review of IMPT treatment planning, see Schwarz [26].

Because the DNA of healthy cells is repaired to a higher degree than that of malignant cells, dividing the radiation therapy treatment into treatment fractions increases the chances of achieving complication-free tumor control. Thus, treatments are typically divided into 30–40 fractions. When the patient is positioned for a treatment, there is an inevitable risk that the patient position relative to the beams

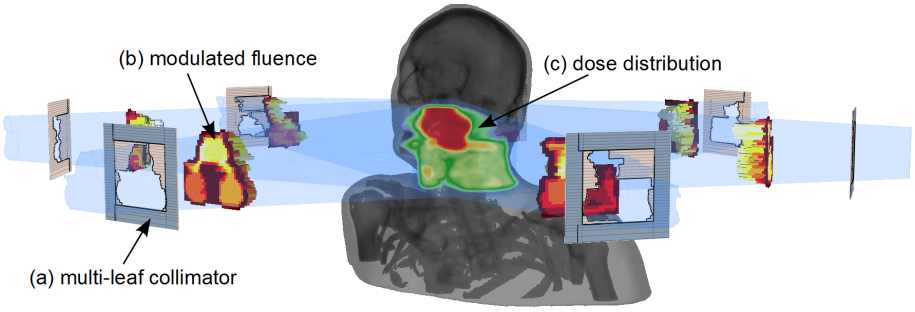


Figure 26.1. *An IMRT plan for a head and neck case. Each beam is delivered as a sequence of beam profiles shaped by a multi-leaf collimator made of a blocking material, see (a). The superposition of beam profiles enables modulation of the fluence from a single direction, see (b). The superposition of doses from multiple beams yields a 3D modulated dose, see (c). Illustration adapted from Fredriksson [14].*

differs from what was planned. The conventional way of handling this uncertainty is to deliver high dose to an enlarged target volume, while trying to protect enlarged regions encompassing the sensitive organs. Such a use of margins of course aggravates the conflicts between the high and low dose regions. More elaborate methods that explicitly take the effects of the possible errors into account during the planning of radiation therapy treatments can yield plans that are robust to errors while delivering lower doses than margin-based plans. Many of these methods are based on robust optimization.

For photon-mediated IMRT, robust optimization methods are usually employed in order to dispense with margins and instead better exploit the fact that the patient geometry changes under and between treatment fractions: if the target doses in some fractions become too low, this can be compensated by higher doses in other fractions. This can lead to solutions that are almost as robust as margin-based plans, but that deliver lower total dose.

Proton beams are highly dependent on the density of the traversed medium. If the patient is slightly misaligned or the patient anatomy has changed between planning and delivery of the treatment, the delivered dose distribution might be deformed compared to the planned one. Because errors can result in deformed dose distributions, margins do not provide the intended robustness for IMPT plans. Methods for robust IMPT planning have therefore generally been oriented towards compensating for the lack of functioning margins by ensuring that the worst case outcome becomes as beneficial as possible. This can lead to solutions that are more robust than margin-based IMPT plans while at the same time delivering lower doses to healthy structures.

26.2 ▪ Treatment planning and optimization

The goal of radiation therapy treatment planning is to find a treatment plan that has as high a probability as possible of curing the disease. Computed tomography (CT) images provide a 3D representation of the patient geometry that guides the treatment planning process. The clinician delineates the boundaries of the regions of interest (ROIs) on the CT images. The ROIs are typically the cancerous regions that are to be irradiated and the healthy organs—the organs at risk (OARs)—that are to be spared during the treatment.

An IMRT treatment depends on tens of thousands of variables representing the settings of the treatment machine, such as settings for the collimating leaves that shape the beam profile, shown as (a) in Figure 26.1. Determining all of these manually would be impractical. Therefore, optimization is used to find settings that result in a high-quality treatment. From the optimization variables, the resulting dose absorbed by the patient, measured in Gray, is calculated. The dose is scored on the basis of an

objective function that measures the quality of the dose. For dose computation and optimization to be made possible, the patient geometry is discretized into a grid of volume-elements called voxels. For a survey of IMRT treatment plan optimization, see Romeijn et al. [24].

For brevity, the following exposition will be concerned with a patient that has one target with voxels indexed by the set $\mathcal{T}$ and one OAR with voxels indexed by the set $\mathcal{O}$. The set of all voxels in the irradiated region is denoted by $\mathcal{V}$. The inclusion of additional structures is straight-forward.

26.3 ▪ Uncertainties

Among the many sources of uncertainty that can affect a radiation therapy treatment are errors in the alignment of the patient during the CT image acquisition [29], erroneously delineated ROIs [12], changes in the ROI geometries [6], and errors in the positioning of the patient relative to the beams [16]. The effects of motion on photon-mediated IMRT plans have been reviewed by Webb [31], and the effects on IMPT plans by Lomax [19].

To great generality, the uncertainties can be modeled as a random variable S picking an error scenario from the set $\mathcal{S}$ of possible scenarios. The scenario set must generally be discretized to become fit for use in tractable optimization problems, and it will henceforth be assumed that the number of elements in $\mathcal{S}$, denoted by $|\mathcal{S}|$, is finite.

The expected value of a random variable Y under a probability measure π is denoted by $\mathbb{E}_\pi[Y]$ and is given by

$$\mathbb{E}_\pi[Y] = \sum_{s\in\mathcal{S}} \pi_s y_s,$$

where y_s is the value that Y takes under scenario s (in the present case of finitely many scenarios, π is a probability mass function). The a priori probability measure, prescribing the historically measured or assumed probability to each scenario, is denoted by p. The standard deviation is only considered under p and is denoted by

$$\sigma(Y) = \sqrt{\sum_{s\in\mathcal{S}} p_s \left(y_s - \mathbb{E}_p[Y]\right)}.$$

26.4 ▪ Scenario doses

The dose distribution $d(x;s)$, illustrated by (c) in Figure 26.1, is a function of the optimization variable vector x and of the scenario s. All methods that are reviewed in the following section rely on a linear relationship between the dose distribution and the optimization variables, i.e.,

$$d(x;s) = D(s)x$$

for some matrix $D(s)$. The dose $d_i(x;s)$ to voxel i is then given by $d_i(x;s) = D_i(s)x$, where $D_i(s)$ is the ith row of $D(s)$. A common setting in which linearity holds is when the variable vector x represents a discretization of the beam fluences into area elements, illustrated by (b) in Figure 26.1. Each column of the matrix $D(s)$ is then the vector of voxel dose depositions that results when the fluence is set to unity in one of the beam fluence area elements. In IMPT, such a linear representation corresponds to deliverable machine settings, whereas in photon-mediated IMRT, the beam fluences are only indirectly controllable via superposition of multi-leaf collimator shapes, see (a) in Figure 26.1. A conversion step from beam fluences into deliverable machine settings is thus required when the linear representation is used.

26.5 ▪ Robust optimization methods used in treatment planning

This section provides a survey of robust optimization methods that have been used to generate robust treatment plans. The robust optimization methods are presented in a distilled form that highlights their distinguishing features. Several natural variations and extensions of the methods are possible (and have been used): e.g., all considered methods can be modified in order to emphasize the nominal scenario, corresponding to no error, by the addition of terms penalizing nominal dose deviations to the objective, or to emphasize the reduction of the total dose by the addition of terms penalizing any dose to $\mathcal{V}$. Note that methods such as stochastic programming have also been used to achieve robustness [28].

26.5.1 ▪ Distributional robustness via linear programming

Lung cancer treatments are affected by intrafraction motion—that patient motions occur during the delivery of the treatment [17]. The effect on the dose of intrafraction motion is an averaging of the doses $d(x;s)$ delivered during different phases of the motion, modeled by the scenario s, weighted by the proportion of time, p_s, spent in each phase. The use of a margin that covers the position of the target over all phases can yield fully robust target coverage. However, some phases can be less favorable than others. If the optimization algorithm is informed about how the patient geometry changes over the phases, it may be able to deliver a sufficient dose to the target while reducing the dose to healthy tissues.

Because intrafraction motion results in an averaging of the phase doses, optimizing an averaged dose distribution may seem like the path forward. Doing so, however, results in heterogeneous dose distributions, the robustness of which is highly dependent on that the averaging realized during the delivery of the treatment is the same as that applied during planning. This sensitivity to model uncertainty has been found under various alterations of the treatment plan optimization problem [8, 13, 28].

Chan et al. [8] address the problem that the time spent in each phase can differ from the times assumed during the treatment planning. They introduce a method in which the fraction of time spent in each phase is considered uncertain and only known to lie within some bounds. The fraction of time π_s that is spent in phase s is bound by $a_s \leq \pi_s \leq b_s$, where the bounds are such that $0 \leq a_s \leq p_s \leq b_s \leq 1$ holds for the a priori distribution p. An illustration of this is shown in Figure 26.2. The uncertainty

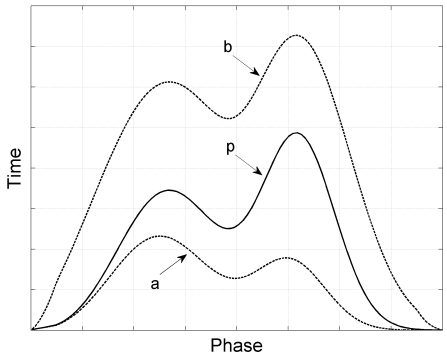


Figure 26.2. *An illustration of the a priori distribution p of times spent in each phase s from the set $\mathcal{S}$ and the bounds a and b on the true distribution π.*

set $\mathcal{U}$ of possible fractions of time spent in the phases is then defined by

$$\mathcal{U} = \left\{ \pi \in \mathbb{R}^{|\mathcal{S}|} : a \leq \pi \leq b, \sum_{s \in \mathcal{S}} \pi_s = 1 \right\}. \tag{26.1}$$

They minimize the total dose delivered under the a priori distribution p subject to constraints requiring the dose to each target voxel to exceed a reference dose level under all distributions π from $\mathcal{U}$. This is formulated mathematically as

$$\begin{array}{ll} \underset{x \geq 0}{\text{minimize}} & \sum_{i \in \mathcal{V}} \mathbb{E}_p [d_i(x;S)] \\ \text{subject to} & \mathbb{E}_\pi [d_i(x;S)] \geq \delta_i \quad \forall \pi \in \mathcal{U}, \quad \forall i \in \mathcal{T}. \end{array} \tag{26.2}$$

Note that optimization of expectations is the fundamental formulation of the intrafraction motion problem, and that (26.2) introduces robustness by the constraints for all π in $\mathcal{U}$. When $a \neq b$, this problem has infinitely many constraints because $\mathcal{U}$ contains infinitely many points. However, the objective and the constraints are linear, and the uncertainty set $\mathcal{U}$ is polyhedral, so there exists a robust counterpart to (26.2) that is a linear program with finitely many constraints (see Ben-Tal and Nemirovski [2] and Bertsimas and Sim [3]). To derive the robust counterpart, the first observation to make is that for each $i \in \mathcal{T}$, if the constraint holds for the worst π in $\mathcal{U}$, then it holds for all of them. This yields the equivalence

$$\mathbb{E}_\pi [d_i(x;S)] \geq \delta_i \quad \forall \pi \in \mathcal{U} \quad \Leftrightarrow \quad \min_{\pi \in \mathcal{U}} \left\{ \sum_{s \in \mathcal{S}} \pi_s d_i(x;s) \right\} \geq \delta_i \tag{26.3}$$

where it has been used that $\mathbb{E}_\pi [d_i(x;S)] = \sum_{s \in \mathcal{S}} \pi_s d_i(x;s)$. By strong duality of linear programming, the value of the minimum in (26.3) equals the value of its dual

$$\max_{\lambda_i, \mu_i, \nu_i} \left\{ \begin{array}{l} \lambda_i + a^T \mu_i - b^T \nu_i : \\ \lambda_i + \mu_{i,s} - \nu_{i,s} \leq d_i(x;s) \quad \forall s \in \mathcal{S} \\ \mu_i, \nu_i \geq 0 \end{array} \right\}, \tag{26.4}$$

where λ_i is a scalar whereas μ_i and ν_i are vectors of length $|\mathcal{S}|$. The value of any $\lambda_i + a^T \mu_i - b^T \nu_i$ satisfying the constraints in (26.4) is bounded from above by the minimum in (26.3), so the maximum operation can be discarded and the robust counterpart of (26.2) can be formulated as

$$\begin{array}{ll} \underset{x, \lambda, \mu, \nu}{\text{minimize}} & \sum_{i \in \mathcal{V}} \mathbb{E}_p [d_i(x;S)] \\ \text{subject to} & \lambda_i + a^T \mu_i - b^T \nu_i \geq \delta_i \quad \forall i \in \mathcal{T} \\ & \lambda_i + \mu_{i,s} - \nu_{i,s} \leq d_i(x;s) \quad \forall s \in \mathcal{S}, \quad \forall i \in \mathcal{T} \\ & x, \mu, \nu \geq 0, \end{array} \tag{26.5}$$

which is a linear programming problem with finitely many constraints.

The plans resulting from this method provide target coverage under all distributions π in $\mathcal{U}$ of times spent in each phase. At the same time, they have been found to deliver substantially lower total dose than plans planned with margins [8].

26.5.2 ▪ Probabilistic robustness via second-order cone programming

During the course of a radiation therapy treatment, the patient geometry changes. The interfraction motion—the changes to the geometry occurring between the treatment fractions—is due to that the

patient may be positioned differently during each treatment fraction, and that the patient anatomy changes due to organ motion, tumor shrinkage, and weight loss.

The total dose delivered to the patient is the sum of n fraction doses $d(x;S_j)$ for $j=1,\ldots,n$, where the random variable S_j selects the scenario in fraction j. Under the assumption that $S_1,\ldots,S_n$ are independent and identically distributed, distributed like S, and using that the number of fractions n is fairly large (typically 30–40), the central limit theorem implies that the total dose to a given voxel is approximately normally distributed.

An approach that has been used to find plans that are robust to interfraction motion and setup errors is to optimize towards achieving high probability that the dose to each voxel, considered individually, is above or below the desired dose level. For an OAR voxel i, achieving a low probability ϵ that the total dose exceeds some dose level $\delta^{\mathcal{O}}$ can be formulated as the constraint

$$\mathbb{P}\left(\sum_{j=1}^{n} d_i(x;S_j) > \delta^{\mathcal{O}}\right) \leq \epsilon. \tag{26.6}$$

When the total dose $\sum_{j=1}^{n} d_i(x;S_j)$ is assumed to be normally distributed, it is fully characterized by its expected value $\mu = n\mathbb{E}_p[d_i(x;S)]$ and standard deviation $\sigma = \sqrt{n}\sigma(d_i(x;S))$, and (26.6) is equivalent to

$$\mathbb{P}\big(Z > (\delta^{\mathcal{O}} - \mu)/\sigma\big) \leq \epsilon \quad \Leftrightarrow \quad (\delta^{\mathcal{O}} - \mu)/\sigma \geq z_{1-\epsilon},$$

where Z denotes a standard normal deviate and $z_{1-\epsilon}$ is chosen such that $\mathbb{P}(Z > z_{1-\epsilon}) = \epsilon$. This expression can be rewritten as

$$\mu + z_{1-\epsilon}\sigma \leq \delta^{\mathcal{O}}. \tag{26.7}$$

Similarly, requiring a target voxel i to achieve a lower probability than ϵ of falling below the level $\delta^{\mathcal{T}\text{low}}$ is equivalent to requiring that

$$\mu - z_{1-\epsilon}\sigma \geq \delta^{\mathcal{T}\text{low}}. \tag{26.8}$$

The left-hand sides of (26.8) and (26.7) can be used to define modified dose distributions

$$\begin{aligned} d_i^{\text{low}}(x) &= \left(n\mathbb{E}_p[d_i(x;S)] - z_{1-\epsilon}\sqrt{n}\sigma(d_i(x;S))\right)_+, \\ d_i^{\text{high}}(x) &= n\mathbb{E}_p[d_i(x;S)] + z_{1-\epsilon}\sqrt{n}\sigma(d_i(x;S)), \end{aligned} \tag{26.9}$$

where $n\mathbb{E}_p[d_i(x;S)]$ and $\sqrt{n}\sigma(d_i(x;S))$ have been substituted for respectively μ and σ, and the positive part is taken of $d_i^{\text{low}}(x)$ because dose cannot be negative.

Chu et al. [11] relax the constraints (26.7) and (26.8) into objectives by penalizing the worst violation of these constraints over the voxels for each structure. They solve the optimization problem

$$\begin{aligned} \underset{x\geq 0}{\text{minimize}} \quad & w^{\mathcal{T}\text{low}} \max_{i\in\mathcal{T}} \left\{\left(\delta^{\mathcal{T}\text{low}} - d_i^{\text{low}}(x)\right)_+\right\} + \\ & w^{\mathcal{T}\text{high}} \max_{i\in\mathcal{T}} \left\{\left(d_i^{\text{high}}(x) - \delta^{\mathcal{T}\text{high}}\right)_+\right\} + \\ & \sum_{s\in\mathcal{S}} w_s^{\mathcal{T}} \max_{i\in\mathcal{T}} \left\{\left(\delta_s^{\mathcal{T}} - d_i(x;s)\right)_+\right\} + \\ & w^{\mathcal{O}} \max_{i\in\mathcal{O}} \left\{\left(d_i^{\text{high}}(x) - \delta^{\mathcal{O}}\right)_+\right\}, \end{aligned} \tag{26.10}$$

where $w^{\mathcal{T}\text{low}}$, $w^{\mathcal{T}\text{high}}$, $w_s^{\mathcal{T}}$, and $w^{\mathcal{O}}$ are penalty weights for failing to reach the minimum target dose level $\delta^{\mathcal{T}\text{low}}$, for exceeding the maximum target dose level $\delta^{\mathcal{T}\text{high}}$, for failing to reach the minimum target dose level $\delta_s^{\mathcal{T}}$ in scenario s, and for exceeding the maximum OAR dose level $\delta^{\mathcal{O}}$, respectively.

The nonlinear program (26.10) can be rearranged into a second-order cone problem, which is more practical to solve: The maxima and positive parts in (26.10) can be avoided by the introduction of auxiliary linear variables and constraints (see, e.g., Vanderbei [30]). The modified dose distributions contain standard deviation terms that require second-order cone constraints. The objective in (26.10) becomes reducible to a linear function when auxiliary linear variables y^{low} and y^{high} are substituted for $d_i^{\text{low}}(x)$ and $d_i^{\text{high}}(x)$, and constraints on the form $y_i^{\text{low}} \leq d_i^{\text{low}}(x)$ and $y_i^{\text{high}} \geq d_i^{\text{high}}(x)$ are introduced. These constraints are equivalent to

$$z_{1-\epsilon}\sqrt{n}||RA_i x||_2 \leq n\mathbb{E}_p[d_i(x;S)] - y_i^{\text{low}} \quad \text{and} \quad z_{1-\epsilon}\sqrt{n}||RA_i x||_2 \leq y_i^{\text{high}} - n\mathbb{E}_p[d_i(x;S)],$$

where the matrix A_i consists of the rows $D_i(s)$ of the dose matrix for all scenarios s in $\mathcal{S}$ and $R = P^{1/2}(I - ep^T)$, in which $P = \text{diag}(p)$, $I \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ is the identity matrix, and $e \in \mathbb{R}^{|\mathcal{S}|}$ is the vector of all ones. After the introduction of auxiliary variables and constraints, problem (26.10) can thus be equivalently formulated as a second-order cone programming problem, see Chu et al. [11] for details.

The formulation of the second-order cone constraints indicates a relationship to robust programming with ellipsoidal uncertainty sets (see Ben-Tal and Nemirovski [2]). Indeed, the program may be interpreted in terms of robust linear programming. The dose to voxel i is then modeled to be given by $a_i^T x$ for some uncertain vector a_i from the ellipsoidal uncertainty set $\mathcal{U}_i$, given by

$$\mathcal{U}_i = \left\{ n\mathbb{E}_p[D_i(S)] + W_i u_i : u_i^T u_i \leq 1 \right\}.$$

For each voxel i from the set $\mathcal{O}$ of OAR voxels, the optimization problem is subject to constraints

$$a_i^T x \leq \delta^{\mathcal{O}} \quad \forall a_i \in \mathcal{U}_i \quad \Leftrightarrow \quad n\mathbb{E}_p[d(x;S)] + \max_{u:u^T u\leq 1} \left\{ u^T W_i^T x \right\} \leq \delta^{\mathcal{O}}.$$

The maximum over u equals $||W_i^T x||_2$. Thus, when $z_{1-\epsilon}\sqrt{n}RA_i$ has been substituted for W_i^T, this constraint is seen to be equivalent to (26.7).

Ólafsson and Wright [22] use a model similar to the model (26.10) of Chu et al. [11], but take into account uncertainty due to Monte Carlo dose matrix computation in addition to organ motion. They solve the problem

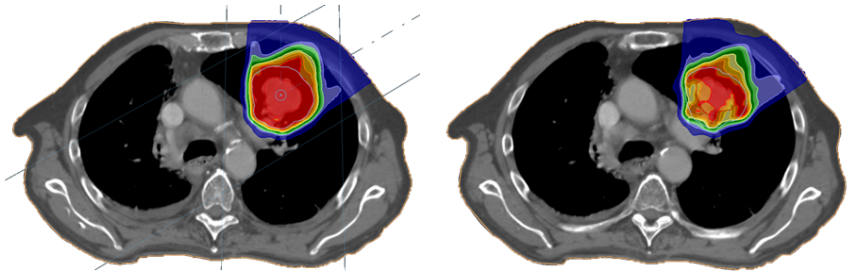
$$\begin{aligned}
\underset{x\geq 0}{\text{minimize}} \quad & \sum_{i\in\mathcal{T}} w_i^{\mathcal{T}\text{low}} \left(\delta_i^{\mathcal{T}} - n\mathbb{E}_p[d_i(x;S)] \right)_+ + \\
& \sum_{i\in\mathcal{T}} w_i^{\mathcal{T}\text{high}} \left(n\mathbb{E}_p[d_i(x;S)] - \delta_i^{\mathcal{T}} \right)_+ + \\
& \sum_{i\in\mathcal{O}} w_i^{\mathcal{O}} \left(d_i^{\text{high}}(x) - \delta_i^{\mathcal{O}} \right)_+ \\
\text{subject to} \quad & d_i^{\text{low}}(x) \geq \delta_i^{\mathcal{T}\text{low}}, \qquad \forall i \in \mathcal{T} \\
& d_i^{\text{high}}(x) \leq \delta_i^{\mathcal{T}\text{high}}, \qquad \forall i \in \mathcal{T}.
\end{aligned} \tag{26.11}$$

Like (26.10), this program can be equivalently formulated as a second-order cone program.

The plans optimized with probabilistic robustness requirements were found to provide similar robustness with respect to target coverage as planning with margins (when evaluated against the assumed probability distribution p), while reducing the total dose and the doses to the OARs [11, 22].

26.5.3 ▪ Worst case robustness via nonlinear programming

For margins to provide robustness against errors, the effects of the errors must be well approximated by rigid translations of the dose distribution. If not, there is no guarantee that the dose distribution will cover the target after an error has occurred. The beam doses of IMPT treatments are often drastically



(a) Margin-based, planned setup (b) Margin-based, shifted setup

Figure 26.3. *Plan optimized to deliver dose to an enlarged target region. (a): the planned dose distribution covers the target; (b): the dose distribution resulting when the patient is misaligned yields too low dose to the target.*

deformed as a result of errors. Figure 26.3 shows the effects of a setup error on the dose distribution of a plan reached by margin-based planning. The sensitivity to errors of IMPT has motivated the development of methods that generate plans that are robust in the sense intended by margin-based planning. Such methods do not exploit that the target may be located differently in different treatment fractions, such as the methods described in Sections 26.5.1 and 26.5.2. Instead, these methods try to achieve target coverage and OAR sparing under the, say, 95 % most probable error scenarios that can occur, even if the same scenario occurs in each fraction. This is similar to how margins are specified to provide a given probability of target coverage. Thus, this type of method can be thought of as performing inverse planning of margins, because instead of specifying margins in order to achieve robustness against uncertainties, the treatment planner specifies the uncertainties and leaves it to the optimization algorithm to determine where to deposit dose in order to achieve robust plans.

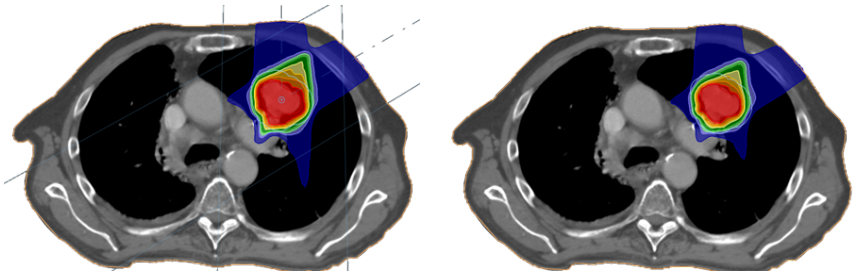
To this end, Fredriksson et al. [15] consider a nonlinear treatment plan optimization problem and minimize the penalty of the objective function under the worst case scenario. The problem is formulated as

$$\underset{x\geq 0}{\text{minimize}} \quad \max_{s\in\mathcal{S}} \left\{ w^{\mathcal{T}} f^{\mathcal{T}}(d(x;s)) + w^{\mathcal{O}} f^{\mathcal{O}}(d(x;s)) \right\}, \tag{26.12}$$

where $f^{\mathcal{T}}$ and $f^{\mathcal{O}}$ are nonlinear functions penalizing deviation from the planning goals of the target and OAR doses, respectively, such as those described by Oelfke and Bortfeld [21]. Because of the maximum in the objective, this problem is not continuously differentiable, which can result in problems with convergence for gradient-based optimization solvers. This disadvantage is alleviated when the problem is posed on epigraph form, i.e., when an auxiliary variable λ is substituted for the objective, and this variable is constrained by $\lambda \geq w^{\mathcal{T}} f^{\mathcal{T}}(d(x;s)) + w^{\mathcal{O}} f^{\mathcal{O}}(d(x;s))$ for all scenarios s in $\mathcal{S}$. Chen et al. [10] use a similar method in a multiobjective optimization framework, but apply the maximum over $\mathcal{S}$ to each objective constituent $f^{\mathcal{T}}$ and $f^{\mathcal{O}}$ individually.

It has been empirically found that the worst scenarios are typically at the boundary of the uncertainty region [7, 9, 13], i.e., that the worst scenarios are those corresponding to the largest deviations from the planning image. This observation can be used to reduce the computational cost of solving (26.12), by removing non-boundary scenarios from $\mathcal{S}$.

Minimizing the worst case penalty has been found to result in more robust target coverage than heuristic methods using margins, uniform beam doses, and planning with the densities in the target region overridden, while at the same time yielding lower doses to healthy structures [15]. Figure 26.4 shows the effects of a setup error on the dose distribution of a plan reached by worst case optimization according to (26.12).



(a) Worst case optimized, planned setup (b) Worst case optimized, shifted setup

Figure 26.4. *Plan optimized to deliver dose to the target in the worst case scenario. (a): the planned dose distribution covers the target; (b): the dose distribution resulting when the patient is misaligned still covers the target.*

26.5.4 ▪ Voxel-wise worst case robustness via linear and nonlinear programming

The concept of worst case dose distributions was introduced by Lomax et al. [20] as a tool to evaluate the robustness of IMPT plans. They define modified dose distributions d^{low} and d^{high} according to

$$d_i^{\text{low}}(x) = \min_{s\in\mathcal{S}} \{d_i(x;s)\} \quad \text{and} \quad d_i^{\text{high}}(x) = \max_{s\in\mathcal{S}} \{d_i(x;s)\}.$$

How much a treatment plan can deteriorate is bounded by how well d^{low} meets the minimum target dose goals and how well d^{high} meets the maximum OAR dose goals. Note the similarity to the modified dose distributions (26.9) used to bound the probability of over- and underdosage.

Optimization of the worst case dose distributions was first considered by Unkelbach et al. [27] and later by Pflugfelder et al. [23]. The optimization problem using the worst case dose distribution takes the form

$$\underset{x\geq 0}{\text{minimize}} \quad w^{\mathcal{T}^{\text{low}}} f^{\mathcal{T}^{\text{low}}}(d^{\text{low}}(x)) + w^{\mathcal{T}^{\text{high}}} f^{\mathcal{T}^{\text{high}}}(d^{\text{high}}(x)) + w^{\mathcal{O}} f^{\mathcal{O}}(d^{\text{high}}(x)), \tag{26.13}$$

where $f^{\mathcal{T}^{\text{low}}}$ penalizes deviations from the planning goals specifying what the target dose should exceed, while $f^{\mathcal{T}^{\text{high}}}$ and $f^{\mathcal{O}}$ penalize deviations from the goals specifying what respectively the target and OAR doses should fall below. For the program to be convex, $f^{\mathcal{T}^{\text{low}}}$ should be non-increasing and convex whereas $f^{\mathcal{T}^{\text{high}}}$ and $f^{\mathcal{O}}$ should be non-decreasing and convex (see, e.g., Boyd and Vandenberghe [5, Section 3.2.4]).

Unkelbach et al. [27] solve (26.13) on the form of a robust linear programming problem. They consider errors that are independent per beam, which results in a scenario set $\mathcal{S}$ of exponential size in the number of beams. By using linear programming duality in a way similar to that transforming (26.2) into (26.5), they reduce the number of constraints used to specify d^{low} and d^{high} from an exponential into a polynomial number.

Pflugfelder et al. [23] optimize the worst case dose distributions in a nonlinear programming setting. They apply the penalty functions to d^{low} and d^{high} directly, without introducing auxiliary variables and constraints to alleviate the discontinuous gradients that result from the minima and maxima. These discontinuities have not been observed to pose any practical problems. A plausible explanation is that the number of voxels is generally large so when one voxel is affected by a discontinuity, many others are not.

The nonlinear functions $f^{\mathcal{T}^{\text{high}}}$, $f^{\mathcal{T}^{\text{low}}}$, and $f^{\mathcal{O}}$ that have been employed in the literature often penalize dose deviations linearly or quadratically, i.e., are on the form

$$f^{\mathcal{T}^{\text{low}}}(d^{\text{low}}) = \sum_{i\in\mathcal{T}} \left(\delta^{\mathcal{T}^{\text{low}}} - d_i^{\text{low}}\right)_+^q \quad \text{and} \quad f^{\mathcal{O}}(d^{\text{high}}) = \sum_{i\in\mathcal{O}} \left(d_i^{\text{high}} - \delta^{\mathcal{O}^{\text{high}}}\right)_+^q,$$

where q is 1 or 2 (and with $f^{\mathcal{T}^{\text{high}}}$ on the same form as $f^{\mathcal{O}}$, but for the target region). This hints at the close relationship between the voxel-wise worst case formulation (26.13) and the probabilistic formulations (26.10) and (26.11) of Chu et al. [11] and Ólafsson and Wright [22]. The main difference between these formulations lies in the choice of risk measure used to define the modified dose distributions d^{low} and d^{high}: the probabilistic formulations use mean-deviation risk measures, whereas the voxel-wise worst case formulation uses a max-norm risk measure. See Ruszczyński and Shapiro [25] for details concerning general optimization of risk measures.

Optimization of the worst case dose distributions has been found to result in more robust target coverage and higher doses to healthy structures than optimization of the nominal dose distribution (without a margin) [23, 27]. When compared to conventional planning with margins, it not only resulted in more robust target coverage, but also in lower doses to healthy structures [18]. The worst case dose distribution has, however, been found to be overly conservative for evaluation compared to the worst case scenario [7].

26.6 ▪ Conclusion

Robust optimization has great potential for improving radiation therapy treatments. For photon-mediated IMRT, robust optimization methods like those in Sections 26.5.1 and 26.5.2 can yield treatment plans that exploit the fact that the patient geometry changes during the course of the treatment. Although margin-based plans can provide more robust target coverage, they do so by extending the high-dose region, which is arguably a crude means of increasing the robustness. The cost of doing so is exhibited by the plans generated by robust optimization methods, which provide almost as robust target coverage, but reduce the doses to the OARs as well as the total dose substantially.

For IMPT, where margins are often insufficient to generate robust plans, more advanced methods are indispensable. Here, robust optimization not only provides lower doses to the healthy structures, but moreover provides more robust target coverage than conventional methods. The robust optimization methods that have been applied to IMPT planning, including those in Sections 26.5.3 and 26.5.4, have been constructed to provide the type robustness intended by margins. There is likely even more to be gained by methods that take into account that the patient anatomy changes during the treatment.

In conclusion, the success of robust optimization methods in radiation therapy points to the benefits of (a) including additional information during the optimization, and (b) utilizing the information in relevant ways. Still, robust optimization of radiation therapy treatment plans remains an active area of research; there is more information to be utilized and further ways to do so.

Bibliography

[1] American Medical Association. *Physician Characteristics and Distribution in the U.S.* Americal Medical Association, Atlanta, Georgia, 2010. (Cited on p. 1)

[2] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Oper. Res. Lett.*, 25(1):1–13, 1999. (Cited on pp. 5, 7)

[3] D. Bertsimas and M. Sim. The price of robustness. *Oper. Res.*, 52(1):35–53, 2004. (Cited on p. 5)

[4] T. Bortfeld. IMRT: a review and preview. *Phys. Med. Biol.*, 51(13):R363–R379, 2006. (Cited on p. 1)

[5] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, United Kingdom, 2004. (Cited on p. 9)

[6] K. Britton, Y. Takai, M. Mitsuya, K. Nemoto, Y. Ogawa, and S. Yamada. Evaluation of inter- and intrafraction organ motion during intensity modulated radiation therapy (IMRT) for localized prostate cancer measured by a newly developed on-board image-guided system. *Radiat. Med.*, 23(1):14–24, 2005. (Cited on p. 3)

[7] M. Casiraghi, F. Albertini, and A. Lomax. Advantages and limitations of the 'worst case scenario' approach in IMPT treatment planning. *Phys. Med. Biol.*, 58(5):1323–1339, 2013. (Cited on pp. 8, 10)

[8] T. Chan, T. Bortfeld, and J. Tsitsiklis. A robust approach to IMRT optimization. *Phys. Med. Biol.*, 51(10):2567–2583, 2006. (Cited on pp. 4, 5)

[9] T. Chan, J. Tsitsiklis, and T. Bortfeld. Optimal margin and edge-enhanced intensity maps in the presence of motion and uncertainty. *Phys. Med. Biol.*, 55(2):515–533, 2010. (Cited on p. 8)

[10] W. Chen, J. Unkelbach, A. Trofimov, T. Madden, H. Kooy, T. Bortfeld, and D. Craft. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. *Phys. Med. Biol.*, 57(3):591–608, 2012. (Cited on p. 8)

[11] M. Chu, Y. Zinchenko, S. Henderson, and M. Sharpe. Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. *Phys. Med. Biol.*, 50(23):5463–5477, 2005. (Cited on pp. 6, 7, 10)

[12] C. Fiorino, M. Reni, A. Bolognesi, G. Cattaneo, and R. Calandrino. Intra-and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiother. Oncol.*, 47(3):285–292, 1998. (Cited on p. 3)

[13] A. Fredriksson. A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization. *Med. Phys.*, 39(8):5169–5181, 2012. (Cited on pp. 4, 8)

[14] A. Fredriksson. *Robust optimization of radiation therapy accounting for geometric uncertainty*. PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2013. (Cited on p. 2)

[15] A. Fredriksson, A. Forsgren, and B. Hårdemark. Minimax optimization for handling range and setup uncertainties in proton therapy. *Med. Phys.*, 38(3):1672–1684, 2011. (Cited on p. 8)

[16] C. Hurkmans, P. Remeijer, J. Lebesque, and B. Mijnheer. Set-up verification using portal imaging; review of current clinical practice. *Radiother. Oncol.*, 58(2):105–120, 2001. (Cited on p. 3)

[17] K. Langen and D. Jones. Organ motion and its management. *Int. J. Radiat. Oncol. Biol. Phys.*, 50(1):265–278, 2001. (Cited on p. 4)

[18] W. Liu, X. Zhang, Y. Li, and R. Mohan. Robust optimization of intensity modulated proton therapy. *Med. Phys.*, 39(2):1079–1091, 2012. (Cited on p. 10)

[19] A. Lomax. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. *Phys. Med. Biol.*, 53(4):1043–1056, 2008. (Cited on p. 3)

[20] A. Lomax, E. Pedroni, H. Rutz, and G. Goitein. The clinical potential of intensity modulated proton therapy. *Z. Med. Phys.*, 14(3):147–152, 2004. (Cited on p. 9)

[21] U. Oelfke and T. Bortfeld. Inverse planning for photon and proton beams. *Med. Dosim.*, 26(2):113–124, 2001. (Cited on p. 8)

[22] A. Ólafsson and S. Wright. Efficient schemes for robust IMRT treatment planning. *Phys. Med. Biol.*, 51(21):5621–5642, 2006. (Cited on pp. 7, 10)

[23] D. Pflugfelder, J. Wilkens, and U. Oelfke. Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy. *Phys. Med. Biol.*, 53(6):1689–1700, 2008. (Cited on pp. 9, 10)

[24] H. Romeijn and J. Dempsey. Intensity modulated radiation therapy treatment plan optimization. *TOP*, 16(2):215–243, 2008. (Cited on p. 3)

[25] A. Ruszczyński and A. Shapiro. Optimization of risk measures. In *Probabilistic and randomized methods for design under uncertainty*, pages 119–157. Springer, London, United Kingdom, 2006. (Cited on p. 10)

[26] M. Schwarz. Treatment planning in proton therapy. *Eur. Phys. J. PLUS*, 126(7):1–10, 2011. (Cited on p. 1)

[27] J. Unkelbach, T. Chan, and T. Bortfeld. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. *Phys. Med. Biol.*, 52(10):2755–2773, 2007. (Cited on pp. 9, 10)

[28] J. Unkelbach and U. Oelfke. Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions. *Phys. Med. Biol.*, 49(17):4005–4029, 2004. (Cited on p. 4)

[29] M. van Herk. Errors and margins in radiotherapy. *Semin. Radiat. Oncol.*, 14(1):56–64, 2004. (Cited on p. 3)

[30] R. Vanderbei. *Linear programming: Foundations and extensions*. Kluwer, Boston, Massachusetts, 1996. (Cited on p. 7)

[31] S. Webb. Motion effects in (intensity modulated) radiation therapy: a review. *Phys. Med. Biol.*, 51(13):R403–R425, 2006. (Cited on p. 3)