Math. Program., Ser. B 101: 387–413 (2004)

Digital Object Identifier (DOI) 10.1007/s10107-004-0530-y

Michael C. Ferris · Meta M. Voelker

Fractionation in radiation treatment planning

Received: March 6, 2002 / Accepted: April 6, 2004
Published online: 7 July 2004 – © Springer-Verlag 2004

Abstract. Radiotherapy treatment is often delivered in a fractionated manner over a period of time. Emerging delivery devices are able to determine the actual dose that has been delivered at each stage facilitating the use of adaptive treatment plans that compensate for errors in delivery. We formulate a model of the day-to-day planning problem as a stochastic program and exhibit the gains that can be achieved by incorporating uncertainty about errors during treatment into the planning process. Due to size and time restrictions, the model becomes intractable for realistic instances. We show how heuristics and neuro-dynamic programming can be used to approximate the stochastic solution, and derive results from our models for realistic time periods. These results allow us to generate practical rules of thumb that can be immediately implemented in current planning technologies.

Key words. Fractionation – Adaptive Radiation Therapy – Dynamic Programming

1. Introduction

In radiation therapy, ionizing radiation is delivered to cancerous tissue, damaging the DNA and interfering with the ability of the cancerous cells to grow and divide [55, 63]. Healthy cells are also damaged by the radiation but they are more able to repair the damage and return to normal function. Since both cancerous and healthy cells are affected by radiation, dose distributions need to be designed that expose the tumor to enough radiation for treatment while simultaneously avoiding excessive radiation to surrounding healthy tissue and, in particular, to nearby organs. By cross-firing beams from a number of directions, the radiation damage is concentrated in the patient's tumor but is less severe and more widely distributed in the surrounding healthy regions.

Given a particular delivery mechanism, a treatment plan corresponds to settings of the machine that facilitate the delivery of the target dose distribution. Optimization techniques can be used to design such plans, see [4, 6, 46, 57] and the included references. Schematically, the problems have the form:

$$\min_{x,y} f(x - T) \text{ subject to } x = \sum_{a \in \mathcal{A}} D_a y_a,\ y \geq 0,\ (x, y) \in X.$$

The objective f typically measures the weighted difference between the dose delivered x and a target (idealized) distribution T, where the dose is a superposition of dose

M.C. Ferris: Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, WI 53706, USA. e-mail: ferris@cs.wisc.edu

M.M. Voelker: Alphatech, Inc., 3811 North Fairfax Drive, Suite 500, Arlington, VA 22203, USA. e-mail: meta.voelker@dc.alphatech.com

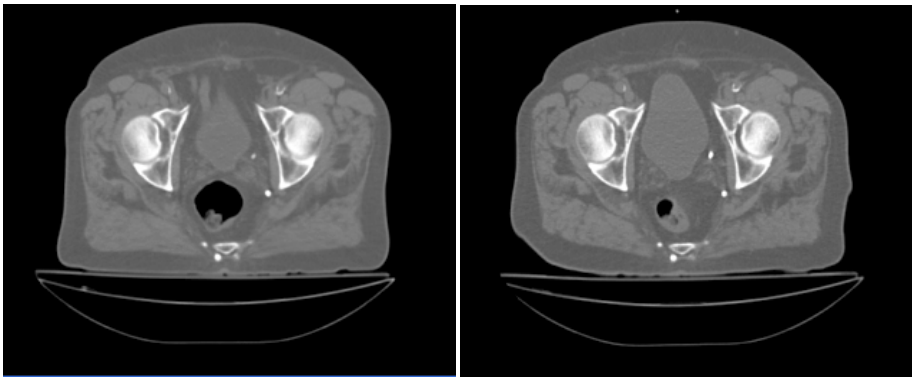
distributions D_a, weighted by delivery times y_a. In many cases, it is desirable to use different penalties for under and overdosing, and not to penalize small errors at all; such modifications can be easily incorporated into the objective function using standard modeling techniques. In conformal radiation therapy [55, 63], the set $\mathcal{A}$ runs over a collection of angles from which radiation is fired, whereas in IMRT [3, 62], the set $\mathcal{A}$ represents pencil beams delivered through a multileaf collimator. A variety of constraints are represented by $(x, y) \in X$: the problems are complicated by the complexities of the delivery mechanisms and the physicians' need to impose quality assurance measures on the resulting dose distribution. Furthermore, there is a large amount of data D_a that needs to be manipulated to obtain sufficient detail of the dose on the target area. More information is provided in [55].

While these problems remain at the forefront of cancer treatment planning and many techniques have been proposed for the large varieties of machines (for example, see [7, 10, 11, 14, 16, 28, 41, 54, 58, 33, 29, 30, 45, 34, 32, 31, 57, 12, 8, 44, 9, 10, 56, 36, 61, 49]), many of which take from minutes to hours to solve, we will not focus on this aspect of the problem. Instead, our focus in this paper is on the day-to-day planning problem and the stochastic issues that are inherent in such problems. Once a treatment plan is decided on, it is delivered to the patient in a fractionated manner (typically 5 treatments per week for a period of 4-9 weeks). In current clinical practice, the total dose is divided equally between the fractionated treatments. Generally, the use of fractionation is known to increase the probability of controlling the tumor and decrease damage to normal tissues surrounding the tumor. However, movement of the patient or the internal organs during or between treatment sessions can result in failure to deliver adequate radiation to the tumor (leading to recurrence of the disease) or to painful and debilitating damage to surrounding tissues [39]. Even before the treatment starts, the physician faces uncertainty in prescribing the radiation dose, because the extent of the cancerous cells is often not known with complete precision. If the tumor does not lie exactly in the region specified when designing the treatment plan, some cancerous regions may be dramatically underdosed. Such "cold spots" can lead to the survival of clonogenic tumor cells and ultimately to recurrence of the patient's disease [5, 39].

Displacement of the tumor from the target region and uncertainty about its exact position and extent are caused by several factors:

- A mistake in interpretation of the imaged data, or the presence of microscopic extensions of the tumor that cannot be viewed with current imaging technology;
- Movement and/or shrinkage of the patient's internal organs from day to day, between treatment sessions;
- Errors in setting up (registering) the patient on the treatment device [13, 59];
- Delivery errors or differences between planned and actual delivered dosages;
- Patient movement (usually due to breathing) while the dose is being delivered.

For example, in breast cancer cases, the breast cannot be positioned exactly the same from day to day. Also, breathing motion can move the breast out of the treatment field if this motion is not accounted for properly in the planning and delivery. As a second example, we illustrate a prostate cancer case in Figure 1. If the patient drinks a glass of



(a) Fraction 1 (b) Fraction 9

Fig. 1. CT scan for a prostate case; the black region represents the rectum, the grey area above it between the bone structures is the bladder. Both sensitive structures (bladder and rectum) move radically between treatment periods

water, the bladder may become larger and push on the prostate, changing the relative shapes and locations of these organs. The resulting delivery can then compromise the bladder, and more crucially, allow portions of the cancerous tissue to receive insufficient irradiation.

The traditional method to deal with uncertainty and patient movement in radiation treatment is to place a margin around the tumor and consider the resulting volume as the target [17, 19, 18]. Under this approach, we begin with a clinical target volume (CTV) determined by adding a margin to the gross tumor volume (GTV) to take into consideration "potential 'subclinical' invasion" [18]. The CTV is an oncological concept that specifies the pure target. From the CTV, a planning target volume (PTV) is established. To obtain the PTV, an internal margin (IM) and a set-up margin (SM) are combined in some fashion with the CTV. The IM is defined to take into consideration physiologic variations; the SM is defined to take into consideration uncertainties in technical factors such as patient set-up and mechanical stability [18]. As noted in [19, 18], simply adding the CTV, IM and SM to obtain the PTV may result in an excessively large PTV that could result in normal tissue complications. Thus, typically a global safety margin is defined that depends upon the situation at hand: when sensitive structures (organs at risk) are nearby, a smaller margin is used. The definition of the safety margin usually involves a compromise. For example for organs at risk, a planning organ at risk volume (PRV), analogous to the PTV for the target, is defined [19, 18]. When the PRV and the PTV overlap, then the safety margin becomes a compromise between the two volumes, as determined by the radiation-oncology team [18]. The advantage to this approach is that small displacements can occur and the tumor will still be dosed. However, normal tissue and/or organs at risk will also be dosed, due to the use of the margins.

Besides considering PTV margins, other clinical methods have been investigated [66]. Kubo and Hill [27] look at synchronizing the radiotherapy beam with respiration to minimize patient movement during treatment. Wong et al. [65] study active breathing control,

involving immobilization of the patient to minimize movement during treatment. Keall et al. [26] consider motion adaptive x-ray therapy, which adapts the beam to follow target motion. All of these clinical techniques focus on restricting internal movement due to respiration, but do not consider set-up issues like patient registration errors.

Statistical approaches have also been presented to deal with uncertainty in radiation treatment. Löf, Lind and Brahme [39] address the issue of modifying radiation beams to minimize the difference between the desired dose and the delivered dose, under the assumption that uncertainty in the process is governed by a known stochastic process. They build on the work of Lind et al. [37] (who earlier had solved the problem for special cases) by using symmetry and numerical integration techniques for small problems, and a Monte Carlo integration method for general cases. Similarly, Li and Xing [35] consider reducing the "hard margins" of PTV by representing random organ motion in terms of a spatial probability distribution, specifically a three-dimensional Gaussian. For fractionated stereotactic radiotherapy (radiotherapy with relocatable fixed beam heads), Zavgorodni [70] modifies the margin approach by convolving the dose with the probability density distribution of the the isocenter (where the beams meet).

Other statistical approaches use feedback from the system to adjust beam profiles. Löf, Lind and Brahme [40] incorporate dynamic optimization techniques into their work of [39]. Internal variations (such as organ movement) are dealt with using stochastic optimization, as before. Dynamic optimization is added to automatically adjust beam profiles and patient location in subsequent treatments to account for current set-up errors. Re-optimization has also been explored. In [69], Yan et al. discuss the conceptual idea of re-optimizing by adapting the margins treatment. Wu et al. [66] generalize this approach by modifying the original treatment plan as well. There is a growing literature on adaptive radiotherapy, see for example [20, 42, 64, 68], where the treatment plans are modified based on imaging information obtained prior to, or even during each treatment fraction [42, 50, 52, 53] to reconstruct the original planned dose.

The procedure we present requires feedback from the system. At this time, mechanisms to determine the actual dose delivered during individual treatments are quite primitive. More advanced imaging devices are currently being developed that can generate more accurate delivery information, highlighting where the delivered treatment may be inaccurate. For example, a helical tomotherapy system is able to produce megavoltage CT images of the patient [51] and dose reconstruction [21, 25, 22, 24, 23, 43] is being developed to verify the dose delivered to the patient on a daily basis [47, 48, 67]. Thus, in principle any variation between the delivered and the predicted dose distributions can be compensated for in subsequent treatments. The purpose of this paper is to exploit this knowledge to improve the overall treatment.

The paper aims to adapt the treatment plan on a day-to-day basis to account for organ movement and deformation, shrinkage of the tumor, or errors in the original prescription or treatment delivery. Rather than modifying a predetermined treatment plan, we build the plan as treatments progress, focusing on the total dose to be delivered. In our work we attempt to take uncertainty into account at the planning stage, and develop a technique that allows fractionation to decrease errors, rather than increase them. We develop a control mechanism for the treatment course, leaving the implementation of the daily dosage to a specialized planning tool. To find the control, we use neuro-dynamic

programming, particularly a rollout policy, to improve upon simple heuristic policies (see Section 2.3). Our focus is on determining appropriate doses to use that hedge against errors in the delivery process; we leave the particular implementation to specific tools used by application experts for particular machines.

In the next section, we consider the mathematical framework in which we will be working and describe techniques for solving the day-to-day planning problem. These techniques include neuro-dynamic programming (NDP) and heuristic policies, one of which is the current method of choice. Our application of NDP makes critical use of simulation to estimate the "cost-to-go" from a particular state with a certain number of remaining deliveries allowed. We next present examples and discuss their results, showing how the NDP ideas can improve upon the heuristic policies. We define rules of thumb, which allow for practical implementations of solutions suggested by NDP while still maintaining most of the improvements. Finally, we show how our methods perform on actual patient data under re-planning, assuming both accurate and implementable delivery. The paper concludes with our recommendations for planning methods and some suggestions of open research issues.

2. Model formulations

To describe the problem more precisely, we introduce some notation and a simplified model that captures the salient features of interest. Let $\mathcal{I}$ be a collection of voxels (pixels, points) and let $T(i)$, $i \in \mathcal{I}$ represent the required final dosage (target). Suppose the course lasts N periods (stages), and the actual dose delivered (the state) after k days is x_k. This state evolves as a discrete-time dynamic system:

$$x_{k+1} = f(x_k, u_k, w_k), \quad k = 0, 1, \ldots, N-1.$$

Here u_k is the control (the dose that will be delivered in the kth period) to be selected from a collection $U(x_k)$, and w_k is a random disturbance drawn from a set W. In the application, we assume that these random disturbances come from errors in the delivery process (such as patient movement) or errors in the setup (such as patient registration errors). For this reason, we assume that w_k corresponds to a shift to u_k. Further, since each treatment is delivered separately, the errors that arise pertain only to a particular treatment and time stage, and so w_k is independent over stages. A key issue to note is that the controls are nonnegative since dose cannot be removed from the patient.

At the end of N stages, the state x_N should minimize a terminal cost G. For ease of exposition we assume that $G(x)$ is a linear combination of the differences between the current dose and the target at each voxel, that is

$$G(x) = \sum_{i \in \mathcal{I}} c(i) \, |x_N(i) - T(i)| \, .$$

Here, the vector c weights the importance of hitting the target value for each voxel. We typically use similar values of c for distinct areas in the target, such as the location(s) of the tumor, sensitive structures like organs, and normal tissue. In practice, larger values

of c correspond to tumor areas and/or sensitive structures. This gives us the following mathematical model:

$$\begin{array}{ll} \min_u & E(G(x_N)) \\ \text{subject to} & x_{k+1}(i) = x_k(i) + u_k(i + w_k), \quad \forall i \in \mathcal{I}, k = 0, 1, \ldots, N-1 \\ & u_k \in U(x_k), u_k \geq 0, w_k \in W, \end{array} \tag{1}$$

with x_0 given and $E(\cdot)$ representing expectation.

2.1. Dynamic programming

Dynamic programming can be used to solve (1). Since the dose delivered on a given day impacts how much dose is still needed in subsequent days, the choice of control at each time stage contributes to the final cost. Dynamic programming applies backwards recursion. Starting at the last stage, the optimal control to apply for each state x is determined, then the second-to-last stage is considered, and so on, working backward through the stages.

As a means of determining the optimal control for a state, we consider the cost-to-go from the state x after k stages have elapsed, $J_k(x)$. As noted in [1, 2], the cost-to-go functions satisfy the dynamic programming recursion

$$J_k(x) = \min_{u \in U(x)} E[g_k(x, u, w) + J_{k+1}(f(x, u, w))] \tag{2}$$

with initial condition

$$J_N(x) = G(x), \tag{3}$$

where $g_k(x, u, w)$ represents the immediate cost of applying control u at stage k.

The individual controls $\bar{u}_k$ can be found in the backward recursion:

$$\bar{u}_k(x) = \arg \min_{u \in U(x)} E[g_k(x, u, w) + J_{k+1}(f(x, u, w))]. \tag{4}$$

If we wanted to use equations (2), (3), and (4) in the radiation treatment application, we would need to calculate $J_k(x)$ and $\bar{u}_k(x)$ for every possible state x at each stage k. A standard technique for doing this is to discretize the state space for x and form a lookup table for J_k and u_k over this discretization. Even for the simple examples that we describe in Section 3, this becomes unmanageable.

2.2. Heuristic policies

One approach to avoid the computational burden outlined above is to apply simple heuristic policies.

Several heuristic control techniques immediately spring to mind. First, there is the simple plan to deliver

$$u_k := T/N$$

at each stage and not account at all for disturbances in the delivery. This plan can be used when treatment errors cannot be measured directly, and is currently the method of choice. We refer to this plan as the constant policy. Note that the implementation on a particular machine of this policy only needs one optimization to be performed at the start of the process. However, even when a voxel has been overdosed at stage k, the constant policy continues to add dose at subsequent time stages. An alternative is to only add dose if the current dose is less than the target dose. We refer to this modification as the constant-plus policy; this latter policy would require re-planning at every stage.

If the distribution of the random disturbances is known, then we can construct a policy u^e whose expected delivery is the target distribution

$$E(u^e) = T.$$

In practice, u^e can be calculated for discrete probability distributions by solving a linear system, and will probably deliver more than the constant policy near the boundary of the target. It is then possible to deliver

$$u_k := u^e/N$$

at each stage. Additionally, we can take a convex combination of this policy with the constant policy; we call such a policy a modified constant policy.

An alternative to these (constant) policies is to attempt to compensate for the error delivered in the previous time by spreading the error over the remaining time stages. At each time stage, we divide the residual over the remaining time stages:

$$u_k := \max(0, T - x_k)/(N - k).$$

We refer to this plan as the reactive policy. Since the reactive policy takes into consideration the residual at each time stage, we expect that the reactive policy will perform better than a constant policy. Note, though, that the reactive policy requires knowledge of x_k and re-planning at every stage k.

In actual treatment plans, individual doses are subject to an upper bound, applied in order to limit burning and allow for healthy tissue to recover between treatments. For this reason, we assign a cutoff value that restricts the dose prescribed by each control to such an upper bound. For testing purposes, we set the cutoff to be $2T_{\max}/N$, which is double the dose prescribed by the constant policy in the worst case. Such a value allows for a large dose to be prescribed, while still ensuring that the dose is not unreasonably large. Although this is used in all our computations, very little changes if this upper bound is not applied.

2.3. Neuro-dynamic programming (NDP)

To overcome the "curse of dimensionality", various approximation techniques have been proposed for the solution of problems similar to (1). A particularly simple one is a rollout policy, an instantiation of neuro-dynamic programming [1, 2]. This approach can also be thought of as a heuristic improvement mechanism.

The key idea is to iterate (4) forwards over stages, finding the control to apply immediately by optimization and replacing $J_{k+1}(f(x, u, w))$ by an approximation. This is much closer to the way a decision maker would work in practice: only today's decision is needed precisely; the remaining decisions can be approximated. One approximation assumes the existence of a good base (heuristic) policy that will be used at all future stages. Then for $k = 0, 1, \ldots,$ we choose the control $\bar{u}_k$ to use at stage k by solving

$$\bar{u}_k(x_k) = \arg \min_{u \in U(x_k)} E[g_k(x_k, u, w) + H_{k+1}(f(x_k, u, w))] \tag{5}$$

where H_{k+1} is the approximation of J_{k+1} obtained by simulating the future and applying the base policy at all decision points in the future. Under appropriate assumptions, it has been shown [2] that the policy

$$(\bar{u}_0(x_0), \bar{u}_1(x_1), \ldots, \bar{u}_{N-1}(x_{N-1}))$$

outperforms the given base policy. While this approach is probably the simplest form of neuro-dynamic programming, it has proved to be very effective in a variety of practical examples.

The minimization in (5) is carried out by complete enumeration in the setting where $U(x)$ is a finite set. Defining the Q-factor

$$Q_k(x_k, u) := E[g_k(x_k, u, w) + H_{k+1}(f(x_k, u, w))]$$

the minimization could be carried out by comparisons between $Q_k(x_k, u)$ and $Q_k(x_k, \tilde{u})$ for $u, \tilde{u} \in U(x_k)$. Rather than calculating H_{k+1} at all possible states $f(x_k, u, w)$ using multiple simulations, we instead use a single simulation to estimate the Q-factor Q_k as $\tilde{Q}_k$. Since simulation is involved, difference calculations between $\tilde{Q}_k(x_k, u)$ and $\tilde{Q}_k(x_k, \tilde{u})$ will be prone to errors. These errors can be alleviated somewhat by simulating the Q-factor differences $Q_k(x_k, u) - Q_k(x_k, \tilde{u})$ instead of taking the difference between two simulated Q-factors [1]. Essentially, we simulate differences that are calculated from the same realizations of w. The simulation code we use generates 10000 paths through the simulation tree between stage k and stage N. While in our examples this enabled us to ensure we chose the correct u, more sophisticated techniques based on the variances of the differences, or ranking would be better. For each path, the Q-factor differences are calculated; at the end, the average of these differences determines $\bar{u}_k$.

Thus our NDP rollout policy is generated as follows. We begin by choosing the base policy for the calculation of H_{k+1}. As the base policy is applied at all later stages, it should be a heuristic policy that performs well. In our numerical experiments, we use the reactive policy, described above in Section 2.2. Starting at a given x_0, for each pair of controls u and $\tilde{u} \in U(x_0)$, we simulate (concurrently) the Q-factor differences $Q_0(x_0, u) - Q_0(x_0, \tilde{u})$, and choose the $\bar{u}(x_0)$ that makes all these differences negative. (In our example these differences actually only involve the terminal state x_N.) We then calculate x_1 using the state dynamics, and repeat the entire procedure at the next stage.

In the radiation treatment application that is expressed in (1), we assume no immediate costs in applying individual controls and so $g_k(x, u, w) = 0$ (if the costs for missing the target area are additive, g could account for these explicitly when they occur). In practice, the procedure could also implement an on-line policy choice since

the simulation model could be modified as time stages elapse. In the current setting, we approximate the future by simulation and choose the policy to apply right now by optimization. After applying this policy, we wait for time to elapse and repeat the same process at the next stage. For a particular radiation therapy, the choice of the current control may itself involve a lengthy optimization similar to that outlined in Section 4.2, and the error produced in delivery will be provided to the decision-maker as a by-product during delivery. (For example, observed anatomical changes may also be incorporated into later simulations.)

3. Model computations

In this section we analyze the behavior of small models. Section 3.1 suggests alternative approaches when re-planning is not allowed at every stage. A variety of approaches that deliver the same dosage at each stage are compared under disturbances during delivery. Section 3.2 extends the analysis to investigate approaches that allow re-planning at each stage, and shows how these approaches improve upon the other techniques. Due to the added computational burden of determining such methods, we investigate "rules of thumb" in Section 3.3 that allow plans to be determined without forward simulations. We believe that the distribution of the errors may affect the form of our solutions. We use the term volatility loosely to measure the variance of the distribution. The rules of thumb are promising for low volatility cases, but are less applicable under high volatility.

3.1. Modified constant policy

We investigate the use of the modified constant policy as described in Section 2.2. We consider a case where $\mathcal{I} = \{1, \ldots, 20\}$, with $T(i) = 1$ and $c(i) = 5$ when $3 \leq i \leq 18$, $T(i) = 0$ and $c(i) = 1$ otherwise (see Section 2 for details on this notation).

The probability of shifts for all our models are given by:

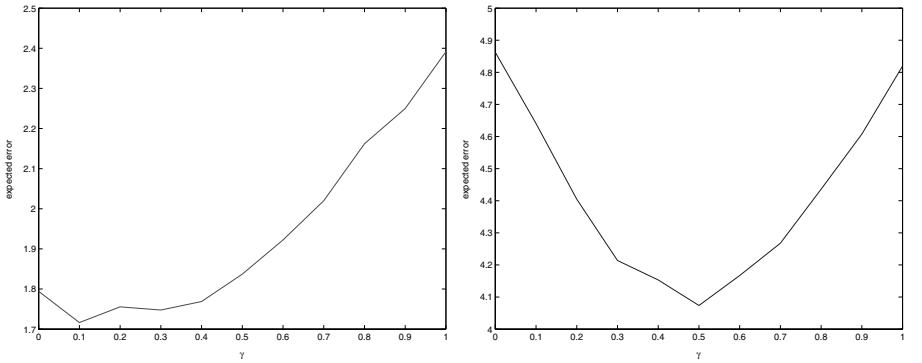
$$w_k(\delta, \mu) = \begin{cases} -2 & \text{with probability } \delta \\ -1 & \text{with probability } \mu \\ 0 & \text{with probability } 1 - 2(\delta + \mu) \\ 1 & \text{with probability } \mu \\ 2 & \text{with probability } \delta. \end{cases} \tag{6}$$

for different choices of δ and μ.

In this section, low volatility refers to the shift probabilities given by $\delta = 0.05$ and $\mu = 0.1$, while high volatility corresponds to $\delta = 0.1$ and $\mu = 0.2$. In this simple example, we can calculate u^e from Section 2.2 easily, and can evaluate the terminal cost as an average over 1000 simulations. The modified constant policy depends on the volatilities: we show in Figure 2 the results of runs at these two volatilities for a convex combination of u^e with the constant policy u^c parameterized by:

$$u = \gamma u^c + (1 - \gamma) u^e. \tag{7}$$

Note that in the low volatility case u^e significantly outperforms u^c but in the high volatility case this is no longer true. Clearly, our approaches must be able to deal with



(a) Low volatility (b) High volatility

Fig. 2. Expected error for modified constant policy, parameter γ and $N = 20$

different levels of volatility in delivery. Dependent on the volatility we see improvement for particular choices of γ over both u^e and u^c. We also note that separate calculations demonstrated that the value of γ decreases as N increases as we would expect; for the low volatility case we found that $\gamma = 1$ for $N = 4$ and 6, $\gamma = 0.5$ for $N = 10$ and $\gamma = 0.35$ for $N = 14$.

For completeness, Figure 3 shows the terminal cost associated with all the policies we investigated that require only a single planning step as a function of N. We recommend using the modified constant policy in this situation provided there are good estimates of the probability distribution of disturbances.

Most of the error for the constant policy is due to underdosing of the target area, not overdosing of the sensitive structures. By choosing a multiple of T/N (greater than 1) to deliver at each stage (instead of T/N itself), we can reduce this error at the cost of increasing the total dose delivered to the patient. However, on targets with extensive interiors, including the one analyzed above, this leads to significant overdose in the these regions; we do not discuss such approaches further.

3.2. Simple NDP rollout policy examples

To gain intuition regarding the potential of the re-planning approaches, we first consider two simple, one-dimensional targets under different weighting and probability distributions, pictured in Figure 4. For both targets, $\mathcal{I} = \{1, 2, \ldots, 9\}$ and we allow a maximum shift of 2 voxels. In both targets, the "spikes" of dose 0.8 represent tumor locations. Thus, it is important that these areas receive as much of the 0.8-prescribed dose as possible, and so these areas will have a relatively high weighting in the objective. The 0.1 areas can represent sensitive structures (which can be exposed to a certain level of radiation) or normal tissue, depending upon the particular weighting scheme employed. We apply 3 different weighting schemes to the spike target in Figure 4(a). Moving from easiest to hardest, these schemes are:

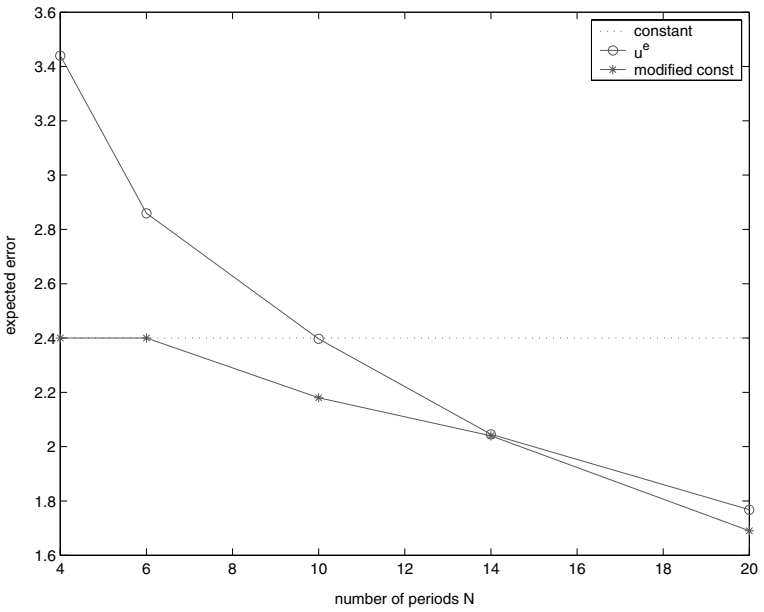
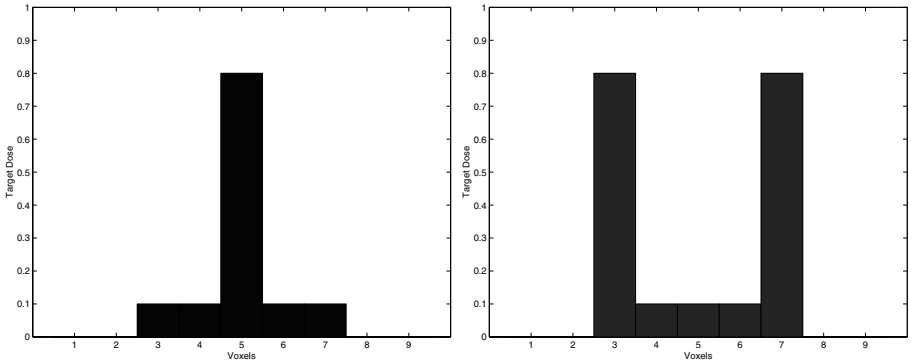


Fig. 3. Single plan policy results



(a) Spike target

(b) Double spike target

Fig. 4. Example targets

- the smooth weighting: $c = [1, 1, 1, 1, 10, 1, 1, 1, 1]$, which only enforces the 0.8-dosage, allowing for more variation in the other voxels (including a "building" up to the spike);
- the nonsymmetric weighting: $c = [1, 1, 1, 1, 10, 5, 5, 1, 1]$, which allows for a build-up to the spike on the left-hand-side, but enforces the spike structure rigidly on the right-hand-side; and

– the spike weighting: $c = [1, 1, 5, 5, 10, 5, 5, 1, 1]$, which enforces the spike structure rather rigidly.

For the double spike target of Figure 4(b), we apply the double spike weighting scheme: $c = [1, 1, 10, 5, 5, 5, 10, 1, 1]$, which enforces high dosage on the target edges and the low dosage in the center. The examples have been chosen to simulate practical cases of interest in the application area. We consider first the low volatility case where $\delta = 0.02$ and $\mu = 0.08$ in (6) for every stage k.

We apply the NDP rollout approach using the reactive policy as the base policy. We require a rich collection of heuristics for the finite set $U(x_k)$. We include in this collection the constant and reactive policies, and we add what we refer to as categorical policies. For these policies at stage k, we calculate the residual target for each voxel i by $\max\{0, T(i) - x_k(i)\}$. Then, the voxels are divided into three categories by comparing their residual target to the maximum residual:

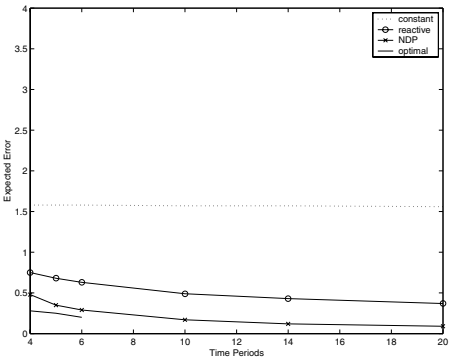
$$\max_{i \in \mathcal{I}} \max\{0, T(i) - x_k(i)\}.$$

The three categories correspond to voxels whose residual target is less than 40% (low residual), between 40% and 70% (medium residual), and greater than 70% (high residual) of this maximum value. In each category, we apply one of three controls. Either we apply 0 dosage, 0.4 of the residual target, or $1/(N-k)$ of the residual target. This yields an additional 26 policies for $U(x_k)$ (as the reactive policy is the categorical policy with $1/(N-k)$ applied in each category).

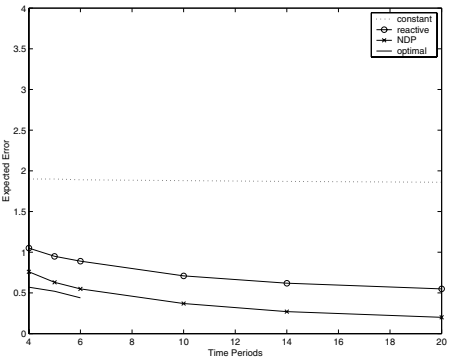
Figure 5 displays simulated results of each example. For each graph, the constant policy, reactive policy, and NDP rollout policy results are displayed, as well as the optimal results for time stages 4, 5, and 6. The optimal results come from reformulating the problem as a stochastic linear programming and solving it exactly. The curse of dimensionality precluded exact solution with more time stages with this approach, although it may be possible to extend to a few more stages using scenario reduction or sampling techniques [15, 38].

Note that the alphabetic ordering of targets (a) to (d) are increasingly difficult and lead to larger errors, independent of the optimization scheme chosen. Common to all examples is the poor performance of the constant policy. The reactive policy performs better than the constant policy, but not as well as the NDP rollout policy. The level of improvement depends upon the difficulty of the target. We see that the reactive policy gives a large improvement over the constant policy — the error is nearly halved. NDP does even better, yielding about a 50% drop in the reactive policy error at larger time stages, and achieving near-optimal results at smaller time stages. As time advances, the improvement for both the NDP and reactive policies becomes greater: where constant remains almost level, reactive and NDP continue to drop as we move to later time stages. Further, NDP decreases faster than the optimal results do, suggesting that it may become optimal at later time stages.

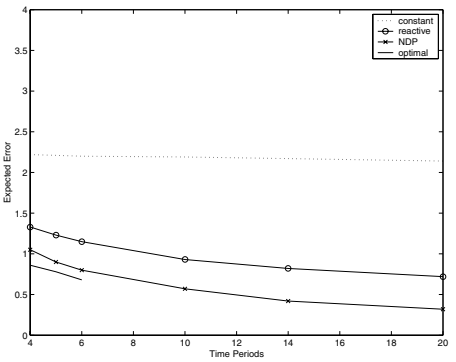
In addition to three-category policy choices, we also experimented with two-category policy choices. Under these policies, the voxels were classified as either less than 50% of the maximum residual or more than 50% of the maximum residual. To maintain approximately the same number of policies, we allowed five choices for each category (resulting in a total of 27 policies, including the constant policies). In one experiment,



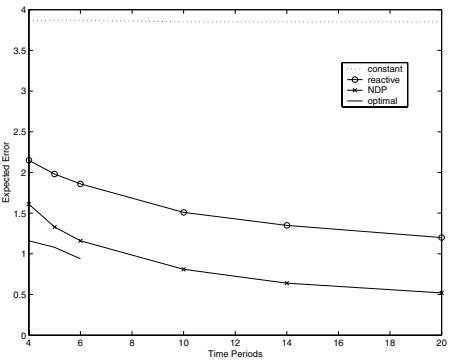
(a) Spike target with smooth weighting.



(b) Spike target with nonsymmetric weighting.



(c) Spike target with spike weighting.



(d) Double spike target/double spike weighting.

Fig. 5. Simple examples under low volatility

we allowed for small multiples: 0, $1/(N-k)$, 0.01, 0.1, or 0.4. In another experiment, we allowed for large multiples: 0, $1/(N-k)$, 0.1, 0.4, or 0.6. Applying these policies to the double spike example (the hardest example), we found very little change in the NDP results. The small-multiple category choices returned approximately the same results as the three-category choices, while the large-multiple category choices returned slightly better results but nothing visually significant on the plot.

These results suggest that significant improvements over the presented NDP results cannot be achieved while choosing from among approximately 30 policies. Note that in addition, we also experimented with many more examples, including different targets, different weighting schemes, and larger targets. The results from these other examples were qualitatively the same. We did find, though, that for higher volatility examples, constant weighting ($c(i) = 1, \forall i \in \mathcal{I}$) resulted in significant hedging and consequential underdose of the target. This strongly suggests that the use of an appropriate weighting

Table 1. Expected error after 20 stages for Section 3.1 example using 1000 simulations

Policy	Mean	Variance
Constant	2.40	0.81
Modified Constant (0.2)	1.75	0.43
Reactive	0.74	0.10
Modified Reactive (3.0)	0.56	0.16
NDP rollout	0.53	0.17

scheme to focus the treatment is imperative. Further computational testing is reported in [60].

We experimented with a different collection of policy choices in the NDP approach, namely replacing the categorical choices with a collection of choices of multiples of the reactive delivery at each stage. Thus, we allow delivery of

$$\alpha_k \max(0, T - x_k)/(N - k) \tag{8}$$

at each stage k, with α_k taken from a finite collection. Repeating the low volatility experiments of Section 3.1 gave the results shown in Table 1. (Note that the value 0.2 corresponds to the value chosen for γ in (7) and the value 3.0 corresponds to the value of α_k in (8).) The modified reactive policy uses a fixed $\alpha_k = 3$ at each stage, whereas the NDP approach updates the value in the manner described above. Note that all the re-planning techniques outperform the single planned methods, both in mean and variance. Furthermore, the more aggressive modified reactive policy (that chooses $\alpha_k = 3$ at every stage except the last two) performs very well. The NDP rollout policy, choosing the multiples using forward simulations, outperforms all the methods at this volatility. For higher volatilities, the NDP approach remains the best, but the aggressive modified policy becomes inferior to the reactive policy due to a significant increase in variance of its solution errors. This is not the case for the NDP approach. Thus, the forward simulations facilitate adapting to uncertainty; it remains to be shown whether the additional computational cost is allowable in the application.

3.3. Rules of thumb

While building simple models and analyzing their properties can lead to great insight into the application at hand, it is important to draw definitive conclusions that are applicable to the real problem. In this section, we endeavor to derive policies u_k that are not derived from a forward simulation from the state x_k. Removing this requirement, allows such policies to be directly implemented in the radiation treatment planning arena by simply modifying the delivery target at each stage k.

For the results of Section 3.2, we applied an outer simulation to generate many paths through the scenario tree and we used an inner simulation (for Q-factor differences) to determine $\bar{u}_k$. To find policies without recourse to simulation, we look for policies that are used most often at stage k. For a particular example, the outer simulation gives a series of possible policies to apply. By considering the average Q-factors for each control, we have an idea of how effective that control is for that example at that stage.

Table 2. Rules of thumb for 20 time period examples

Stage	Low Volatility	Med. Volatility	High Volatility	Simple Rule
1–9	(0, 0, 0.4)	(0, 0, 0.4)	(0, 0, 0.4)	(0, 0, 0.4)
10	(0, 0.09, 0.4)	(0, 0.09, 0.4)	(0, 0.09, 0.4)	(0, 0.09, 0.4)
11	(0, 0.1, 0.4)	(0, 0.1, 0.4)	(0, 0.1, 0.4)	(0, 0.1, 0.4)
12	(0, 0.11, 0.4)	(0, 0.11, 0.4)	(0, 0.4, 0.4)	(0, 0.11, 0.4)
13	(0, 0.4, 0.4)	(0, 0.125, 0.4)	(0, 0.4, 0.4)	(0, 0.4, 0.4)
14–17	(0, 0.4, 0.4)	(0, 0.4, 0.4)	(0, 0.4, 0.4)	(0, 0.4, 0.4)
18	constant-plus	constant-plus	(0.4, 0.4, 0.4)	(0.4, 0.4, 0.4)
19	constant-plus	constant-plus	(0.5, 0.5, 0.5)	(0.5, 0.5, 0.5)
20	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)

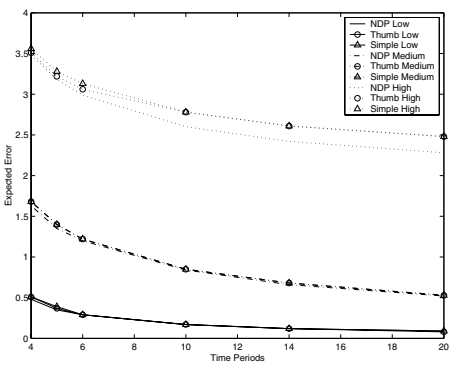
In [60] we tried three different volatility levels, namely low ($\delta = 0.02$, $\mu = 0.08$ in (6) as reported above), medium ($\delta = 0.05$, $\mu = 0.15$) and high ($\delta = 0.05$, $\mu = 0.25$). Averaging these Q-factors across examples with the same volatility and choosing the controls that correspond to the smallest Q-factors at each time stage, we determine a generalized policy for each volatility. We refer to this generalized policy as the "rule of thumb" policy for that volatility. These rules of thumb allow us to remove the dependence on the simulation and provide us with a pre-defined plan to use for a particular volatility.

We can take the generalization further and remove dependence on the volatility by averaging the Q-factors across volatilities as well. We refer to the resulting policy as the "simple rule of thumb". The rules of thumb and simple rules of thumb for $N = 20$ are given in Table 2. The categorical policies (including the reactive policy) are indicated as triplets; the first entry corresponds to the low residual areas; the second entry corresponds to medium residual areas; and the third entry corresponds to the high residual areas. These entries correspond to the multiplier of the residual that is used at all voxels in that area.

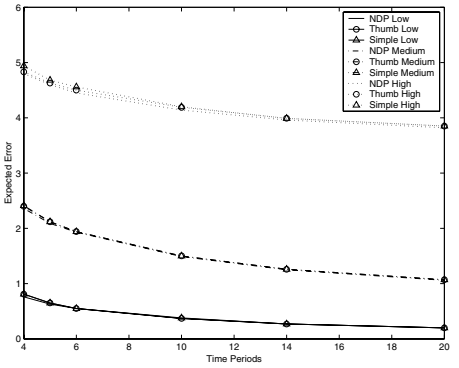
Examining the tables, we notice some general trends in the control choices. First of all, within each table, the controls become more aggressive as we near the final time stages, generally moving from controls in which only the high residual areas are dosed, to controls in which all areas are dosed. Typically, we use the first half of the time periods to work aggressively on the high residual areas and ignore the other areas.

Controls in the middle stages tend to focus on both the medium and high residual areas first. Later stages focus on all three categories, ending in every case aggressively with the reactive policy (to attempt to apply all of the remaining dose). An interesting question arises as to whether the low and medium volatility rules follow this general trend. In these three cases, the rule of thumb makes use of the constant-plus policy in the later time stages. While $1/20$-th of the original target dose is a seemingly rather small amount, we claim that it is very likely to be an aggressive control at the later time stages since the remaining residual is likely to be small, and hence a small fraction of the original dose is in fact a large dose in comparison to the residual. In this case, the removal of overdosing (the difference between constant and constant-plus) is important.

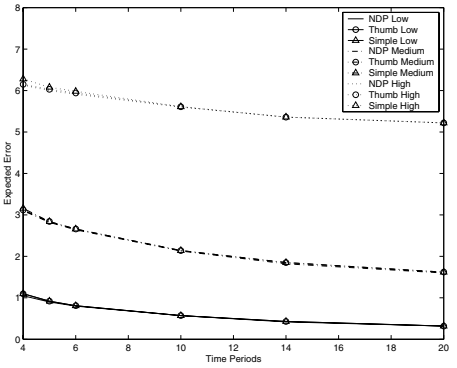
Since they are generalizations, we expect that the rules of thumb and the simple rule of thumb will perform worse than the NDP rollout policy. This is the case, although the differences tend to be so small that they are not noticeable. Figure 6 compares the



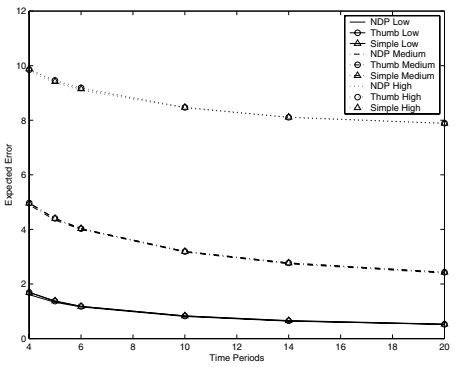
(a) Spike target with smooth weighting.



(b) Spike target with nonsymmetric weighting.



(c) Spike target with spike weighting.



(d) Double spike target with double spike weighting.

Fig. 6. Rules of thumb and simple rules of thumb results for the examples

NDP rollout results to the rules of thumb and simple rules of thumb for each example. In addition Table 3 shows the percentage decrease achieved by the reactive policy, the NDP rollout policy, rules of thumb and simple rules of thumb over the currently-used constant policy at 20 time stages.

The results applying the rules of thumb and simple rules of thumb are almost always indistinguishable from one another. We only see a noticeable difference in the high volatility example of Figure 6(a). It is not surprising that one target suffers under a generalization built from considering all targets. This case, corresponding to the easiest target, probably does not require the same controls as the other targets, and so suffers particularly in the high volatility case where things are more likely to go wrong. Table 3 provides the end point for the results shown in Figure 6. In all cases, the NDP, rules of thumb and simple rules of thumb improved upon the reactive policy results significantly.

Table 3. Percentage decrease over the constant policy at 20 time stages, calculated for the reactive policy, the NDP rollout policy (NDP), the rules of thumb (RoT) and the simple rules of thumb (SRoT)

Target	Volatility	Reactive	NDP	RoT	SRoT
Smooth Spike	Low	76%	94%	94%	94%
Smooth Spike	Medium	51%	83%	81%	83%
Smooth Spike	High	24%	51%	47%	47%
Nonsymmetric Spike	Low	70%	89%	89%	89%
Nonsymmetric Spike	Medium	44%	71%	69%	71%
Nonsymmetric Spike	High	15%	30%	29%	29%
Spike Spike	Low	66%	85%	85%	85%
Spike Spike	Medium	38%	61%	60%	61%
Spike Spike	High	8%	17%	16%	16%
Double Spike	Low	68%	86%	86%	86%
Double Spike	Medium	43%	69%	67%	68%
Double Spike	High	17%	31%	31%	31%

However, the percentage decrease over the constant policy varies very little between the three of them. This suggests that very little is sacrificed in moving to the generalized simple rules of thumb; this is the choice used in Section 4.1.

Note that if the policy pool U is changed, the simulations must be rerun and this process must be repeated on the new results in order to determine appropriate rules of thumb and simple rules of thumb.

The results do have a disturbing trend, in that as the volatility increases the percentage improvement over the constant policy decreases. In order to investigate this further, we consider a much larger example, based on realistic data.

4. Real-life treatment planning example

In this section, we consider how well the NDP simple rule of thumb generalizes to a real-life example. The example that we use in this section is a three-dimensional example drawn from actual patient data.

As in the one-dimensional case, the area of consideration is divided into voxels. We consider an area where the x-axis varies from 0 to 110, the y-axis varies from 0 to 90, and the z-axis varies from 0 to 80. Within this area, there are 747, 667 voxels in normal tissue; 69, 270 voxels in four sensitive structures (spinal cord, liver, left and right kidneys); and 1, 244 voxels in the tumor. Figure 7 shows the layout at $z = 30$.

We consider a 20-stage treatment. Let $\mathcal{T}$ be the set of tumor voxels, $\mathcal{S}$ be the set of sensitive structure voxels, and $\mathcal{N}$ be the set of normal tissue voxels. Upon completion, we would ideally like to have the tumor dosed completely, with no dose delivered to the sensitive or normal tissues:

$$T^*(i, j, k) = \begin{cases} 1 \text{ if } (i, j, k) \in \mathcal{T} \\ 0 \text{ if } (i, j, k) \in \mathcal{S} \cup \mathcal{N}. \end{cases} \tag{9}$$

However, this is not possible as radiation beams must pass through normal and/or sensitive tissues to reach the tumor, see [36] for a description. Instead, application experts allow for some error by relaxing their requirements to the following [36]:

$$T(\mathcal{T}) \in [0.95, 1.07], \tag{10}$$

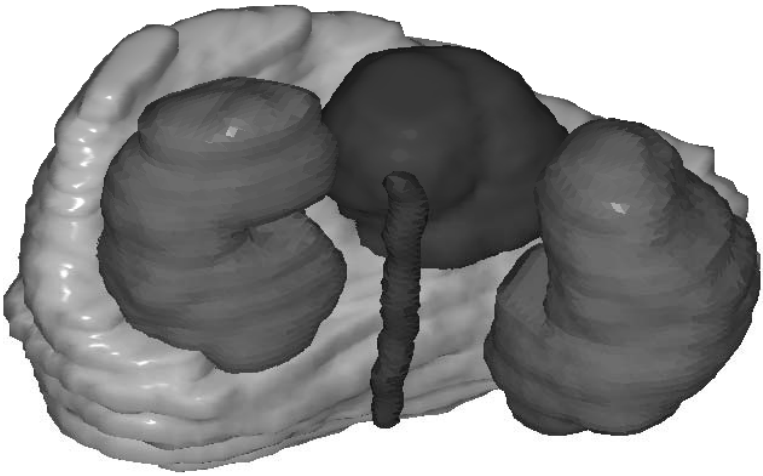


Fig. 7. Pancreas example: The large mass in the backround is the liver, the spinal cord is in the foreground, the kidneys are to the right and left of the darkly shaded target structure

$$90\% \text{ of sensitive tissues should have } T(\mathcal{S}) \leq 0.2 \cdot T^*(\mathcal{T}) = 0.2, \tag{11}$$

and

$$T(\bar{\mathcal{N}}) \text{ should be as small as possible} \tag{12}$$

where $\bar{\mathcal{N}}$ is a reduced set of normal tissue voxels, consisting of those normal tissue voxels around the tumor and a sampling of the other normal tissue voxels. In our case, $\bar{\mathcal{N}}$ consists of 96, 154 voxels.

4.1. Re-planning with perfect delivery

We consider a direct translation from the simple example. As before, we allow for a shift in the delivery of one or two voxels. In the three-dimensional example, we allow for six possible shift directions (up, down, left, right, forward, back); this gives thirteen total possibilities (including no shift). Here again, we consider three volatilities, based on the volatilities from the one-dimensional case: low volatility (probability of a shift is 0.2), medium volatility (probability of a shift is 0.4), and high volatility (probability of a shift is 0.6). In addition, we consider a very volatile case, where the probability of a shift is 0.78. We also assume that the ideal dose can be delivered, and so we measure the error against T^* from equation (9).

Table 4 displays the average results for the constant, reactive and NDP simple rule of thumb policies after 1000 simulations. Displayed are the errors on the tumor, the errors on the sensitive structures, the errors on the normal tissues, and the overall error. The overall error is a linear combination of the three other errors, found by weighting the areas as in the one-dimensional examples: the weight on the tumor is 10, the weight on the sensitive structures is 5, and the weight on the normal tissues is 1.

Table 4. Results of simple shifts on the real-life example

Policy	Volatility	$\mathcal{T}$ Error	$\mathcal{S}$ Error	$\mathcal{N}$ Error	Overall Error
Constant	Low	54.3	0.5	53.8	599.3
Reactive	Low	5.1	0.6	58.4	112.4
Simple NDP	Low	0.3	0.6	60.1	66.1
Constant	Medium	107.7	1.0	106.7	1188.7
Reactive	Medium	18.7	1.1	127.6	320.1
Simple NDP	Medium	5.0	1.2	134.6	190.6
Constant	High	151.7	1.0	150.7	1672.7
Reactive	High	48.5	1.4	201.5	693.5
Simple NDP	High	34.3	1.4	221.1	571.1
Constant	Very High	263.1	4.7	258.4	2912.9
Reactive	Very High	145.6	6.7	403.4	1892.9
Simple NDP	Very High	166.5	7.4	481.9	2183.9

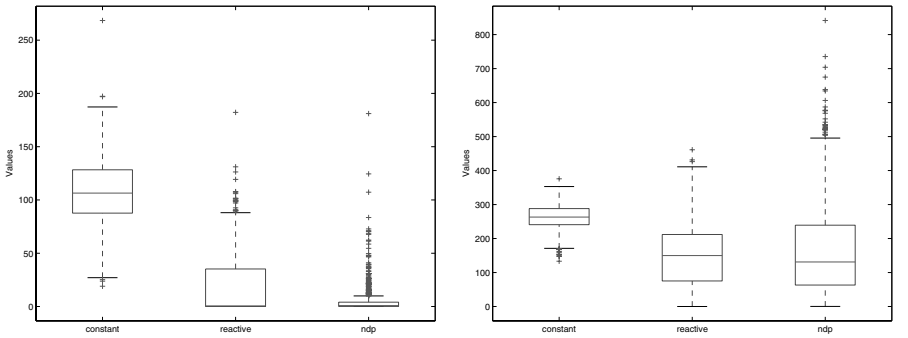
The table shows that the simple NDP rules of thumb significantly outperforms the constant policy on the tumor in all cases. It also does very well compared to the reactive policy, beating the reactive policy in all cases except for the very high volatility. This is not surprising since the simple NDP policy was built from considering only the low, medium and high volatilities, but not the very high volatility. The NDP policy is too aggressive under very high volatility; it produces significantly worse errors on the normal region.

Under the other volatilities, the NDP policy has about the same error as the reactive policy on the sensitive tissue, and slightly worse error on the normal tissue. The degraded performance on the normal tissue is not surprising, considering the simple one-dimensional examples that were used to build the NDP policy. We note that the simple NDP policy was built focusing on the tumor: in the one-dimensional examples, the tumor areas (spikes) were always weighted high. As a result, the NDP policy concentrates on fully dosing the tumor, whereas avoidance of normal and sensitive tissues is a secondary consideration.

Also, we note that the simple NDP policy delivers more overall dose to the patient than the constant policy. In addition, most of this dose is delivered early on, when the residual is high (since we deliver 0.4 of the high residual). Due to the shifts, this higher dose is generally not delivered to the tumor (although some of the tumor may be overdosed). Thus, we would expect that the error on the normal and sensitive tissues to be larger for the simple NDP policy.

Like the NDP policy, the reactive policy may deliver more dose overall than the constant policy. However, in the ideal case of no shifts, the reactive policy reduces exactly to the constant policy. Unlike the NDP policy, where a great deal of dose is delivered early on, the reactive policy policy starts out by delivering little dose early on ($1/20$-th of the residual — exactly the constant policy at the first time stage). Thus, at the last time stage, where we deliver the full residual, the reactive policy may be delivering a large amount of dose to make up for errors earlier on. As a result, a shift in the last policy can be disastrous for the reactive policy.

As noted earlier, unlike the NDP or the reactive policies, the constant policy delivers exactly the same dose at every time stage. Because it does not adjust its dose for errors



(a) Medium volatility

(b) Very high volatility

Fig. 8. Boxplots of target values

on the tumor, the constant policy tends to underdose the tumor. In fact, as can be seen from Table 4, the constant policy error on the tumor is the sum of the errors on the normal and sensitive tissues — the tumor dose was simply shifted to the normal and sensitive tissues instead.

The boxplots of Figure 8 show more detail on the distribution of the target values over the 1000 simulation runs carried out to generate Table 4. In each column, the box has lines at the lower quartile, median and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers and are indicated by crosses. It is clear that as the volatility increases, the effectiveness of a simple NDP policy based on simulations at lower volatilities on a simple model diminishes. Figure 8(a) shows that the simple model is indeed very useful as a predictive mechanism for the real data model when the volatility is not too great; both the mean and the variance are significantly reduced under the simple NDP policy.

4.2. Re-planning with implementable delivery

In the previous examples, we assumed that we could achieve any dose that was necessary for the policies. However, these doses may not be physically implementable. To obtain an implementable dose, we must use a planning tool. Our work thus far has been independent of any planning tools. In this section, we make use of a particular planning tool, presented in [36], to demonstrate how the policies perform under actual re-planning.

Given the ideal dose T^* from equation (9), the planning tool in [36] solves a mixed-integer programming problem to determine the angles (out of a total of 36) from which individual beams of radiation should be delivered and the length of time that each delivery should last. To reduce the solution time, we preselect ten angles to use. This reduces the problem to a linear programming problem:

$$\begin{aligned}
\min_{w,D} \quad & \lambda_t(\|(D_{\mathcal{T}} - \theta_L(\mathcal{T}))_+\|_\infty + \|(\theta_U(\mathcal{T}) - D_{\mathcal{T}})_+\|_\infty) \\
& + \tfrac{\lambda_s}{C_s}\|(D_{\mathcal{S}} - \phi)_+\|_1 + \tfrac{\lambda_n}{C_n}\|D_{\mathcal{N}}\|_1 \\
\text{subject to} \quad & D_\Omega = \sum_{A \in \mathcal{A}} \text{Dose}(A, \Omega) w_A, \ \Omega = \mathcal{T} \cup \mathcal{S} \cup \mathcal{N} \\
& D_{\mathcal{T}} \leq u \\
& 0 \leq w_A \leq M, \qquad \forall A \in \mathcal{A}
\end{aligned} \tag{13}$$

Here, $\mathcal{A}$ is the set of (ten) angles that can be used to deliver radiation. w_A represents the length of radiation exposure time for each angle $A \in \mathcal{A}$, bounded above by M (M is predetermined by application requirements). Dose(A, Ω) is a data matrix that contains the amount of radiation that is delivered to each voxel $(i, j, k) \in \Omega = \mathcal{T} \cup \mathcal{S} \cup \mathcal{N}$ when $w_A = 1$. D_Ω is the total radiation dose delivered to each voxel in Ω. Note that $D_{\mathcal{T}}$ is bounded above by u; we set

$$u = 1.15 \cdot \theta(i, j, k), \quad \forall (i, j, k) \in \mathcal{T},$$

where $\theta(i, j, k)$ is the desired dose at voxel (i, j, k). Here, we do not impose an upper bound on dose that can be delivered to each non-target voxel, but use the objective function to penalize large deviations from the desired dose. However, u could also be used to limit the total dose delivered during each stage: for example, we could set $u = 1/10$ (twice the ideal constant dose) to correspond to the upper bound from the simple one-dimensional examples. θ_L and θ_U are the lower and upper bounds, respectively, on the acceptable tumor dose. From the prescription (10),

$$\theta_L(i, j, k) = 0.95 \cdot \theta(i, j, k), \quad \forall (i, j, k) \in \mathcal{T} \tag{14}$$

and

$$\theta_U(i, j, k) = 1.07 \cdot \theta(i, j, k), \quad \forall (i, j, k) \in \mathcal{T}. \tag{15}$$

ϕ is the acceptable upper bound on the sensitive tissue. From the prescription (11),

$$\phi = 0.2 \cdot (\text{fraction of total dose to be delivered}).$$

(Note that, due to the difficulty in applying this restriction to only 90% of the sensitive tissue, as in the original prescription, we instead apply it to all of the sensitive tissue.) The scalars λ_t, λ_s, and λ_n are parameters that weigh the importance of the prescriptions for the tumor, sensitive structures, and normal tissues, respectively. We set $\lambda_t = \lambda_s = \lambda_n = 10$. C_s and C_n represent the cardinality of the sensitive and normal tissues, respectively, allowing us to compare the average error on the sensitive and normal tissues to the maximum error on the tumor.

To apply the constant policy, we need only solve model (13) once. If we solve model (13) with

$$\theta(i, j, k) = 1, \quad \forall (i, j, k) \in \mathcal{T}$$

and

$$\phi(i, j, k) = 0.2, \quad \forall (i, j, k) \in \mathcal{S},$$

Table 5. Results of re-planning on the real-life example under very high volatility

Policy	$\mathcal{T}$ Error	$\mathcal{S}$ Error	$\mathcal{N}$ Error	Overall Error
Constant	613.5	0.0	2743.9	8878.9
Reactive	41.2	242.6	6633.0	8258.0
Simple NDP	75.2	62.3	6454.5	7518.0

we find the implementable dose for all of T^*. We can then divide the resulting dose by 20 to obtain the implementable dose to deliver at each time stage. (Note that each time stage's dose is implementable: we set the weights for each stage to $w_A/20$.)

To apply the reactive policy, we must re-plan at each time stage. This requires 20 solutions to model (13), each with a possibly different θ and ϕ. For time stage k, we set

$$\theta(i, j, k) = \max\left\{0, \frac{T^*(i, j, k) - x_k(i, j, k)}{N - k}\right\} \quad \forall (i, j, k) \in \mathcal{T}$$

and

$$\phi(i, j, k) = 0.2/(N - k) \quad \forall (i, j, k) \in \mathcal{S}.$$

Like the reactive policy, the simple NDP policy requires 20 solutions to model (13). To obtain θ, we use the simple rule categorical multipliers from Table 2: for (i, j, k) in category c (= low, medium, or high residual),

$$\theta(i, j, k) = m_k(c)(T^*(i, j, k) - x_k(i, j, k))$$

where $m_k(c)$ is the k-th multiple for category c. ϕ is similar to that for reactive, except we conform to the prescription and do not restrict a percentage of the sensitive tissue in each structure:

$$\phi(i, j, k) = \begin{cases} 0.2 \cdot m_k(\text{low}) & \text{for the smallest } 0.90 \cdot m_k(\text{low}) \text{ of each structure} \\ \infty & \text{for the largest } 0.10 \cdot m_k(\text{low}) \text{ of each structure.} \end{cases}$$

We apply the realistic shifts from the previous section, under the very high probability (probability of a shift is 0.78). Table 5 shows the results for the constant, reactive and simple NDP policies after 20 simulations. The errors given correspond to errors on the prescriptions (10), (11) and (12).

As shown in Table 5, the reactive and simple NDP policies significantly outperform the constant policy on the tumor, cutting the error by more than one-sixth. However, the error on the sensitive and normal tissues is also significantly higher — around three times more on the normal tissue, and a great deal more on the sensitive structures. Note also that the reactive policy does better than the simple NDP policy on the tumor, but the reactive's overall error is worse. This comes from the fact that the simple NDP policy performs significantly better on the sensitive structures, cutting that error by about one-fourth.

The improvement of the simple NDP policy over the constant and reactive policies shown in Table 5 is not as great as when perfect delivery was used. This suggests that the simple NDP policy suffers more from re-planning error than the reactive policy, further suggesting that the reactive policy generates dose distributions that are easier to plan. Considering a more advanced delivery tool, such as IMRT, may improve the NDP

results since these methods can deliver dose distributions having steep gradients at the boundaries of structures, similar to the deliveries that are necessary in our modified target distributions. In addition, incorporating a "re-planning error" term in the simulation during the building of the NDP policies may yield more robust simple NDP policies. Both of these are issues for future study.

For immediate use, we suggest that either the reactive policy or the NDP rollout approach be employed, as long as the errors on the sensitive and normal tissues are acceptable. If these errors are too high, a different weighting scheme that puts more emphasis on avoiding the sensitive and/or normal tissues can be used. The process though, remains the same. On a day-to-day basis, the treatment planner, knowing x_k, can calculate the dose required at each voxel in the manner outlined in Section 2.2, accounting for the stochastic errors that have occurred. Once this dose distribution is known, existing planning tools can be used to implement this on particular machines, as we have demonstrated here.

5. Conclusion and future work

Day-to-day treatment planning is a complex procedure that can significantly benefit from knowledge of the errors that occur during the delivery process. While dynamic programming and stochastic optimization would undoubtedly lead to better plans, they are currently intractable for application problems of realistic size and complexity.

This paper proposes a solution based on neuro-dynamic programming, coupled with heuristic policies that are based on the particular application. In terms of increasing planning complexity, we suggest the following approaches are the most promising:

- The modified constant policy (Section 2.2): if no re-planning is allowed through the course of treatment, this policy performs well provided the user has a good estimate for the distribution of the errors involved in delivery.
- If re-planning is allowed, and the case has low volatility, the simple rules of thumb (derived from the NDP rollout policy applied to simple examples) perform very well and outperform all the single plan policies and the reactive policy under perfect delivery. The practical implementation of both the simple rule and the reactive policies are exactly the same. First of all, knowledge of x_k is required. Given this information we can calculate the actual dose that should be delivered at each voxel $i \in \mathcal{I}$ by determining which category the voxel resides in, and then multiplying the residual $T(i) - x_k(i)$ by the categorical multiplier. Knowing the dose at every voxel i is all the data that is required to specify a plan optimization that determines how to implement that particular dose on a specific machine. As mentioned in the introduction, we allow existing planning tools to perform this step, and we believe this is a key advantage of our approach.
- When the volatility increases, or delivery devices are incapable of delivering the categorical policies generated above, the reactive policy of Section 2.2 performs well if the user is unwilling to perform forward simulations. The modified reactive policy (Section 3.2) can outperform this method for certain volatilities but its variance increases rapidly making it a dangerous choice for high volatility.

– If re-planning is allowed, the NDP rollout policy performs the best on small examples and under perfect planning. It is able to compensate for errors effectively provided some distributional information is available. Improved estimation of the cost-to-go function leads to better solutions. The principal disadvantage of the NDP rollout policy is the time needed to calculate the estimate H_k for the cost-to-go using simulation.

We propose some simple ideas to improve the accuracy of the estimation and reduce the time overhead.

– The backwards recursion step of dynamic programming can be used to determine the policies that will be used in the last few steps optimally. While this is impractical to carry out for all stages N, it can be used to provide very good estimates of cost-to-go with few stages remaining. Simulation can be used to extrapolate these values back to earlier stages.
– Simulation generates an upper bound on the cost-to-go value. By estimating a lower bound on this value using simple arguments, the accuracy of the estimate can be improved with smaller computational times. Determining valid estimates is a topic for future research.
– Limiting the number of times re-planning is carried out will reduce the size of the problem instances significantly. It is not known how much this reduction will affect the quality and robustness of the solution methods. Additionally, reduced re-planning intervals could be used only to generate the cost-to-go estimates, invoking a new policy at every stage.
– Is it possible to extract features of the problem to allow cost-to-go estimation based on these features, instead of a full blown simulation involving all the voxels in the problem? We believe such features as variance of the values of x_k, drop off near the boundary of the target, etc may be sufficient to calculate the cost-to-go estimates to suitable accuracy. Furthermore, the addition of new policies in the NDP approach may lead to better solutions. For example, if we know what features give good cost-to-go values, we can generate new policies to modify x_k appropriately. As a specific example, having uniform values for all voxels in x_k generates smaller cost-to-go values, and a "fill-up" policy that only delivers to voxels that are below a certain level may be useful to employ at certain stages.

While these approaches provide a variety of planning techniques that offer increasingly accurate treatment of errors in return for increasingly complex planning problems, it remains an open question as to which method (or methods) will be the most applicable in clinical practice.

Acknowledgements. This material is based on research partially supported by the National Science Foundation Grants ACI-0113051 and CCR-9972372, the Air Force Office of Scientific Research Grant F49620-01-1-0040, Microsoft Corporation and the Guggenheim Foundation. The authors would like to thank Geng Deng for his help in producing some of the computational results and figures given in this paper.

References

1. Bertsekas, D.P.: Differential training of rollout policies. In: Proceedings of the 35th Allerton Conference on Communication, Control, and Computing, 1997, Available as PDF document from `http://www.mit.edu:8001//people/dmitrib/Diftrain.pdf`

2. Bertsekas, D.P., Tsitsiklis, J.N.: Neuro-Dynamic Programming. Athena Scientific, Belmont, Massachusetts, 1996
3. Bortfeld, T.: Current status of IMRT: physical and technological aspects. Radiother. Oncol. **61** (2), 291–304 (2001)
4. Bortfeld, T., Schlegel, W.: Optimization of beam orientations in radiation therapy: some theoretical considerations. Phys. Med. Biol. **38** (2), 291–304 (1993)
5. Bortfeld, T.R., Jiang, S.B., Rietzel, E.: Effects of motion on the total dose distribution. Semin. Radiat. Oncol. **14**, 41–51 (2004)
6. Brahme, A.: Treatment optimization: Using physical and radiobiological objective functions. In: Radiation Therapy Physics, A.R. Smith, (ed.), Springer-Verlag, Berlin, 1995, pp. 209–246
7. Cho, P.S., Lee, S., Marks, R.J., Oh, S., Sutlief, S., Phillips, H.: Optimization of intensity modulated beams with volume constraints using two methods: cost function minimization and projection onto convex sets. Med. Phys. **25** (4), 435–443 (1998)
8. D'Souza, W.D., Meyer, R.R., Ferris, M.C., Thomadsen, B.R.: Mixed integer programming models for prostate brachytherapy treatment optimization. Med. Phys. **26** (6), 1099 (1999)
9. D'Souza, W.D., Meyer, R.R., Thomadsen, B.R., Ferris, M.C.: An iterative sequential mixed-integer approach to automated prostate brachytherapy treatment optimization. Phys. Med. Biol. **46**, 297–322 (2001)
10. Ferris, M.C., Lim, J.-H., Shepard, D.M.: Optimization approaches for treatment planning on a Gamma Knife. SIAM J. Optim. **13**, 921–937 (2003)
11. Ferris, M.C., Lim, J.-H., Shepard, D.M.: Radiosurgery treatment planning via nonlinear programming. Ann. Oper. Res. **119**, 247–260 (2003)
12. Ferris, M.C., Meyer, R.R., D'Souza, W.: Radiation treatment planning: Mixed integer programming formulations and approaches. Optimization Technical Report 02-08, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 2002
13. Fitchard, E.E., Aldridge, J.S., Reckwerdt, P.J., Mackie, T.R.: Registration of synthetic tomographic projection data sets using cross-correlation. Phys. Med. Biol. **43**, 1645–1657 (1998)
14. Hamacher, H.W., Küfer, K.-H.: Inverse radiation therapy planning – a multiple objective optimization approach. Disc. Appl. Math. **118**, 145–161 (2002)
15. Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. Preprint 01-8, Institut für Mathematik, Humboldt-Universität, Berlin, 2001
16. Holmes, T., Mackie, T.R.: A filtered backprojection dose calculation method for inverse treatment planning. Med. Phys. **21** (2), 303–313 (1994)
17. International Commission on Radiation Units and Measurements Inc. Prescribing, recording, and reporting photon beam therapy. ICRU Report 50, 1993
18. International Commission on Radiation Units and Measurements Inc. ICRU report 62: Prescribing, recording and reporting photon beam therapy (supplement to ICRU report 50). ICRU News, 1999. Available as HTML document from `http://www.icru.org/n_992_4.htm`
19. International Commission on Radiation Units and Measurements Inc. Prescribing, recording, and reporting photon beam therapy (supplement to ICRU report 50). ICRU Report 62, 1999
20. Jaffray, D.A., Siewerdsen, J.H., Wong, J.W., Martinez, A.A.: Flat-panel cone-beam computed tomography for image-guided radiation therapy. Int. J. Radiat. Oncol. Biol. Phys. **53**, 1337–1349 (2002)
21. Kapatoes, J.M., Olivera, G.H., Balog, J.P., Keller, H., Reckwerdt, P.J., Mackie, T.R.: On the accuracy and effectiveness of dose reconstruction for tomotherapy. Phys. Med. Biol. **46**, 943–966 (2001)
22. Kapatoes, J.M., Olivera, G.H., Reckwerdt, P.J., Fitchard, E.E., Schloesser, E.A., Mackie, T.R.: Delivery verification in sequential and helical tomotherapy. Phys. Med. Biol. **44**, 1815–1841 (1999)
23. Kapatoes, J.M., Olivera, G.H., Ruchala, K.J., Mackie, T.R.: On the verification of the incident energy fluence in tomotherapy IMRT. Phys. Med. Biol. **46**, 2953–2965 (2001)
24. Kapatoes, J.M., Olivera, G.H., Ruchala, K.J., Reckwerdt, P.J., Smilowitz, J.B., Balog, J.P., Keller, H., Mackie, T.R.: A feasible method for clinical delivery verification and dose reconstruction in tomotherapy. Med. Phys. **28**, 528–542 (2001)
25. Kapatoes, J.M., Olivera, G.H., Ruchala, K.J., Reckwerdt, P.J., Smilowitz, J.B., Balog, J.P., Pearson, D.W., and Mackie, T.R.: Database energy fluence verification and the importance of on-board CT imaging in dose reconstruction. In: Proceedings of the 13th International Conference on the Use of Computers in Radiation Therapy, W. Schlegel, T. Bortfeld, (eds.), Springer Verlag, Heidelberg, Germany, 2000, pp. 294–296
26. Keall, P.J., Kini, V.R., Vedam, S.S., Mohan, R.: Motion adaptive x-ray therapy: a feasibility study. Phys. Med. Biol. **46** (1), 1–10 (2001)
27. Kubo, H.D., Hill, B.C.: Respiration gated radiotherapy treatment: a technical study. Phys. Med. Biol. **41** (1), 83–91 (1996)
28. Küfer, K.-H., Hamacher, H.W., Bortfeld, T.: A multicriteria optimization approach for inverse radiotherapy planning. Symposium on Operations Research (OR 2000). Available as PDF document from `http://www.uni-duisburg.de/FB5/BWL/WI/or2000/sektion01/kuefer.pdf`

29. Langer, M., Brown, R., Urie, M., Leong, J., Stracher, M., Shapiro, J.: Large scale optimization of beam weights under dose-volume restrictions. Int. J. Radiat. Oncol. Biol. Phys. **18**, 887–893 (1990)
30. Langer, M., Morrill, S., Brown, R., Lee, O., Lane, R.: A comparison of mixed integer programming and fast simulated annealing for optimized beam weights in radiation therapy. Med. Phys. **23**, 957–964 (1996)
31. Lee, E.K., Fox, T., Crocker, I.: Optimization of radiosurgery treatment planning via mixed integer programming. Med. Phys. **27**, 995–1004 (2000)
32. Lee, E.K., Fox, T., Crocker, I.: Integer programming applied to intensity-modulated radiation therapy treatment planning. Ann. Oper. Res. (Optimization in Medicine) **119**, 165–181 (2003)
33. Lee, E.K., Gallagher, R.J., Silvern, D., Wuu, C.S., Zaider, M.: Treatment planning for brachytherapy: an integer programming model, two computational approaches and experiments with permanent prostate implant planning. Phys. Med. Biol. **44**, 145–165 (1999)
34. Lee, E.K., Zaider, M.: Mixed integer programming approaches to treatment planning for brachytherapy – application to permanent prostate implants. Ann. Oper. Res. (Optimization in Medicine) **119**, 147–163 (2003)
35. Li, J.G., Xing, L.: Inverse planning incorporating organ motion. Med. Phys. **27** (7), 1573–1578 (2000)
36. Lim, J.-H., Ferris, M.C., Wright, S.J., Shepard, D.M., Earl, M.A.: An optimization framework for conformal radiation treatment planning. Optimization Technical Report 02-10, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 2002.
37. Lind, B.K., Källman, P., Sundelin, B., Brahme, A.: Optimal radiation beam profiles considering uncertainties in beam patient alignment. Acta Oncol. **32**, 331–342 (1993)
38. Linderoth, J.T., Shapiro, A., Wright, S.J.: The empirical behavior of sampling methods for stochastic programming. Optimization Technical Report 02-01, Computer Science Department, University of Wisconsin, Madison, Wisconsin, 2002
39. Löf, J., Lind, B.K., Brahme, A.: Optimal radiation beam profiles considering the stochastic process of patient positioning in fractionated radiation therapy. Inverse Probl. **11**, 1189–1209 (1995)
40. Löf, J., Lind, B.K., Brahme, A.: An adaptive control algorithm for optimization of intensity modulated radiotherapy considering uncertainties in beam profiles, patient set-up and internal organ motion. Phys. Med. Biol. **43**, 1605–1628 (1998)
41. Mackie, T.R., Holmes, T., Swerdloff, S., Reckwerdt, P., Deasy, J.O., Yang, J., Paliwal, B., Kinsella, T.: Tomotherapy: a new concept for the delivery of dynamic conformal radiotherapy. Med. Phys. **20** (6), 1709–1719 (1993)
42. Mackie, T.R., Kapatoes, J., Ruchala, K., Lu, W., Wu, C., Olivera, G., Forrest, L., Tome, W., Welsh, J., Jeraj, R., Harari, P., Reckwerdt, P., Paliwal, B., Ritter, M., Keller, H., Fowler, J., Mehta, M.: Image guidance for precise conformal radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. **56**, 89–105 (2003)
43. McNutt, T.R., Mackie, T.R., Paliwal, B.R.: Analysis and convergence of the iterative convolution/superposition dose reconstruction technique for multiple treatment beams and tomotherapy. Med. Phys. **24**, 1465–1476 (1997)
44. Meyer, R.R., D'Souza, W.D., Ferris, M.C., Thomadsen, B.R.: MIP models and BB strategies in brachytherapy treatment optimization. J. Global Optim. **25**, 23–42 (2003)
45. Morrill, S.M., Lam, K.S., Lane, R.G., Langer, M., Rosen, I.I.: Very fast simulated annealing in radiation therapy treatment plan optimization. Int. J. Radiat. Oncol. Biol. Phys. **31**, 179–188 (1995)
46. Niemierko, A.: Optimization of 3D radiation therapy with both physical and biological end points and constraints. Int. J. Radiat. Oncol. Biol. Phys. **23**, 99–108 (1992)
47. Olivera, G.H., Fitchard, E.E., Reckwerdt, P.J., Ruchala, K.J., Mackie, T.R.: Delivery modification as an alternative to patient repositioning in tomotherapy. In: Proceedings of the 13th International Conference on the Use of Computers in Radiation Therapy, W. Schlegel, T. Bortfeld, (eds.), Springer Verlag, Heidelberg, Germany, 2000, pp. 297–299
48. Olivera, G.H., Ruchala, K., Lu, W., Kapatoes, J., Reckwerdt, P., Jeraj, R., Mackie, T.R.: Evaluation of patient setup and plan optimization strategies based on deformable dose registration. Int. J. Radiat. Oncol. Biol. Phys. **57** (2), S188–189 (2003)
49. Preciado-Walters, F., Rardin, R., Langer, M., Thai, V.: A coupled column generation, mixed-integer approach to optimal planning of intensity modulated radiation therapy for cancer. Tech. rep., Industrial Engineering, Purdue University, 2002
50. Ruchala, K.J., Olivera, G.H., Kapatoes, J.M.: Limited-data image registration for radiotherapy positioning and verification. Int. J. Radiat. Oncol. Biol. Phys. **54**, 592–605 (2002)
51. Ruchala, K.J., Olivera, G.H., Mackie, T.R.: A comparison of maximum-likelihood and iterative filtered backprojection reconstruction algorithms for megavoltage CT on a tomotherapy system. Phy. Med. Biol. 1999
52. Ruchala, K.J., Olivera, G.H., Schloesser, E.A., Mackie, T.R.: Megavoltage CT on a tomotherapy system. Phys. Med. Biol. **44**, 2597–2621 (1999)
53. Ruchala, K.J., Olivera, G.H., Schloesser, E.A., Reckwerdt, P.J., Mackie, T.R.: Megavoltage CT image reconstruction during tomotherapy treatments. Phys. Med. Biol. **45**, 3545–3562 (2000)

54. Sauer, O.A., Shepard, D.M., Mackie, T.R.: Application of constrained optimization to radiotherapy planning. Med. Phys. **26**, 2359–2366 (1999)
55. Schlegel, W., Mahr, A. (eds.): 3D Conformal Radiation Therapy - A Multimedia Introduction to Methods and Techniques. Springer-Verlag, Berlin, 2001
56. Shepard, D.M., Chin, L.S., DiBiase, S.J., Naqvi, S.A., Lim, J., Ferris, M.C.: Clinical implementation of an automated planning system for Gamma Knife radiosurgery. Int. J. Radiat. Oncol. Biol. Phys. **56**, 1488–1494 (2003)
57. Shepard, D.M., Ferris, M.C., Olivera, G., Mackie, T.R.: Optimizing the delivery of radiation to cancer patients. SIAM Rev. **41**, 721–744 (1999)
58. Shepard, D.M., Olivera, G., Reckwerdt, P.J., Mackie, T.R.: Iterative approaches to dose optimization in tomotherapy. Phys. Med. Biol. **45** (1), 69–90 (2000)
59. Verhey, L.J.: Immobilizing and positioning patients for radiotherapy. Semin. Radiat. Oncol. **5** (2), 100–113 (1995)
60. Voelker, M.M.: Optimization of Slice Models. PhD thesis, University of Wisconsin, Madison, Wisconsin, Dec. 2002
61. Webb, S.: Optimisation of conformal radiotherapy dose distributions by simulated annealing. Phys. Med. Biol. **34** (10), 1349–1370 (1989)
62. Webb, S.: Inverse planning for IMRT: the role of simulated annealing. In: The Theory and Practice of Intensity Modulated Radiation Therapy, E. Sternick, (ed.), Advanced Medical Publishing, 1997
63. Webb, S. The Physics of Conformal Radiotherapy: Advances in Technology. Institute of Physics Publishing Ltd., 1997
64. Wong, J.R., Grimm, S.L., Oren, R., Uematsu, M.: Image-guided radiation therapy of primary prostate cancer by a CT-linac combination: prostate movements and dosimetric considerations. Int. J. Radiat. Oncol. Biol. Phys. **57** (2), S334–335 (2003)
65. Wong, J.W., Sharpe, M.B., Jaffray, D.A., Kini, V.R., Roberson, J.M., Stromberg, J.S., Martinez, A.A.: The use of active breathing control (ABC) to reduce margin for breathing motion. Int. J. Radiat. Oncol. Biol. Phys. **44** (4), 911–999 (1999)
66. Wu, C., Jeraj, R., Olivera, G.H., Mackie, T.R.: Re-optimization in adaptive radiotherapy. Phys. Med. Biol. **47**, 3181–3195 (2002)
67. Wu, C., Olivera, G.H., Jeraj, R., Keller, H., Mackie, T.R.: Treatment plan modification using voxel-based weighting factors/dose prescription. Phys. Med. Biol. **48**, 2479–2491 (2003)
68. Xing, L., Cotrutz, C., Hunjan, S., Boyer, A.L., Adalsteinsson, E., Spielman, D.: Inverse planning for functional image-guided intensity-modulated radiation therapy. Phys. Med. Biol. **47**, 3567–3578 (2002)
69. Yan, D., Vicini, F., Wong, J., Martinez, A.: Adaptive radiation therapy. Phys. Med. Biol. **42**, 123–132 (1997)
70. Zavgorodni, S.F.: Treatment planning algorithm corrections accounting for random setup uncertainties in fractionated stereotactic radiotherapy. Med. Phys. **27** (4), 685–690 (2000)