Sampling Issues for Optimization in Radiotherapy*

Michael C. Ferris† Rikhardur Einarsson‡ Ziping Jiang§
David Shepard‡

July 17, 2006

*This material is based on research supported by the National Science Foundation Grant ACI-0113051 and the Air Force Office of Scientific Research Grant FA9550-04-1-0192

†Corresponding author, Computer Sciences Department, 1210 West Dayton Street, University of Wisconsin, Madison, WI 53706, U.S.A., email: ferris@cs.wisc.edu, tel: 608-262-4281

‡Industrial Engineering Department, 1513 University Ave. University of Wisconsin, Madison, WI 53706, U.S.A.

§Department of Radiation Oncology, 22 South Greene St. University of Maryland School of Medicine, Baltimore, MD 21201, U.S.A.

Abstract

A wide variety of optimization problems and techniques are used in radiation treatment planning. The problems typically involve large amounts of data, derived from simulations of patient anatomy and the properties of the delivery device. We investigate a three phase approach for their solution based on sampling of the underlying data that determines optimal beam angles, wedge orientations and delivery intensities in patient examples. Phase I uses multiple coarse samplings of the data and linear programming to adapt the sampling and determine a collection of promising angles to use. Phase II solves the adapted sample problems as mixed integer programs using only the promising angles. Phase III refines the sampling further, and fixes most of the discrete decision variables to reduce computation times. Particular emphasis will be given to general principles that are applicable to large classes of treatment planning problems. Specific examples show enormous increase in speed of planning, without detriment to the solution quality.

Keywords: Radiation treatment planning, adaptive sampling, optimization

The optimization of radiation therapy for cancer treatment has become an active research topic in recent years (Bortfeld et al., 1990; Bortfeld and Schlegel, 1993; Bortfeld et al., 1994; Chen et al., 2002; Intensity Modulated Radiation Therapy Collaborative Working Group, 2001; Jordan and Williams, 1994; Tervo and Kolmonen, 2000; Webb, 1998; Wu and Zhu, 2001; Xiao et al., 2003). Several survey articles cover the essential elements of the problem, see (Holder, 2004; Lodwick et al., 1998; Rosen et al., 1990; Shepard et al., 1999a). Many types of cancer are treated by aiming beams of ionizing radiation at the patient from a number of different angles. The fundamental goal is to apply a significant total dose of radiation to the cancerous region (the tumor) while sparing from excessive radiation the surrounding normal tissues (especially organs at risk near the tumor).

In this paper, we will concentrate on the use of mixed integer programming models for the solution of 3D conformal radiation therapy (3DCRT) treatment planning problems. While the form of these models is specific to particular treatment devices, this treatment procedure is widespread amongst clinics throughout the world, and the issues related to their solution are very similar to those that arise in other forms of treatment, such as Intensity Modulation Radiation Therapy (IMRT) (including Tomotherapy(Mackie et al., 1993; Olivera et al., 1999) and IMAT (Yu, 1995)). Further details on the delivery device and the resulting optimization models can be found in Section 1.

There are several very effective commercial solution engines for mixed integer programming, including CPLEX and XPRESS. While these techniques are very effective for large classes of models arising in a plethora of applications, their use in radiation therapy has been limited to this date for two reasons. Firstly, due to the large scale and dense nature of the problem instances, the solution techniques take large amounts of computational time, and in many cases these resource requirements are more than a planner is allowed. Large amounts of data are present in the models due to constraints that are expressed in a per voxel sense (the region of interest is divided into a collection of small three dimensional regions, typically called voxels). It is not uncommon for cases to have more than 60,000 voxels in the target and organs at risk. Secondly, and perhaps more seriously, the computational times have a very large variance, with solution times ranging from seconds to hours. The focus of this paper is on reducing both the mean and the variance of the computational times without adversely affecting solution quality.

Our approach to tshis problem is based on careful sampling of the voxels.

Figure 1 shows the effect of sampling on solution times for a pelvis example consisting of a prostate target, with the bladder and rectum as organs at risk of compromise (further details in Section 3). In this case, we sample at the rate indicated on the organs at risk, twice that rate on the target, and one tenth of that rate on the remaining normal tissue. In this setting, 1% sampling corresponds to 625 voxels. The boxplot is generated using 100 replications at each sample rate and brings out two points clearly. Firstly, the solution times increase dramatically as the sampling rate increases highlighting the enormous benefits of sampling. Secondly, the boxplot shows the large variance in solution times mentioned above. For example, at the 21% sampling rate, while the 25th and 75th percentiles are at 110 and 280 seconds respectively, the minimum solution time is less than 20 seconds while the maximum solution time for a particular replication is over 1200 seconds.

Figure 2 shows the flip side of the story and motivates our work. The gain in solution time comes at a price - when the sample becomes coarse, the mean and variance in the objective value increases considerably. For sampling rates up to 15% the variability of the objective values is too high and can lead in many cases to unacceptable solutions. Thus, running a single replication has a high risk of serious shortcomings in solution quality. Above a threshold rate of sampling, the reduction in means and variance tends to diminish (i.e. at 19% in Figure 2) due to better coverage.

As shown by Figure 2, naive sampling does not work. It is clear that as sampling becomes progressively coarse, the quality of the solutions deteriorates dramatically. Our computational approach attempts to rectify this situation by using information from the solution of linear and mixed integer sampled subproblems to refine the sample and restrict the variable domain to gain significant reduction in computational time, without the accompanying loss of solution quality.

After outlining the computational model in Section 1, we detail our sampling approach in Section 2.1, describing different approaches for inter- and intra-organ sampling. We propose a new technique for adaptively choosing samples based on the solutions of linear programming relaxations of our mixed integer model. These relaxations are also used to determine the initial sample size. The use of multiple samples and multiple phases is described in Section 2.2, building on our previous work in Lim et al. (2006). Our general three phase approach is also outlined in this section. Subsequently, we detail each phase, describing new ways in which we use the weights determined from the linear program solutions to remove unpromising angles, and make

optimization choices using ranking from true objective values of sampled solutions. Section 3 demonstrates the efficiency of our approach on a number of patient examples showing impressive gains in speed across all examples.

1 Computational Model

As mentioned above, we will limit our planning tool to 3DCRT. Conformal radiotherapy treats a cancer patient using a linear accelerator delivering beams of radiation to a body from a variety of angles at different intensity levels, using an adaptive shaping of the beam. Enhanced conformation to the target delivers more dose to the tumor while sparing surrounding normal tissue. A beam's eye view (Mohan, 1995) of the target (determined for a particular tumor for each angle $A \in \mathcal{A}$ from which radiation is delivered) is physically shaped using a multi-leaf collimator (Bortfeld et al., 1994; Brahme, 1993; Brewster et al., 1995; Du et al., 1994).

A metallic wedge filter (Leavitt et al., 1990) can be attached in front of the collimator that attenuates the intensity of the radiation in a linear fashion from one side to the other. Several orientations are possible represented by $F \in \mathcal{F}$. This is useful in changing the dose distribution along curved patient surfaces; for example breast cases. Non-coplanar beams and different energy levels can also be used to change the dose that is delivered. We limit our discussion here to beam angle selection; the ideas are perfectly generalizable to these other features at the cost of increased data and more binary decision variables. For future reference, $\mathcal{A}$ and $\mathcal{F}$ represent the collection of angles and fields respectively.

Manipulating the various features constitutes the delivery plan. In this paper we concentrate on a mixed integer programming (MIP) approach; other techniques based on linear programming, nonlinear optimization, simple heuristics and iterative approaches have also been extensively used. For example, we can use a binary variable $\psi_A \in \{0, 1\}$ to determine whether we use angle A or not (i.e. we use it if $\psi_A = 1$). This can be enforced in a linear model using

$$0 \leq w_{A,F} \leq W\psi_A$$

where W is an upper bound on the exposure from angle A, F indexes the different wedge fields and $w_{A,F}$ represents the exposure time from a particular angle with the given wedge field. To limit the number of angles we use in

our plan to p, the following constraint suffices:

$$\sum_A \psi_A \leq p$$

The superposition principle is used to generate the total dose delivered to a given voxel i:

$$Dose(i) = \sum_{A,F} w_{A,F} D_{A,F}(i)$$

where $D_{A,F}(i)$ is the dose data, determining the dose delivered to voxel i from angle A through wedge field F. In our treatment planning system the dose distribution is calculated via a Monte Carlo simulation. Different cutoff values for which $D_{A,F}(i)$ is set to zero determine the density of the data and hence of the optimization problem; frequently higher cutoff levels lead to a "terma" dose approximation (Shepard et al., 1999b).

The complete model is:

$$\begin{array}{rll} \min_{w,\psi} & f(Dose) & \\ \text{s.t.} & Dose(i) = \displaystyle\sum_{A\in\mathcal{A},F\in\mathcal{F}} w_{A,F} D_{A,F}(i), & i \in O_j,\ O_j \in \mathcal{O}, \\ & 0 \leq w_{A,F} \leq W\psi_A, & \forall A \in \mathcal{A},\ \forall F \in \mathcal{F}, \\ & \displaystyle\sum_{A\in\mathcal{A}} \psi_A \leq p, & \\ & \psi_A \in \{0,1\}, & \forall A \in \mathcal{A}. \end{array} \tag{1}$$

Here $\mathcal{O}$ runs over the collection of organs including the target (PTV), the organs at risk and the remaining "normal" tissue. In many cases, apart from the regions of interest, there are many voxels present in the problem data that reflect "other" tissue. We shall refer to these voxels as "normal tissue", although in some cases this term is also used to describe the organs at risk.

The objective function can take on many different forms. In this work we utilize the following piecewise linear objective function:

$$f(Dose) = \sum_{O_j\in\mathcal{O}} \sum_{i\in O_j} \frac{\lambda_{O_j}}{\text{card}\,(O_j)} \max\left(\pm Dose(i) - \theta_{O_j}, 0\right)$$

where (with a slight abuse of notation) $\mathcal{O}$ runs over a collection of structures (with possible replications to allow multiple penalization terms for the same organ), λ represents objective weightings and θ represent threshold values (the $\pm$ term allows either for over or underdosing terms). These values can

be set to approximate dose volume histogram constraints (see Langer et al., 1990), or other piecewise linearizations can be used to implement so-called CVAR constraints (Romeijn et al., 2006), or standard infinity or L_1 norm constraints. Further details on these choices are available in Lim et al. (2006). Clearly, different settings of these parameters model different facets of the physicians problem and correct choices are crucial to generate appropriate solutions. Some issues that we will ignore here are overlapping structures (where a single voxel is in more than one structure, which may occur due to planning error margins) and how to treat normal tissue that is close to the target. We simply choose to include the voxels from overlapping structures into all the objectives in which they occur, and we remove a small rind from the normal tissue around target. Both of these choices could be made differently, and much of their effects can be compensated for by different choice of λ and θ. Similarly, this paper ignores the clinical effects of homogeneity or minimum/maximum dose constraints on the PTV, but allows for parameterizations of λ and θ to replicate different effects.

Throughout the paper we will assume this standard model and consider a single optimization of it. We will further assume that the optimization model is parameterized in a correct fashion so that smaller values of the objective correspond to preferable solutions.

Our approach replaces the objective function f by a sample approximation. For each structure O_j we let

$$\tilde{f}(Dose, O_j, \mathcal{S}(O_j)) = \sum_{i \in \mathcal{S}(O_j)} \frac{\lambda_{O_j}}{\text{card}\left(\mathcal{S}(O_j)\right)} \max\left(\pm Dose(i) - \theta_{O_j}, 0\right),$$

where $\mathcal{S}$ represents a sampling function of its argument. Note that with this notation the objective function $f(Dose)$ can be rewritten as

$$f(Dose) = \sum_{O_j \in \mathcal{O}} \tilde{f}(Dose, O_j, O_j).$$

We replace this expression with

$$\sum_{O_j \in \mathcal{O}} \tilde{f}(Dose, O_j, \mathcal{S}(O_j))$$

Since $Dose(i)$ is only needed for those voxels in the sample, this approxima-

tion leads to significant computational savings in the optimization problem:

$$\begin{array}{rll} \min_{w,\psi} & \sum_{O_j \in \mathcal{O}} \tilde{f}(Dose, O_j, \mathcal{S}(O_j)) & \\ \text{s.t.} & Dose(i) = \sum_{A \in \mathcal{A}, F \in \mathcal{F}} w_{A,F} D_{A,F}(i), & i \in \mathcal{S}(O_j),\ O_j \in \mathcal{O}, \\ & 0 \leq w_{A,F} \leq W \psi_A, & \forall A \in \mathcal{A},\ \forall F \in \mathcal{F}, \\ & \sum_{A \in \mathcal{A}} \psi_A \leq p, & \\ & \psi_A \in \{0,1\}, & \forall A \in \mathcal{A}. \end{array} \tag{2}$$

However, we note that given a particular weight vector w, it is normally very easy and computationally inexpensive to calculate the value of the function $f(Dose)$ and its per organ components

$$\tilde{f}(Dose, O_j, O_j)$$

evaluated as a sum over all the voxels instead of over the sampling of the voxels. This choice (which we will term the true objective value) will be independent of the particular sample and will ensure our comparisons of different sampling schemes are accurate.

The choice of the sampling function $\mathcal{S}$ is the subject of the remainder of this paper.

2 Computational Approach

2.1 Sampling method

Sampling consists of two parts, namely intra-organ and inter-organ effects. Inter-organ effects determine the relative rates of sampling between different organs. Intra-organ effects determine where we sample within a particular organ. We deal with these effects separately.

We first investigate how many voxels to place in each structure (inter-organ sampling). Small structures can generate problems for sampling. For example, (horizontal patient slices of) structures like the spinal cord are small compared to structures like the liver. If we sample aggressively, we may end up with very few (or no) voxels in the sampled structure, and hence our ability to impose constraints on the structure is severely restricted. For this purpose, we suggest allowing a planner to specify a minimum sample size on each structure.

The "normal" tissue typically contains many more voxels than any other structure. In many cases there are unspecified but implicit upper bound constraints on this tissue - a physician does not want to have "hot spots" on the normal tissue, as this may give rise to complications, blemishes or unnecessary discomfort. We shall call this problem the "streaking effect". Thus, while the (large number of) normal tissue voxels are a prime candidate for aggressive sampling, we try to ensure that our solutions do not display streaking effects.

Figure 3 indicates some of the effects of changing the relative sample rates in each of the structures on the true objective value on the pelvis case. Method 1 uses the same number of sample points in each structure, including the normal tissue. The other three methods sample in a uniform manner proportionally to the original structure size, with the exception of the normal tissue. For such "leftover" tissue, we limit the size of the sample to 2, 5 or 20 times the size of the largest other structure sample. Several points can be drawn from these results. Firstly, method 1 does not work well, and is outperformed by the proportional sampling techniques. Secondly, there is little point oversampling the normal tissue, a sample size of 5 times the largest other structure is more than adequate. Thirdly, as the sample size gets larger, the solution quality improves (see vertical axis on each graph in Figure 3).

Due to concerns regarding the choice of these factors, and whether different structures need finer sampling than others, we show how to adaptively choose the sample sizes in each structure based on feedback from optimization solutions.

The true objective function values give an indication of how accurately the sampled solution reflects the real problem. The values of the true objective function on its constituent structures generate errors

$$error(O_j) = \left| \tilde{f}(Dose, O_j, O_j) - \tilde{f}(Dose, O_j, \mathcal{S}(O_j)) \right|$$

that give information about where the underlying sample is deficient at modeling the particular case. As shown in Figure 4, as the sample size increases, the error between the sampled and true objective function, and its constituent parts tends to decrease (the figure plots the mean error over the 10 replications). However, the errors on some structures are more serious than others, see for example, the bladder and the rectum in Figure 4. This suggests that rather than increasing the samples in each structure proportionally, we

should *adapt* the size of the the samples in each structure separately, based on the results of known optimization solutions.

We suggest a simple adaptive scheme along these lines:

1. Using solutions of (2) for a coarse sampling, generate errors using the above formula.

2. In each structure retain half of the sampled voxels.

3. The remaining K/2 voxels are regenerated and are distributed among the structures according to the ratios

$$\frac{error(O_j)}{\sum_a error(O_a)}$$

4. Adjust these ratios so the minimum and maximum numbers of voxels in each structure are satisfied.

5. Uniformly generate samples within each structure.

Thus the samples are made larger in those organs where the discrepancies are the greatest.

Figure 5 compares uniform sampling against adaptive sampling on the pancreas case of Section 3. The boxplots result from solving the optimization problem as a MIP to optimality, displaying the true objective value over 30 replications. The first two boxplots relate to proportional sampling while the second two relate to adaptive sampling, both with 600 voxels. These results indicate that both adaptive schemes outperform the proportional sampling scheme (some further indication of this is given in Figure 9). While we only show this single result here, the improvements due to adaptive sampling scheme occur for all examples and at all aggressive sample rates. Note that the adaptive samples take longer to solve than the proportional samples (by a factor of approximately 2-4 times).

For intra-organ effects, we consider three options, namely grid sampling, uniform sampling and rind sampling. Grid sampling evaluates the objective function only at voxels within a regular grid. Uniform sampling chooses the voxels uniformly at random. Rind sampling changes the sampling rate in areas close to the periphery of the organ. We first consider the issue of grid sampling compared to uniform sampling.

Returning to Figure 5, we see a comparison of uniform sampling against grid sampling on the pancreas case of Section 3. The first two boxplots relate to proportional sampling with 600 voxels and the size of the normal tissue sample limited to the size of the largest structure, while the second two use the adaptive scheme outlined above. Once the sample sizes were fixed, we either generated solutions using uniform sampling in each organ, or via the following (two step) form of grid sampling. In the first step, we choose a grid size so the resulting sample is slightly larger, but as close as possible to the required sample size. In the second step, this sample of grid points is then uniformly sampled to generate the requested sample size. Thus both the uniform and grid samples have exactly the same number of voxels in each structure.

The results are mixed: grid sampling appears to be better than uniform sampling under proportional inter-organ sampling, while the converse is true under the adaptive scheme, and this is confirmed by a Kolmogorov-Smirnov goodness-of-fit test. These results indicate that the spreading out of sample voxels that is achieved by grid sampling is important when certain structures are undersampled, but that when the inter-organ sample sizes are chosen more carefully (as in the adaptive scheme) then uniform sampling is as good, probably better than grid sampling.

To analyze further the effect of sampling differently within a particular structure, we consider varying the sampling rate in different parts of the target structure. The motivation for such variations is primarily the large changes (gradients in dose distribution) that are likely to occur close to the boundaries of the target and the OAR's. In our test, the target was divided into two volumes: central and rind. For our experiments we used the pelvis example limiting the normal tissue sample to be five times the size of the largest other structure, and set the rind of the target to be every voxel within 2 voxels of the boundary. Then different sampling scenarios were tested and the true objective values used to evaluate the outcome.

Figure 6 shows the change in true objective value as a function of α, the fraction of voxels in the rind of the target. When α is small, there are many more voxels centrally located within the target and the objective function increases. For values of α between 0.65 and 0.95, the objective is lower and does not vary greatly. Since in this example, uniform sampling corresponds to a value $\alpha \approx 0.70$, it seems that rind sampling is not significantly better than uniform sampling. Clearly, further experimentation could be carried out on different cases and modifying the sampling strategies on different

structures, along with different rind sizes, but we believe this is unlikely to produce significant improvements in objective values.

In conclusion, we believe the adaptive inter-organ sampling technique is very reliable and adjusts to differences in problems extremely well. Among the three intra-organ options we investigated we only see small differences and hence choose the simplest, namely uniform sampling within a structure for the remainder of this paper.

2.2 Multiphase optimization

We observe that the optimization problem (2) has two connected components. At a higher level, there is the choice of which angles to deliver radiation from, after which the exposure levels w must be determined to satisfy constraints, many of which are specified at a very fine (voxel) level. As we suggested in Lim et al. (2006), we will split the decision process into several phases. In each phase we will either refine the sample, increase the complexity of the mathematical model or reduce the number of binary decision variables that need to be determined.

Figure 7 justifies the use of a reduced set of angles. In this figure, we show the solution time to optimality for a series of optimization where the angle choice set has different sizes. Since in this pancreas case, we are choosing 3 angles, we consider angle set choices of multiples of 3. The actual sets of angles chosen for Figure 7 were determined by extending the optimal set of three angles using angles that appeared frequently over an extensive set of linear programming solutions. While these sets may not be indicative of sets that our procedure would generate, they are certainly challenging computationally. It is clear that reducing the angle choice set size leads to significant improvements in solution time. The critical issue at hand is how to determine a good reduced set of angles at coarse sampling, allowing us then to optimize with such a set of angles on a finer sample.

Another outcome of the above experiment is Table 1 that shows the sensitivity of the solution to sampling. As we report in Section 3, the optimal solution of the pancreas problem with a full data set (36 angles and complete voxel set) involves the angles 80, 290, 350. Table 1 shows how many of the replications reproduce this solution, and the frequencies for which the optimal true objective value from the sampled problems are close to the true optimal. The table highlights the fact that while a sampled MIP generates a good solution quickly, multiple replications of the solution of the MIP

Table 1: Optimal angle set choice from 10 replications at sample size 3600

Frequency	Number of angles to choose from											
	3	6	9	12	15	18	21	24	27	30	33	36
Optimal set chosen	10	6	4	4	3	3	3	3	3	3	3	3
Objective within 5% of optimal	9	5	5	5	4	4	4	4	4	4	4	4
Objective within 10% of optimal	10	8	9	8	7	7	7	7	7	7	7	7

sampled problems may be needed in order to ensure capturing the optimal solution of the original problem.

The broad outline of our complete sampling approach is as follows:

1. Determine the initial sample size K.

2. Phase I: use all angles with 10 sample LP's of size K. Adapt the samples based on differences between the sampled and true objective values in each organ. Resolve the 10 sample LP's on the adaptive sample.

3. Phase II: use a reduced set of angles determined from Phase I and solve 10 sample MIP's with the same adaptive samples, and determine a further reduced angle set.

4. Phase III: using the reduced angle set from Phase II but with a refined sample, solve a single MIP to determine optimal solution; essentially this is a clean up phase to guarantee problem specific requirements are accurately met.

The use of 10 samples was chosen heuristically: there is a trade off between computational time and treating the distribution of the sample values more accurately. The value of 10 seems to improve reliability of the solutions within an acceptable time limit.

In the sequel we detail each of these parts in turn.

2.3 Initial sample size

While the MIP models can have large solution times, and large variance in solution times, the linear programming relaxation of these models (where ψ_A is allowed to take on any value in $[0, 1]$) is much quicker (allowing larger

sample rates or more replications) and its time variance is much smaller, leading to more predictable solution times. We note that the LP relaxation is solved using the dual simplex for small sample sizes and the barrier method for large sample sizes (see Olafsson and Wright, 2006 for more details on how to reformulate and solve these linear programs).

Our use of the linear program is threefold: initial sample size, adaptive inter-organ sampling and angle reduction.

The initial sample size is chosen as follows. We first guess a trial sample size K. Having K, we generate a single uniform sample of the problem of interest using proportional inter-organ sampling, but with the "normal tissue" limited to the same size as the largest other structure. For this sample, we solve the linear programming relaxation of (2).

We then continue to double K and repeat the above process until the time to solve the resulting linear program becomes unacceptable. This assumes the user has specified a time limit for the complete process that can be used to determine what is an acceptable time for the linear program solution, bearing in mind that the solution time for the MIP is likely to be at least a factor of 10 to 100 longer than the LP time. At this stage, if true objective function with sample size K and $2K$ differ by too much, then inform the user of this fact and terminate. Otherwise, use K as the initial sample size. Clearly, the amount by which the samples can differ is problem dependent here. We have utilized a multiplicative factor of 10 to 20 percent. Typically, very few linear programs are needed to determine the initial sample size.

2.4 Phase I: Linear Programming

Unfortunately, the (sampled or true) objective function value that is returned from the linear program is not immediately useful, since it most likely corresponds to an infeasible MIP solution - we deliver from more than the allowable p angles. However, we will endeavor to show in this section that the linear programming relaxation can be used effectively for two more purposes, namely adaptation of the sample, and gross prediction of the set of promising angles.

For adaptive sampling, we use the solutions of the linear programming relaxations to determine errors between the sampled and the true objective function. The drawbacks of the objective function not being representative of a feasible MIP solution are irrelevant for this purpose; we simply need to discover in which organs our sample poorly approximates the true objective.

Predictor	LP	Proportional Sample				Adaptive Sample			
	reps	9	12	15	18	9	12	15	18
Frequency	10	0.75	0.49	0.26	0.11	0.59	0.31	0.12	0.04
Frequency	20	0.96	0.51	0.03	0.00	0.81	0.14	0.00	0.00
Frequency	30	0.98	0.56	0.01	0.00	0.87	0.10	0.00	0.00
Frequency	40	0.99	0.59	0.00	0.00	0.92	0.07	0.00	0.00
Weights	10	0.88	0.56	0.32	0.13	0.40	0.16	0.07	0.01
Weights	20	0.85	0.53	0.25	0.05	0.28	0.10	0.03	0.00
Weights	30	0.86	0.52	0.22	0.03	0.22	0.07	0.01	0.00
Weights	40	0.87	0.52	0.21	0.02	0.18	0.05	0.00	0.00

Table 2: Average number of errors between suggested set and optimal angle set (size 3); bootstrap sampling, 100 replications, pancreas example, sample size 600

Thus the adaptive scheme of Section 2.1 is applied to the LP solutions, averaging the results over the 10 replications.

To improve our solutions in terms of mean and variance we use multiple samples. After solving the 10 sampled linear programs, we attempt to use the solutions to predict promising angles to use for subsequent optimizations. The collection of angles that are used in at least one of the linear programming solutions is too large to be useful. We need to devise a scheme to further reduce this set. Three measures can be used; namely the true objective, the frequency counts on use of angles, and the weights w that are used from each angle. For the weight calculation, we simply summed up all the weights for that angle using any wedge field. We discount the true objective since the linear programming solutions are not even feasible for the true problem so these values tend to be poor predictors.

We count how many angles are missing from the optimal set to determine how effectively we predict the optimal set of angles. We average this count over 100 replications and report the results for the pancreas case (Table 2) and the pelvis case (Table 3). In these tables, "LP replications" indicates how many linear programs we solved to generate our prediction. The columns relate to the size of our predictor set. When the number of LP replications was greater than 10, we used bootstrap sampling to generate further likely replications.

Since weights and frequency counts are correlated, both tables have similar overall conclusions. In particular, it is better to use the adaptive sample

Predictor	LP	Proportional Sample				Adaptive Sample			
	reps	9	12	15	18	9	12	15	18
Frequency	10	2.06	1.60	1.19	0.86	1.95	1.44	1.08	0.77
Frequency	20	1.99	1.61	1.00	0.75	1.97	1.15	0.92	0.45
Frequency	30	2.00	1.66	1.00	0.81	1.99	1.10	0.94	0.44
Weights	10	2.01	1.50	1.09	0.73	1.99	1.21	0.92	0.64
Weights	20	2.00	1.31	1.00	0.78	1.93	1.02	0.99	0.60
Weights	30	2.00	1.30	1.00	0.83	1.96	1.01	0.99	0.61

Table 3: Average number of errors between suggested set and optimal angle set (size 6); bootstrap sampling, 100 replications, pelvis example, sample size 600

rather than the proportional sample for prediction of angles. The value of increasing the number of linear programming replications is unclear; certainly there is not much benefit at all beyond 20 LP replications (for the patient examples we tested on). We would suggest that it is better to increase the sample size rather than run further LP replications. Thus for the remainder of this paper, we use 10 LP replications. Using adaptive sampling, we believe the above tables demonstrate that using the weights for prediction leads to better predictions, and hence we utilize this scheme in the remainder of our work.

We pursue a cautious approach with regard to dropping angles from consideration, particularly in Phase I. Thus, for the patient example cases we describe in Section 3 we use the prediction of 18 angles from Phase I, rather than a smaller (and more error prone) prediction set.

2.5 Phase II

Phase II utilizes the reduced set of angles predicted from Phase I, and the adaptive samples generated in that phase. We solve 10 replications of the MIP problem with such data to optimality (increasing the optimality gap is a matter for further research but preliminary results show this can seriously degrade the solution).

The main purpose of this phase is to further reduce the size of the predicted angle set. Typically, we set the size of the prediction set from Phase II to be no greater than twice the number of angles requested in the solution $2p$.

In this setting, we believe the weights on each angle that were used in Phase I are not a good predictor, and we turn instead to using frequency count or the true objective value. Figure 8 indicates the set of angles that are identified in the pelvis case (see Section 3) with the frequency $\geq 20\%$ method. Note that as the sampling rate increases, the size of the set identified decreases. Furthermore, the set includes all 6 angles chosen in the optimal solution (namely 40,80,130,240,270,300) at sampling rates above 9%. However, for lower sampling rates, at least one of these angles (typically 40) is dropped from the proposed set.

The true objective values can show the deficiencies of the histogram selection scheme outlined above. In particular, it can happen that a very good solution appears only once in the sampled set, and the histogram approach does not select the associated angles, particularly with coarse sampling. Thus a rule that uses angles if they appear in at least 20% of the sampled solutions may lose the optimal solution, even if one of the samples determined that solution. Our alternative approach calculates the true objective value for each replication, and utilizes the information from this calculation. The sampled solutions are ranked according to their true objective values, and angles are selected from this ordered list until a threshold is reached. The threshold involves a maximum number of angles, and a maximum number of sampled solutions to consider. Thus, for example, we may rank 10 solutions, and select up to 6 angles from the best 5 sampled solutions.

The quality of the predicted set using this alternative approach is significantly better than the frequency approach at low sample rates. For larger sampling rates, the two schemes tend to predict the same set of angles. For this reason, we use the true objective value prediction mechanism in Phase II.

2.6 Phase III

In Phase III, we generate an accurate solution. The set of angles used is restricted from Phase II, and sometimes may have fixed all the angle choices (resulting in Phase III being simply a linear program). We add sample points using the same ratios that we predicted in our adaptive sampling scheme, and redistribute those voxels that exceed the total number of voxels available in a given organ evenly amongst the other organs. Rather than solve multiple instances, we simply solve a single problem in this case. For the results given in the next section we multiplied the initial sample size chosen by a factor of

10 for this phase.

In Figure 9 we can see how the true objective value changes through our process. The first two columns show how powerful adaptive sampling is. Phases I and II use 600 voxels while Phase III involves 6000 voxels. Outliers have even been removed from the first run in Phase I for better readability. The second run in Phase I uses an adaptive sample based on solution information from the first run. This decreases the mean, variance and median of the true objective values dramatically. However, the solutions in Phase I are from relaxations of the MIP and are therefore not feasible for the MIP. This explains why the latter run in Phase I has lower true objective value than that of Phase II. Phase I solutions show that coarse samples still provide very useful information to help adapt samples as coverage deficiencies are clearly identified. In Phase II the increase in true objective value is small, both in mean and variance and Phase III recovers accuracy and has extremely small variance. This motivates the use of a single replication in this Phase.

3 Results on Patient Cases

This section gives the results of applying our sampling scheme to four different patient cases. Throughout our experimentation we used the CPLEX code (version 9.0), with tight tolerances (optcr = 0.001) on optimality. We used the following options for CPLEX. On some problems it is beneficial to change the initial linear programming solver option to use the barrier method instead of the dual simplex code, but for the majority of our sampled cases, these options resulted in smaller solution times:

```
heurfreq -1
probe -1
flowcovers -1
mircuts -1
```

Our four test problems have the following characteristics:

Pelvis case. Prostate: 2,760 voxels and 4 pieces in the objective; Bladder: 1,440 x 2; Rectum: 833 x 2; Normal (for 5 times other largest structure): 13,800 x 1; Normal (for 10 times other largest structure): 27,600 x 1.

Pancreas case. Pancreas: 5,659 x 6; Cord: 515 x 1; Left kidney (ltt): 9,273 x 2; Right kidney (rtt): 6,014 x 2; Liver: 53,615 x 2; Normal: 268,075 x 1.

Case	Avg Time [sec]	STD [sec]	Full Obj.	Angles in 10 degree inc	Variables
Pelvis5	1,018	232	1.6460	4,8,13,24,27,30	32,219
Pelvis10	4,294	1,546	1.6457	4,8,13,24,27,30	46,019
Pancreas	345,930	24,871	0.2316	8,29,35	446,224
Breast	82,736	47,289	2.5549	13,29	335,575
Head, Neck	46,765	12,279	0.6274	4,14,23	195,253

Table 4: Results of different cases with all voxels of structures except the normal structure.

Breast case. Left breast: 38,776 x 6; Heart: 12,552 x 3; Left lung: 10,851 x 3; Normal: 38,776 x 2; Total = 380,417; 80% of total = 304,330.

Head and neck case. tumor: 2,112 x 6; brainstem: 2,381 x 1; cerebellum:13,525 x 1; optic chiasm: 15 x 1; left cerebrum: 55,363 x 2; left eye: 734 x 1, left lacrimal: 15 x 1; left optic nerve: 46 x 1; right cerebrum: 51,698 x 1; right eye: 833 x 1; right optic nerve: 63 x 1. Note the existence of some very small structures in this example.

To benchmark our results, we first attempted to solve the problems to optimality without sampling. In Table 4, we should note that the Pelvis10 case is run with the number of normal voxels being 10 times the max of other structures, and the breast case could only be run on 80% of full data problem. Note also that the number of variables reported here is different from the number of voxels; in some cases the "Dose" variable is eliminated directly, and in other cases we have extra variables to treat the piecewise linear nature of the objective function.

Based on the analysis of the previous sections of this paper, we constructed a script that implements the complete sampling process which we ran on all these four examples. In each case, we show the output of our solution process using a dose volume histogram and a doseplot. Notice that the dose conforms well in the targets, limits dose to the organs at risk and avoid streaking effects. The specific prescription goals of the planner were met in each case.

Pelvis case. More details of the solution procedure are given below:

```
Sample size: 2400
```

```
Phase I  angles:   4, 5, 6, 7, 8, 9,10,11,12,13,14,15,24,27,29,30,31,32
Phase II angles:   6, 8, 9,10,11,13,14,24,27,30,31,32
Phase III angles:  8,11,13,24,27,32
Objective:        1.6498

Total time:     197 secs
Phase I time:    32 secs
Phase II time:   18 secs
Phase III time: 147 secs
```

Note that in this example the solution is not the same as that found by the full optimization. However, the objective values of this solution and the optimal one differ only in the third decimal place. To generate the same solution, the sample size can be increased to 4800, with an accompanying increase in time (to around 500 seconds total).

Pancreas case.

```
Sample size: 600
Phase I  angles:   0, 4, 5, 6, 7, 8, 9,16,17,26,28,29,30,31,32,33,34,35
Phase II angles:   0, 7, 8,29,30,35
Phase III angles:  8,29,35
Objective:        0.2317

Total time:     290 secs
Phase I time:     5 secs
Phase II time:  229 secs
Phase III time:  56 secs
```

The resulting solution is depicted in Figure 10.

Breast case.

```
Sample size: 600
Phase I  angles:   2, 7,12,13,14,19,20,21,26,27,28,29,30,31,32,33,34,35
Phase II angles:  13,29,30
Phase III angles: 13,29
Objective:        2.5577
```

```
Total time:     88 secs
Phase I time:   29 secs
Phase II time:  56 secs
Phase III time:  3 secs
```

The resulting solution is depicted in Figure 11.

Head and Neck case.

```
Sample size: 600
Phase I  angles:  3, 4, 5, 7,10,14,15,16,17,18,19,20,21,22,23,26,27,33
Phase II angles:  4, 5, 7,14,22,23
Phase III angles: 4,14,23
Objective:        0.6277

Total time:     68 secs
Phase I time:    5 secs
Phase II time:  53 secs
Phase III time: 10 sec
```

The resulting solution is depicted in Figure 12.

4 Conclusion

This paper described a new approach to sampling in 3DCRT problems. In Section 3 we demonstrated significant reductions in solution times without loss of quality, as indicated both by objective values, final decision variables, and a series of dose volume histograms and dose plots.

Our results show that very aggressive sampling can be used provided multiple replications are carried out and the sample is chosen adaptively using feedback from optimization solutions. Specific contributions of this work include novel use of linear programming relaxations to adapt the sample (Section 2.1), use of multiple LP replication solutions to do a gross reduction of binary variables, ranking of MIP solutions to further reduce the size of angle set predictions based on true objective values (as compared to frequency counts or sampled objective values), and a detailed discussion and analysis of various choices in the overall sampling scheme.

Our work demonstrates general principles that we believe are applicable to many treatment planning problems. In particular, it is straightforward to extend the 3DCRT code to include energy levels and non-coplanar beams as well as angles and wedges. Its utility in IMRT, tomotherapy and IMAT need further investigation since in many instances these problems are even larger and more complex. This remains a topic for future research, in addition to the open questions we have posed within this paper.

References

Bortfeld, Th., A. L. Boyer, W. Schlegel, D. L. Kahler, and T. J. Waldron (1994). Realization and verification of three-dimensional conformal radiotherapy with modulated fields. *International journal of radiation oncology: Biology, Physics* 30(4): 899–908.

Bortfeld, Th., J. Burkelbach, R. Boesecke, and W. Schlegel (1990). Methods of image reconstruction from projections applied to conformation radiotherapy. *Physics in Medicine and Biology* 25(10): 1423–1434.

Bortfeld, Th. and W. Schlegel (1993). Optimization of beam orientations in radiation therapy: some theoretical considerations. *Physics in Medicine and Biology* 38(2): 291–304.

Bortfeld, Th. R., D. L. Kahler, T. J. Waldron, and A. L. Boyer (1994). X-ray field compensation with multileaf collimators. *International Journal of Radiation Oncology, Biology and Physics* 28(3): 723–730.

Brahme, A. (1993). Optimization of radiation therapy and the development of multileaf collimation. *International Journal of Radiation Oncology, Biology and Physics* 25: 373–375.

Brewster, L., R. Mohan, G. Mageras, C. Burman, S. Leibel, and Z. Fuks (1995). Three dimensional conformal treatment planning with multileaf collimators. *International Journal of Radiation Oncology, Biology and Physics* 33(5): 1081–1089.

Chen, Y., D. Michalski, C. Houser, and J. M. Galvin (2002). A deterministic iterative least-squares algorithm for beam weight optimization in conformal radiotherapy. *Physics in Medicine and Biology.* 47: 1647–1658.

Du, M. N., C. X. Yu, M. Symons, D. Yan, R. Taylor, R. C. Matter, G. Gustafson, A. Martinez, and J. W. Wong (1994). A multileaf collimator field prescription preparation system for conventional radiotherapy. *International Journal of Radiation Oncology, Biology and Physics* 30(3): 707–714.

Holder, A. (2004). Radiotherapy treatment design and linear programming. In Brandeau, M. L., F. Sainfort, and W. P. Pierskalla, editors, *Operations Research and Health Care: A Handbook of Methods and Applications*, pp. 741–774. Kluwer Academic Publishers, Boston.

Intensity Modulated Radiation Therapy Collaborative Working Group (2001). Intensity-modulated radiotherapy: Current status and issues of interest. *International Journal of Radiation Oncology: Biology, Physics* 51(4): 880–914.

Jordan, T. J. and P. C. Williams (1994). The design and performance characteristics of a multileaf collimator. *Physics in Medicine and Biology.* 39: 231–251.

Langer, M., R. Brown, M. Urie, J. Leong, M. Stracher, and J. Shapiro (1990). Large scale optimization of beam weights under dose-volume restrictions. *International Journal of Radiation Oncology, Biology and Physics* 18: 887–893.

Leavitt, D. D., M. Martin, J. H. Moeller, and W. L. Lee (1990). Dynamic wedge field techniques through computer-controlled collimator motion and dose delivery. *Medical Physics* 17(1): 87–92.

Lim, J.-H., M. C. Ferris, S. J. Wright, D. M. Shepard, and M. A. Earl (2006). An optimization framework for conformal radiation treatment planning. *INFORMS Journal on Computing, forthcoming* .

Lodwick, W., S. McCourt, F. Newman, and S. Humphries (1998). Optimization methods for radiation therapy plans. In Borgers, C. and F. Natterer, editors, *Computational Radiology and Imaging: Therapy and Diagnosis*, IMA Series in Applied Mathematics, pp. 229–250. Springer-Verlag, Berlin, Germany.

Mackie, T. R., T. Holmes, S. Swerdloff, P. Reckwerdt, J. O. Deasy, J. Yang, B. Paliwal, and T. Kinsella (1993). Tomotherapy: a new

concept for the delivery of dynamic conformal radiotherapy. *Medical Physics* 20(6): 1709–1719.

Mohan, R. (1995). Field shaping for three-dimensional conformal radiation therapy and multileaf collimation. *Seminars in Radiation Oncology* 5(2): 86–99.

Olafsson, A. and S. J. Wright (2006). Linear programming formulations and algorithms for radiotherapy treatment planning. *Optimization Methods and Software* 21: 201–232.

Olivera, G. H., D. M. Shepard, K. J. Ruchala, J. S. Aldridge, J. M. Kapatoes, E. E. Fitchard, P. J. Reckwerdt, G. Fang, J. Balog, J. Zachman, and T. R. Mackie (1999). Tomotherapy. In Dyk, J. Van, editor, *Modern Technology of Radiation Oncology*, pp. 521–587. Medical Physics Publishing, Madison, Wisconsin.

Romeijn, H. E., R. K. Ahuja, J. F. Dempsey, and A. Kumar (2006). A new linear programming approach to radiation therapy treatment planning problems. *Operations Research* 54: 201–216.

Rosen, I., R. Lane, S. Morrill, and J. Belli (1990). Treatment plan optimization using linear programming. *Medical Physics* 18(2): 141–152.

Shepard, D. M., M. C. Ferris, G. Olivera, and T. R. Mackie (1999a). Optimizing the delivery of radiation to cancer patients. *SIAM Review* 41: 721–744.

Shepard, D. M., G. H. Olivera, L. Angelos, O. A. Sauer, P. J. Reckwerdt, and T. R. Mackie (1999b). A simple model for examining issues in radiotherapy optimization. *Medical Physics* 26.

Tervo, J. and P. Kolmonen (2000). A model for the control of a multileaf collimator in radiation therapy treatment planning. *Inverse Problems* 16: 1875–1895.

Webb, S. (1998). Configuration options for intensity-modulated radiation therapy using multiple static fields shaped by a multileaf collimator. *Physics in Medicine and Biology* 43: 241–260.

Wu, Xingen and Yuping Zhu (2001). A global optimization method for three-dimensional conformal radiotherapy treatment planning. *Physics in Medicine and Biology* 46: 109–119.

Xiao, Y., Y. Censor, D. Michalski, and J. M. Galvin (2003). The least-intensity feasible solution for aperture-based inverse planning in radiation therapy. *Annals of Operations Research* 119: 183–203.

Yu, C. X. (1995). Intensity-modulated arc therapy with dynamic multi-leaf collimation: an alternative to tomotherapy. *Physics in Medicine and Biology* 40(9): 1435–1449.

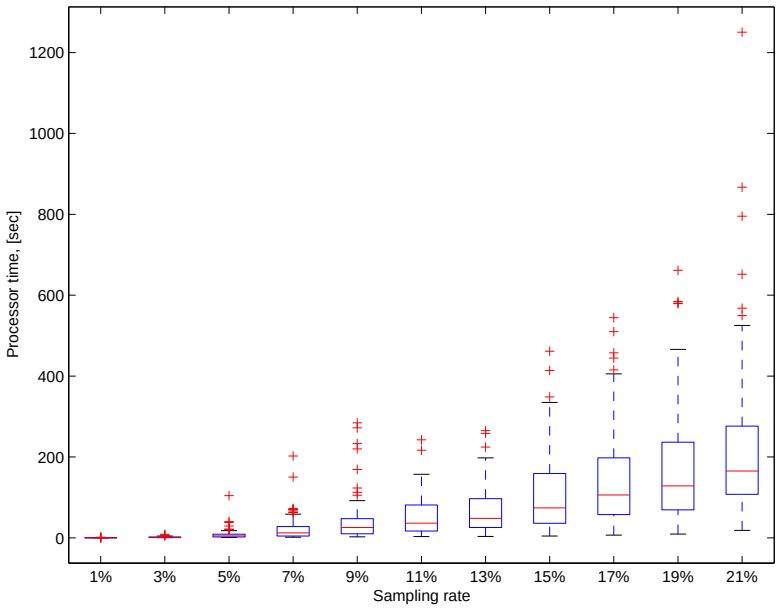

Figure 1: Pelvis example: solution times for various sample rates. The boxes represent the 25th and 75th percentiles, with the median value indicated within each box.

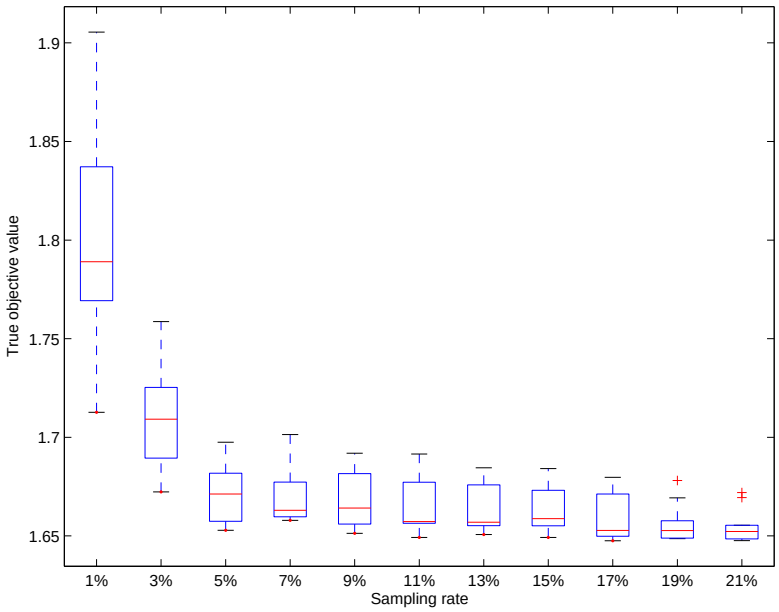

Figure 2: Pelvis example: objective values for various sample rates

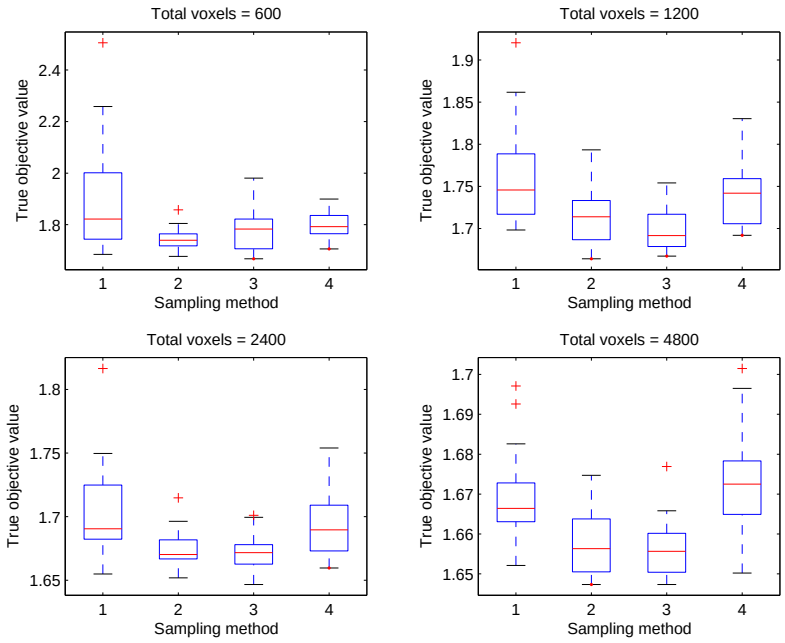

Figure 3: Boxplot of different sampling methods for the pelvis case with 20 replications. Method 1 has equal sampling in each structure, method 2 uses sampling proportional to structure size except normal tissue sample is at most twice as large as any other sample, method 3 is similar to method 2 except factor is 5, method 4 is similar to method 2 except factor is 20

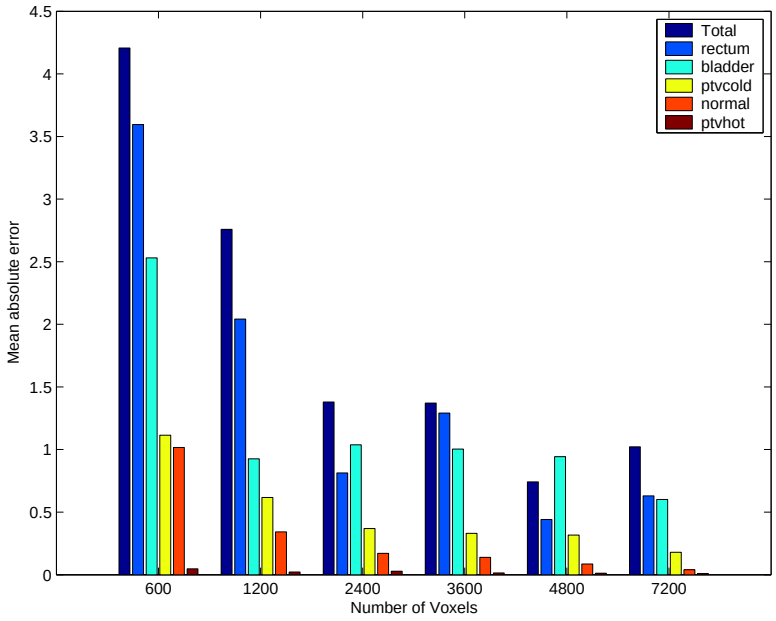

Figure 4: Errors in objective components for the pelvis case, 10 replications using proportional sampling

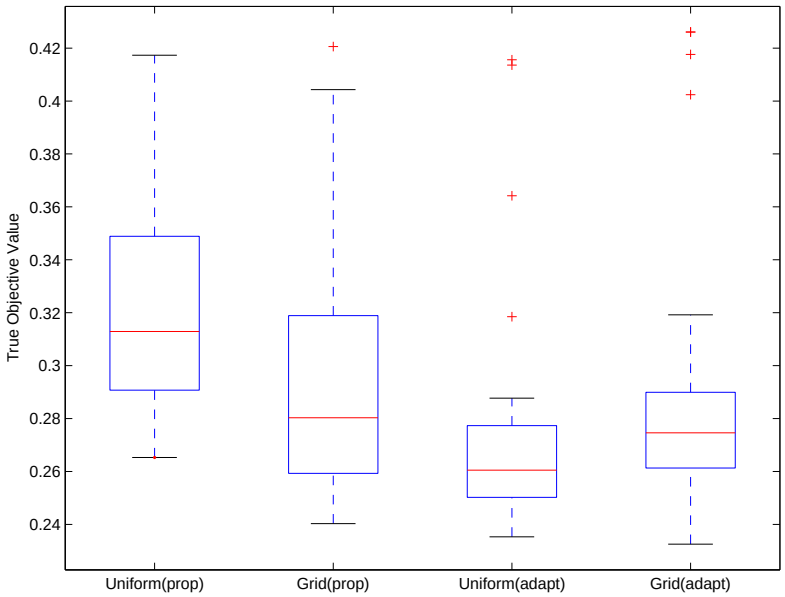

Figure 5: Boxplot of objective values from MIP solution with various types of intra-organ sampling, solved to optimality using 30 replications and 600 total voxels.

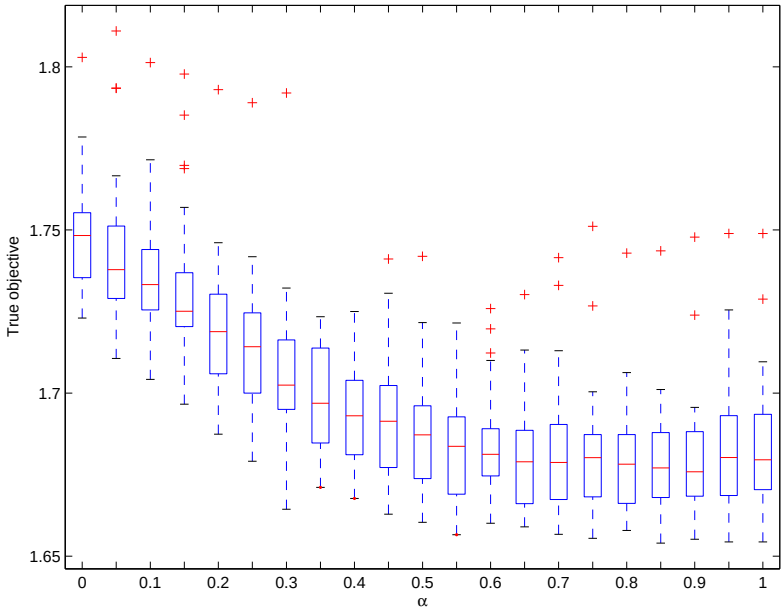

Figure 6: Boxplot of different proportions of sampling in and outside rind of target, total number of voxels 2400, α is fraction of voxels in the rind of the target, 30 replications

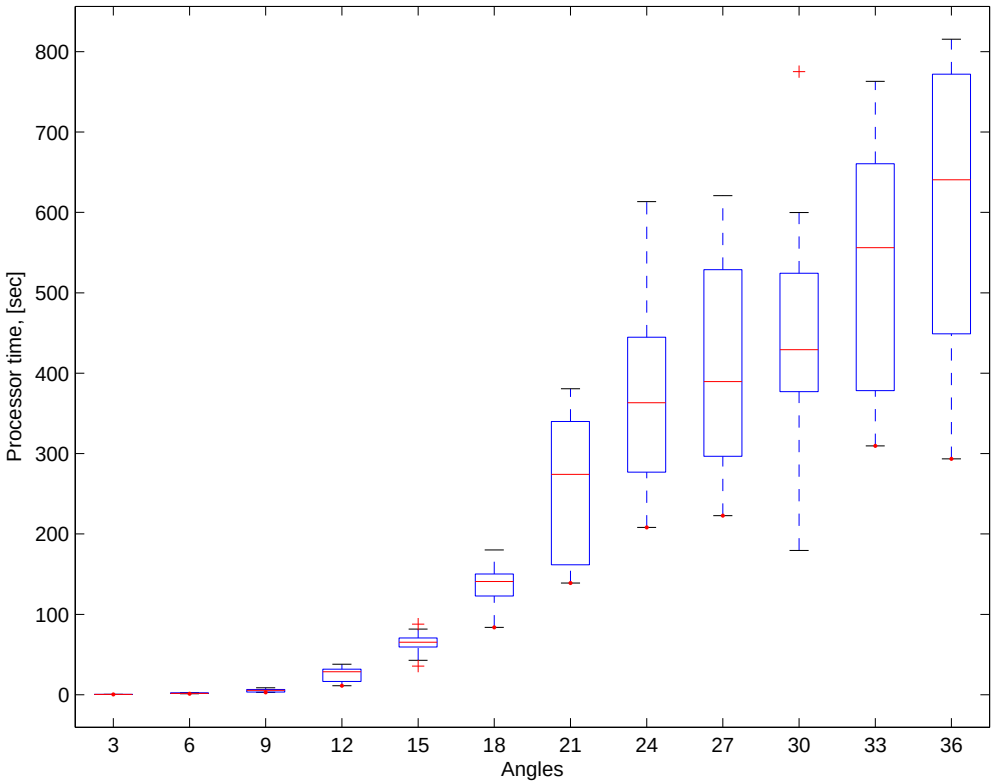

Figure 7: Solution times (to optimality) for MIP choosing 3 angles for pancreas case with 10 replications at sample size 3600 with normal tissue sample the same size as largest other structure sample

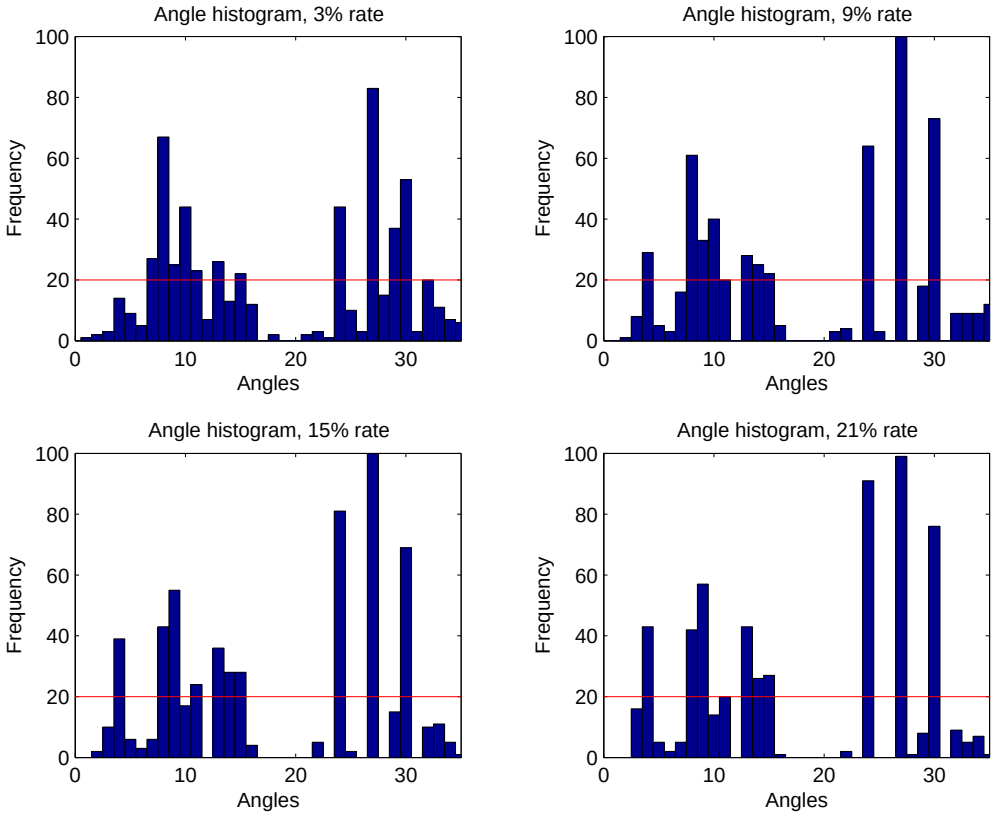

Figure 8: Angle selection using frequency counts

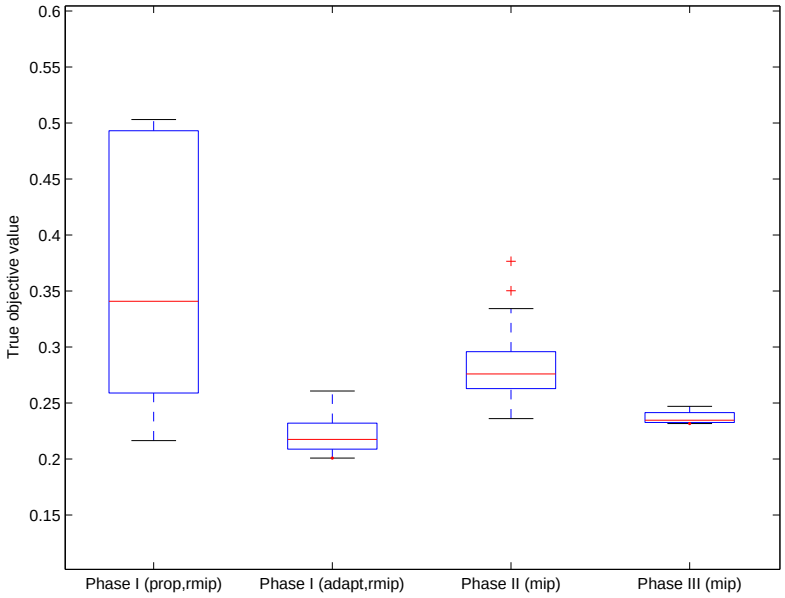

Figure 9: Box plot of true objective values thoughout the multi-phase sampling method in the pancreas case, with 30 replications.

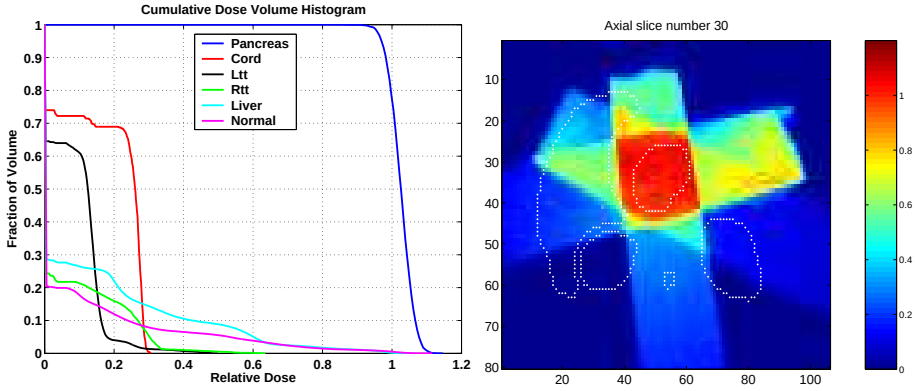

Figure 10: Cumulative dose volume histogram and dose plot of pancreas case solution

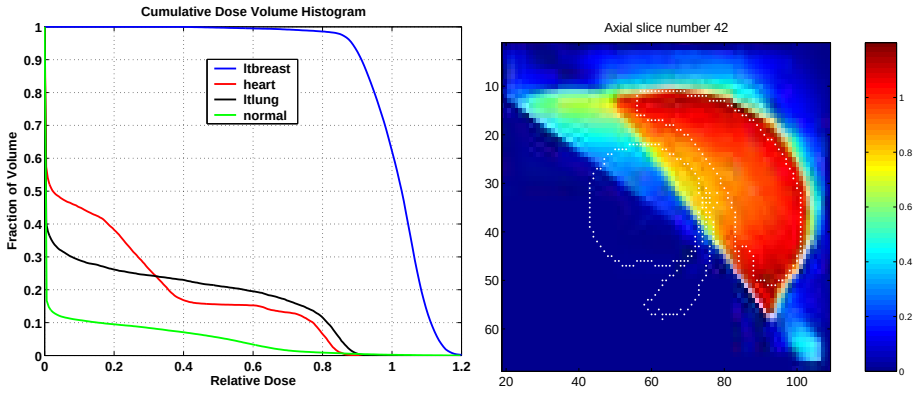

Figure 11: Cumulative dose volume histogram and dose plot of breast case solution

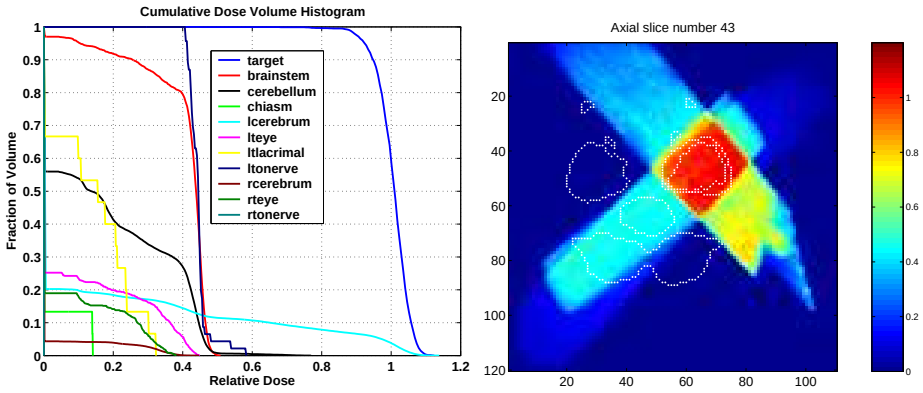

Figure 12: Cumulative dose volume histogram and dose plot of head/neck case solution