Reduced Order Constrained Optimization (ROCO): Clinical Application to Lung IMRT

Hans Stabenau,[1,*] Linda Rivera,[2,†] Ellen Yorke,[1,‡] Jie Yang,[1,§]
Renzhi Lu,[2,¶] Richard Radke,[2,**] and Andrew Jackson[1,††]

[1] *Memorial Sloan-Kettering Cancer Center*

[2] *Rensselaer Polytechnic Institute*

Abstract

Purpose: We use reduced-order constrained optimization (ROCO) to create clinically acceptable IMRT plans quickly and automatically for advanced lung cancer patients. Our new ROCO implementation works with the treatment planning system and full dose calculation used at Memorial Sloan-Kettering Cancer Center, and we have implemented mean dose hard-constraints, along with the point-dose and dose-volume constraints that we used for our previous work on the prostate.

Methods: ROCO consists of three major steps. First, we sample the space of treatment plans by solving a series of optimization problems using penalty-based quadratic objective functions. Next, we find an efficient basis for this space via principal component analysis (PCA); this reduces the dimensionality of the problem. Finally, we solve a constrained optimization problem over this basis to find a clinically acceptable IMRT plan. Dimensionality reduction makes constrained optimization computationally efficient.

Results: We apply ROCO to 12 stage III non-small-cell lung cancer (NSCLC) cases, generating IMRT plans that meet all clinical constraints and are clinically acceptable, and demonstrate that they are competitive with the clinical treatment plans. We also test how many samples and PCA modes are necessary to achieve an adequate lung plan, demonstrate the importance of long range dose calculation for ROCO, and evaluate the performance of non-specific normal tissue ("rind") constraints in ROCO treatment planning for the lung. Finally, we show that ROCO can save time for planners; we estimate that, in our clinic, planning using our approach would save a median of 105 minutes for the patients in our study.

Conclusions: New challenges arise when applying ROCO to the lung site, which include the lack of a class solution, a larger treatment site, an increased number of parameters and beamlets, a variable number of beams and beam arrangement, and the customary use of rinds in clinical plans to avoid high-dose areas outside the PTV. In our previous work, use of an approximate dose calculation in the hard constraint optimization sometimes meant that clinical constraints were not met when evaluated with the full dose calculation. This difficulty has been removed in the current work by using the full dose calculation in the hard constraint optimization. We have demonstrated that ROCO offers a fast and automatic way to create IMRT plans for advanced NSCLC, which extends our previous application of ROCO to prostate cancer IMRT planning.

7 Keywords: optimization, IMRT, constrained optimization, dimensionality reduction

I. INTRODUCTION

Intensity-modulated radiotherapy (IMRT) has revolutionized the treatment of cancers in the last decade: it allows a higher dose to be delivered to a tumor while protecting nearby radiation-sensitive normal tissues, yielding better local control and fewer post-treatment complications than previous techniques[1–3] However, the process of obtaining a clinically acceptable IMRT plan for a difficult treatment site is often slow and labor-intensive, requiring hours of expert time in a manual trial-and-error loop in which the parameters of the optimization score function are repeatedly adjusted. Long planning times place a severe stress on available resources in a busy clinic, and can result in treatment delays, acceptance of sub-optimal plans or, in the worst case, errors due to time pressure. In this paper, we apply a method called reduced-order constrained optimization (ROCO) to greatly reduce the amount of time required to obtain a clinically acceptable IMRT plan. By minimizing the trial-and-error effort characteristic of current IMRT planning, it allows treatment planners to focus on clinical tradeoffs between tumor coverage and normal organ doses. We have previously applied ROCO to prostate cancer cases[4]; in this paper, we improve our application of ROCO and report new results on a more challenging treatment site, the lung.

Lung cancer accounts for the most cancer-related deaths in both men and women in the United States. An estimated 157,300 deaths, accounting for about 28% of all cancer deaths, are expected to occur in 2010[5]. Radiation therapy is the main curative treatment for inoperable non-small cell lung cancer (NSCLC), but it remains a technically challenging procedure with very low 5-year survival rates ($< 10\%$)[6]. IMRT is promising for treatment of NSCLC compared to traditional radiotherapy or 3D-CRT since it may enable dose escalation to the tumor[7]; however, the organs at risk (OARs) are sensitive to radiation, including the lungs, esophagus, and spinal cord. Since the sizes and locations of lung cancers are diverse, unlike prostate cancer, a standard multi-field class solution for IMRT is not used. Typical treatment plans for locally advanced (stage III) lung cancer feature prescription doses of 1.8–2 Gy/fraction delivered by 3–5 coplanar treatment beams of 6 MV photons, occasionally with the addition of non-coplanar beams. For the locally advanced NSCLC cases we examine in this paper, we estimate that it takes an expert planner around 3 hours to create a clinically acceptable IMRT plan (not counting time spent contouring structures and selecting beam directions).

In this paper, we describe our implementation of ROCO, which we have integrated with the clinical treatment planning system at Memorial Sloan-Kettering Cancer Center (MSKCC), and our results from retrospective application of ROCO to 12 locally-advanced lung cancer cases. The anonymized clinical data (image sets, structure contours, and clinical treatment plans) for these patients were provided by MSKCC under IRB approval. ROCO consists of three main steps, after beam directions have been selected. First, random sets of score function parameters are chosen via latin hypercube sampling, and these plans are optimized using the clinical score-function-based optimization. Second, principal component analysis (PCA) isolates the important modes of variation in the intensity matrices, which shifts the independent variables of the problem to the few dominant PCA modes. Sampling and PCA modes are generated for each patient individually, not as class solutions. The third step is hard-constrained optimization. Dimensionality reduction by PCA makes it feasible to rapidly and automatically locate plans with clinically acceptable PTV coverage and normal tissue protection in the space spanned by the sampled plans. Using the MSKCC planning system, the overall process takes approximately 30 minutes per patient on a modest desktop workstation (an Intel Core 2 Duo, clock speed 2.33 GHz, with 3.5 GB of RAM).

Advanced lung cancer cases present new challenges when compared to our previous work on the prostate. For prostate cases, because the relationship between the PTV and the rest of the anatomy varies relatively little from patient to patient, the same beam directions were used for each patient. Stage III lung tumors, on the other hand, show extremely variable geometries and can grow to considerable size, growing outside of the lung proper and into the mediastinum; additionally, single or multiple tumors can appear in a variety of geometries near OARs such as the heart, esophagus, spinal cord, and brachial plexus. Because of this, ROCO used the clinical beam directions chosen by the planner in each case. Our current implementation is integrated with the clinical MSKCC treatment planning system in order to make it flexible enough to deal with different treatment sites besides the prostate, whereas the software previously described used data exported from, and performed calculations outside of, the treatment planning system[4] (which caused difficulty because of discrepancies in dose calculation).

Fig. 1 shows a single CT image slice of a representative lung cancer patient in our study, with contours for the different OARs, together with a 3D representation of the CT images showing the tumor wrapping around the esophagus. The dimension of the space of possible

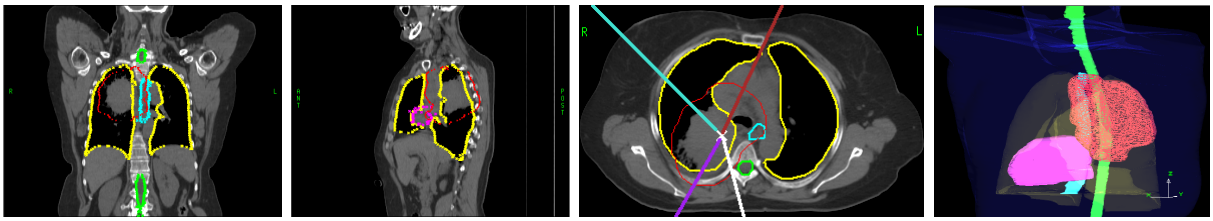

FIG. 1. The left three panels show CT image slices in the treatment plane for patient #8 in our study; the rightmost panel shows a beam's eye view of the same case. The PTV is shown in red, the lungs in yellow, the spinal cord in green, the heart in pink, and the esophagus in cyan. The solid lines in the third panel show the beam directions.

treatments is larger for these locally-advanced lung cases than for prostate cases, because the larger treatment fields contain a greater total number of beamlets. For prostate cases there are on the order of 10^3 beamlets, and for the lung cases that we consider, there are about 10^4. Finally, IMRT for NSCLC often includes "rind" structures to prevent hot spots in non-specific normal tissues. Table I summarizes the major differences pertaining to treatment planning between the prostate cases we had considered previously and the stage III NSCLC cases considered in this paper.

The standard clinical approach to inverse IMRT planning is to combine all the evaluation criteria specified by the physician into a scalar value using a weighted sum of several terms (i.e., costlets[8]). Each term includes a dose parameter (i.e., a minimum or maximum limit) or a pair of dose-volume parameters (i.e., a point on a DVH curve), and reflects a clinical objective. The weight of each term is the relative penalty imposed by the planner for not satisfying each objective. Such a formulation is easy to implement and can be optimized quickly using gradient information, e.g., by Newton's methods[9] or conjugate gradient algorithms[10]. Because the result of a penalty-based optimization is not guaranteed to satisfy the clinical criteria, we refer to such an optimization scheme as an "unconstrained optimization".

In practice, unconstrained optimizations require a great deal of heuristic trial and error to arrive at parameter settings such that the resulting plan is clinically acceptable[11]. The planner uses the weights (or "importance factors") in the objective function to try to "steer" the optimization algorithm to more clinically desirable solutions[12], but this can be difficult since the process of adjusting these weights is inherently imprecise and unintuitive[13]. The role of dose limits in IMRT optimization is also confusing, since it has been observed that in

Criterion	Prostate Case	Lung Case
Beams (geometry)	5 (class solution)	4–9
Beamlets	$\sim 10^3$	$\sim 10^4$
Median PTV volume	~ 160 cm^3	~ 380 cm^3
PTV/OAR relationships	Similar	Variable
Non-specific normal tissue sparing	Beam arrangement	"Rind" structures
Optimization parameters	~ 30	~ 50
OARs	3–5	5–10

TABLE I. Comparison of IMRT treatment planning complexity in prostate and lung treatments. Hot spots in non-specific normal tissues around the prostate are avoided by beam arrangements and small PTV size, so rinds are not usually required.

an unconstrained optimization, dose limits more stringent than the clinical limits are required to obtain convergence to an acceptable plan (see, e.g.,[10,14,15]). The inverse planning process of obtaining a clinically acceptable IMRT plan for a difficult site can take several hours, largely due to the manual process of adjusting the parameters in the objective function[10,13,16].

In our previous work[17], we applied sensitivity analysis to identify key parameters of an unconstrained IMRT objective function that have a strong impact on the resultant dose distribution. We then applied an outer loop over the sensitive parameter set to find the parameters such that the minimizer of the corresponding objective function gave the best score of a scalar function of plan quality. While this method quickly produced plans that generally satisfied the clinical constraints, it still suffered from (1) using a scalar-valued objective function to approximate a fundamentally hard-constrained problem, and (2) requiring training data to identify the sensitive set, assuming a generalizable class solution for the treatment site. The ROCO algorithm has neither shortcoming.

While hard-constrained optimization for IMRT planning has been proposed previously (e.g., using mixed-integer programming[18]), it is typically prohibitively time-consuming due to the huge dimensionality of the problem and the difficulty in implementing dose-volume constraints. Another recent focus of interest is multiobjective (MO) optimization, which

allows the planner to choose from a family of Pareto-optimal plans (that is, plans in which no criterion can be improved without worsening the others)[19,20].

The ROCO algorithm makes constrained optimization computationally tractable using four steps:

1. Select the targets and OARs to be included in the score function, and choose the beams whose intensities are to be optimized.

2. Randomly sample sets of score function parameters, apply the clinical optimization to each set, and store the resulting intensity patterns.

3. Apply principal component analysis (PCA) to this set of intensity profiles. The resulting principal components form a basis for the space of plans that contains the optimal plan.

4. Compute the coefficients of the basis vectors that optimize target coverage, subject to clinical constraints.

In the following section, we briefly review each of these steps, placing emphasis on the new features we have added; a more complete treatment is given in our previous paper[4].

II. SUBJECTS AND METHODS

At MSKCC, stage III NSCLC IMRT plans are delivered at 2 Gy/fraction with the sliding window technique. Up to 7 planner-chosen 6 MV beam directions concentrated on the ipsilateral side to geometrically protect the contra-lateral lung are used (in our set of patients, up to 6 beam directions are used). The ROCO algorithm used these beam directions to retrospectively re-plan 12 NSCLC patients who had already been treated with IMRT; PTV volumes ranged from 194 to 820 cm^3 (median 383 cm^3), with some patients having two PTVs (the tumor and nodal metastases). These patients were selected to have challenging clinical scenarios, i.e., large tumors with mediastinal extent, where the treatment planner had required from 10-50 optimization cycles to come up with an acceptable plan.

Structure	PTV	Lungs	Esophagus	Spinal cord	Brachial plexus
Constraint	$D_{\max} < 110\%$	$D_{\text{mean}} < 20$ Gy	$D_{\text{mean}} < 34$ Gy	$D_{\max} < 50$ Gy	$D_{\max} < 65$ Gy, $D_{05} < 60$ Gy

TABLE II. Clinical organ constraints to be implemented by ROCO for lung plans. The lung mean dose constraint is a proxy for NTCP $< 25\%$ (see text); the esophagus constraint is not enforced clinically by the planner if it cannot be met without compromising coverage. Not shown in the table are nonspecific normal tissue maximum dose constraints: major hot spots ($< 110\%$) outside the PTV are not tolerated.

A. Treatment plan criteria

The current MSKCC clinical evaluation protocol requires that the plan for IMRT treatment of primary lung tumors to 50–80 Gy for the PTV satisfies the conditions in Table II. The mean dose constraint on the paired lungs usually ensures that the Lyman-Kutcher-Burman lung NTCP[21,22] is $\leq 25\%$. The hard constraint on the esophagus D_{mean} is only used clinically by the planner if it can be met without compromising target coverage. Target $D_{\min}$ is not included as a hard constraint on the clinical plans; if D_{95} and V_{95} are $\sim 95\%$ or better, we deem coverage sufficient.

The dose to non-specific normal tissue surrounding the PTV is also of concern: "hot spots" above 100% of prescription outside the PTV are discouraged in clinical plans, while those above 110% are not tolerated. If the dose distribution is insufficiently conformal more than ~ 0.5 cm beyond the PTV, then the plan will be rejected by the treatment planner. Excessive modulation of the intensity profiles, which can lead to delivery problems and unnecessarily increased delivery time, is also not permitted in the clinic.

A difficulty in creating a treatment plan is that the definition of "clinically acceptable" can change depending upon the specific situation under consideration. Certain dose constraints are inflexible (e.g., in our clinic, the spinal cord maximum dose is never permitted to go above 50 Gy). Other constraints, however, such as restrictions on non-specific normal tissue maximum dose ("hot spot" constraints), or mean dose constraint to the esophagus, may be relaxed if the physician is unhappy with the tradeoffs in the plan and desires to improve the coverage of the PTV.

B. Unconstrained optimization

For every patient, we sample the solution space by varying the parameters of a quadratic dose-based objective function, and subjecting it to the unconstrained optimization that has been used for many years in clinical practice at MSKCC[6,10]. This optimization is referred to as "unconstrained" because while the objective function parameters influence the doses to the various structures, an intensity distribution that minimizes such an objective function is not guaranteed to obey any particular constraint.

For the k^{th} target, the corresponding objective function term is:

$$\begin{aligned} F_k^{\text{target}} =& \frac{1}{N_k}\left(\sum_{i=1}^{N_k}(D_i - D_k^{\text{Rx}})^2\right. \\ &+ w_k^{\min}\sum_{\{i|D_i<D_k^{\min}\}}(D_i - D_k^{\min})^2 \\ &\left.+\ w_k^{\max}\sum_{\{i|D_i>D_k^{\max}\}}(D_i - D_k^{\max})^2\right), \end{aligned} \tag{1}$$

where N_k is the number of points in the target, D_i is the dose to the i^{th} point in the target, D_k^{Rx} is the prescription dose, $D_k^{\min}$ and $D_k^{\max}$ are the minimum and maximum dose allowed without penalty, and $w_k^{\min}$ and $w_k^{\max}$ are the penalties (weights) for under- and over-dosing. The parameter set $P_k = \{D_k^{\text{Rx}}, D_k^{\min}, D_k^{\max}, w_k^{\min}, w_k^{\max}\}$ completely specifies the objective function for target k. A similar objective function term is defined for each OAR and rind structure (see Sec. II A and Table II), which also includes parameters D_k^{dv}, D_k^{mean}, w_k^{mean}, and w_k^{dv}, that define the dose-volume-histogram (DVH) and mean dose constraints:

$$\begin{aligned} F_k^{\text{OAR}} =& \frac{1}{N_k}\left(w_k^{\max}\sum_{\{i|D_i>D_k^{\max}\}}^{N_k}(D_i - D_k^{\max})^2\right. \\ &+ w_k^{\text{dv}}\sum_{i=1}^{N_k^{\text{dv}}}(D_i - D_k^{\text{dv}})^2 \\ &\left.+\ w_k^{\text{mean}}N_k(\bar{D}_k - D_k^{\text{mean}})^2\Theta(\bar{D}_k - D_k^{\text{mean}})\right). \end{aligned} \tag{2}$$

The sum in the second term is carried out over the lowest N_k^{dv} doses that are greater than D_k^{dv}, and N_k^{dv} is the minimum number of point dose changes required to bring the k^{th} organ into compliance with the DVH constraint[10]. Mean dose to the k^{th} organ is denoted by $\bar{D}_k$;

the heaviside step function Θ ensures that this term only contributes to the score function when $\bar{D}_k > D_k^{\text{mean}}$.

C. Sampling

Let I_{opt} be an intensity distribution that optimizes PTV coverage while obeying clinical constraints; further suppose that some unknown set of score function parameters P_{opt} cause this plan to be generated by unconstrained optimization. Then we conjecture that if we randomly choose parameter sets P_q in the neighborhood of P_{opt}, the resulting I_q from unconstrained optimization will define a small basis which spans a space containing such an I_{opt}[23].

We choose this neighborhood from clinical experience to include the range of values that planners have used for similar cases. Once a range of parameters has been chosen, Latin hypercube sampling is used to choose N_{samp} parameter sets at which to sample; Latin hypercube sampling is a particular case of stratified sampling that achieves an efficient coverage of the space of input parameters[24].

D. Dimensionality reduction and dose calculation

Given N_{samp} optimized intensity distributions $\{I_1, I_2, \ldots, I_{N_{\text{samp}}}\}$ resulting from the unconstrained optimization using score function parameter sets $\{P_1, P_2, \ldots, P_{N_{\text{samp}}}\}$, the dimensionality of the intensity space can be reduced by linear or nonlinear feature extraction methods. Here, we use Principal Component Analysis (PCA)[25] for the reduced-order approximation. PCA is an orthogonal linear transformation that maps the data to a new coordinate system, such that the dimension with the k^{th} greatest variance is oriented to lie on the k^{th} coordinate (i.e., the k^{th} principal component). This procedure shifts the independent variables of the problem from the approximately 10^4 beamlets that specify the intensity profile of a treatment plan to the N_{modes} PCA modes with the greatest variance. These modes U_k span a reduced solution space.

During unconstrained optimization with conjugate-gradient methods, the MSKCC treatment planning system uses an approximate, short-range kernel for the purposes of calculating the doses to the targets and OARs, so that many evaluations of the dose calculation can be

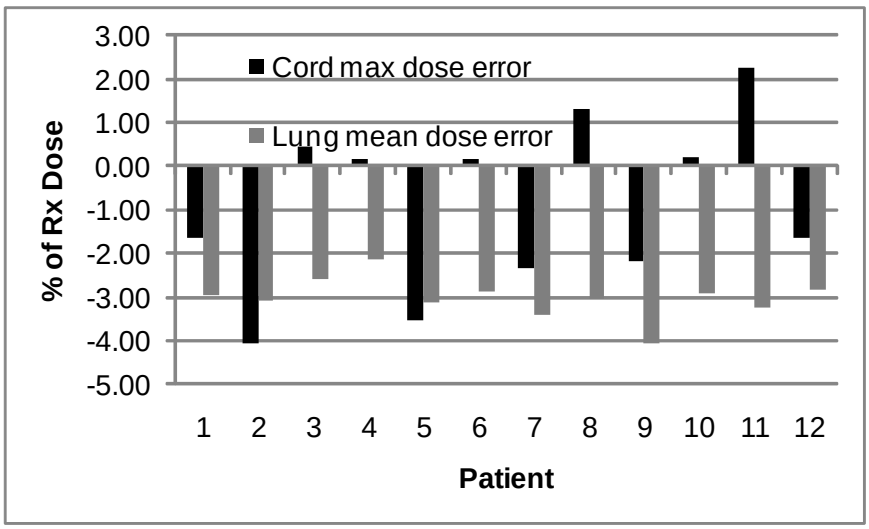

FIG. 2. Importance of long-range dose calculation. The figure shows $\Delta D = (D_{\text{short range}} - D_{\text{full dose}})/D_{\text{Rx}}$ for the spinal cord $D_{\max}$ (black) and mean lung dose (gray) for each patient using the ROCO intensities. In order to get accurate results from ROCO, the dose calculation for each PCA mode used by the constrained optimization must be performed with the full long-range dose kernel.

executed rapidly[10]. Once optimization has completed, a long-range full dose calculation is performed, and then the plan is evaluated based on this calculation.

We have found that, while this approach is sufficient for the sampling step, it is inaqequate for the subsequent steps of ROCO. After the U_k are determined, it is critical to make the dose calculation for the PCA modes as accurate as possible, so that during the constrained optimization, the solver has accurate information about OAR doses and target coverage. Fig. 2 shows that lung mean doses are systematically underestimated by up to 5% when using the short-range approximate dose calculation normally used during the clinical score-function-based optimization. This would result in a systematic overdosing of these tissues in ROCO plans, which was a limitation in our previous work[4]. We have addressed this issue here by using the long-range dose calculation to evaluate the dose distributions for the PCA modes, which costs some time: finding the optimal basis requires less than a minute to complete, but calculating the full dose distributions corresponding to the modes requires another 5–10 minutes. Nevertheless, this constitutes a major improvement in our implementation of ROCO, and makes ROCO optimization suitable for large targets.

E. Constrained optimization

Given the reduced-dimension space that captures the effective degrees of freedom in the intensity variables, our final task is to find a clinically acceptable solution in terms of the reduced basis. For this step, the optimizer has N_{modes} degrees of freedom: the coefficients of the PCA modes. The goal of the optimization is specified as

$$\min \sum_{i \in T} (D_i - D^{\text{Rx}})^2, \tag{3}$$

for the voxels T in the target structures. This causes the optimizer to work toward uniform PTV coverage. The doses to voxel i are given by

$$D_i = \sum_{k=1}^{N_{\text{modes}}} V_{ik} \xi_k + v_i. \tag{4}$$

In this equation, the ξ_k are the coefficients of the principal components, which are the independent variables of the optimization. V_{ik} is the dose to voxel i from principal component k, and v_i is the dose to this voxel from the mean of the samples. The intensities of these modes were determined during the dimensionality reduction step, and the V_{ik} and v_i are obtained by calculating the doses for each intensity mode (the U_k from Sec. II D) and for the mean.

For each organ, the point dose hard constraints are specified by

$$D_i \leq D^{\max} \tag{5}$$

where i runs over the set of voxels in each organ or target. There is no $D^{\min}$ constraint present, because while targets are specified with a $D^{\max}$, dose homogeneity is included as an optimization goal in Eq. (3) above instead of as an explicit $D^{\min}$ constraint. Mean dose constraints are given by

$$\sum_{i=1}^{N_{\text{vox}}} D_i \leq N_{\text{vox}} D^{\text{mean}}, \tag{6}$$

where N_{vox} is the number of voxels in the structure. DVH constraints are implemented using an iterative scheme[4]; briefly, on the first iteration of optimization, no DVH constraints are implemented. If a DVH constraint of the form $D_V \leq Y$ for a dose Y and volume fraction V is then found to be violated after optimization, we apply the constraint $D_i \leq Y$, where i consists of the hottest $N_{\text{vox}} V$ voxels, and repeat the optimization, applying DVH constraints as needed at each step.

There is also an additional constraint, which is that

$$\sum_{k=1}^{N_{\text{modes}}} U_{jk}\xi_k + \mu_j \geq 0, \tag{7}$$

where U_{jk} is the value of beamlet j in mode k, and μ_j is the value of beamlet j in the mean of the samples. This ensures that the set of ξ_k in the solution results in a non-negative intensity distribution.

The dimensionality reduction by PCA makes it feasible to use a quadratic programming solver (ILOG CPLEX) to solve this hard-constrained problem. This step took 1-10 minutes when using 25 PCA modes, and we have found that the calculation time required is approximately linear for up to 200 degrees of freedom. Total time required was therefore approximately 30 minutes per patient; if it is desired to adjust the hard constraints, only the last step needs to be repeated. At the end of the process, the ROCO-optimized plans were leaf-sequenced for clinical delivery and the final clinical dose calculation was performed.

III. RESULTS

A. N_{samp} and N_{modes}

In order to help determine the optimal value for N_{samp}, in Fig. 3 we studied how coverage for the final ROCO plan varies with number of samples for all the patients in our study. These results show that 50 samples are sufficient to achieve the desired 95% PTV coverage, and that a larger number of samples is not likely to result in much improvement. Obtaining the intensity profiles of the 50 samples requires 10-15 minutes of computer time.

We also studied the characterstics of the solutions from ROCO as we vary N_{modes} in two ways. First, we examined how much of the variance in the samples was recovered using the PCA decomposition. The top panel of Fig. 4 shows the fraction of the variance recovered for each patient as a function of the number of modes used, using an original set of 500 samples. From this we determined that 25 modes was sufficient to recover 98% of the variance of the samples in all cases. The bottom panel shows the PTV coverage that we achieved as a function of the number of modes used in the plan. For $N_{\text{modes}} > 25$ we observed only a few % increase in PTV coverage for these patients. While we fixed 25 modes per patient for this study, we note that for some patients, excellent performance was achieved by ROCO

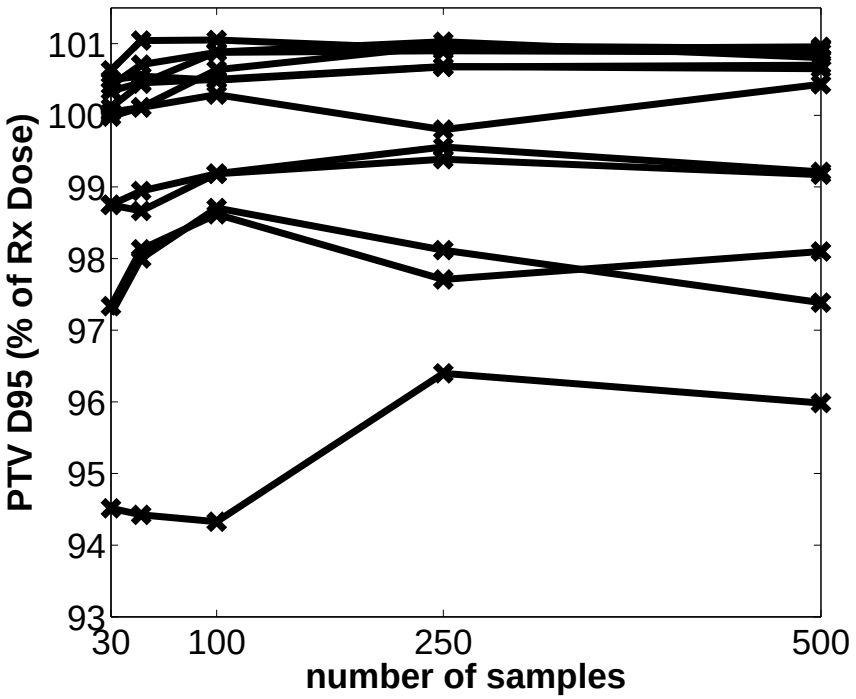

FIG. 3. PTV coverage (solid lines) achieved by ROCO vs number of samples for each patient in our study. After about 50 samples, almost no benefit is seen as the number of samples used is increased. No renormalization of the plans (as in Fig. 6) is performed.

for several patients using a surprisngly small number of modes. For example, patient 6 only required 2 modes to obtain a plan that was close to acceptable.

Next, for each patient, we took the intensity vector for the clinical plans used by the treatment planners $I_{\rm cl}$, and projected it into the reduced-dimension space, which allowed us to measure the projection residual R:

$$R = \frac{\|I_{\rm cl} - {\rm proj}_{U_k}(I_{\rm cl})\|_2}{\|I_{\rm cl}\|_2} \tag{8}$$

The top panel of Fig. 5 shows how R behaves as a function of $N_{\rm modes}$. In this plot, we see that there is initially a decrease in R as we increase the number of modes. Less improvement is seen after 10–25 modes. We chose to use 25 PCA modes for the patients in our study. Similarly, the bottom panel of Fig. 5 shows how R behaves as we vary $N_{\rm samp}$; around $N_{\rm samp} = 50$, the behavior of R is smooth, so we used $N_{\rm samp} = 50$ in our subsequent experiments.

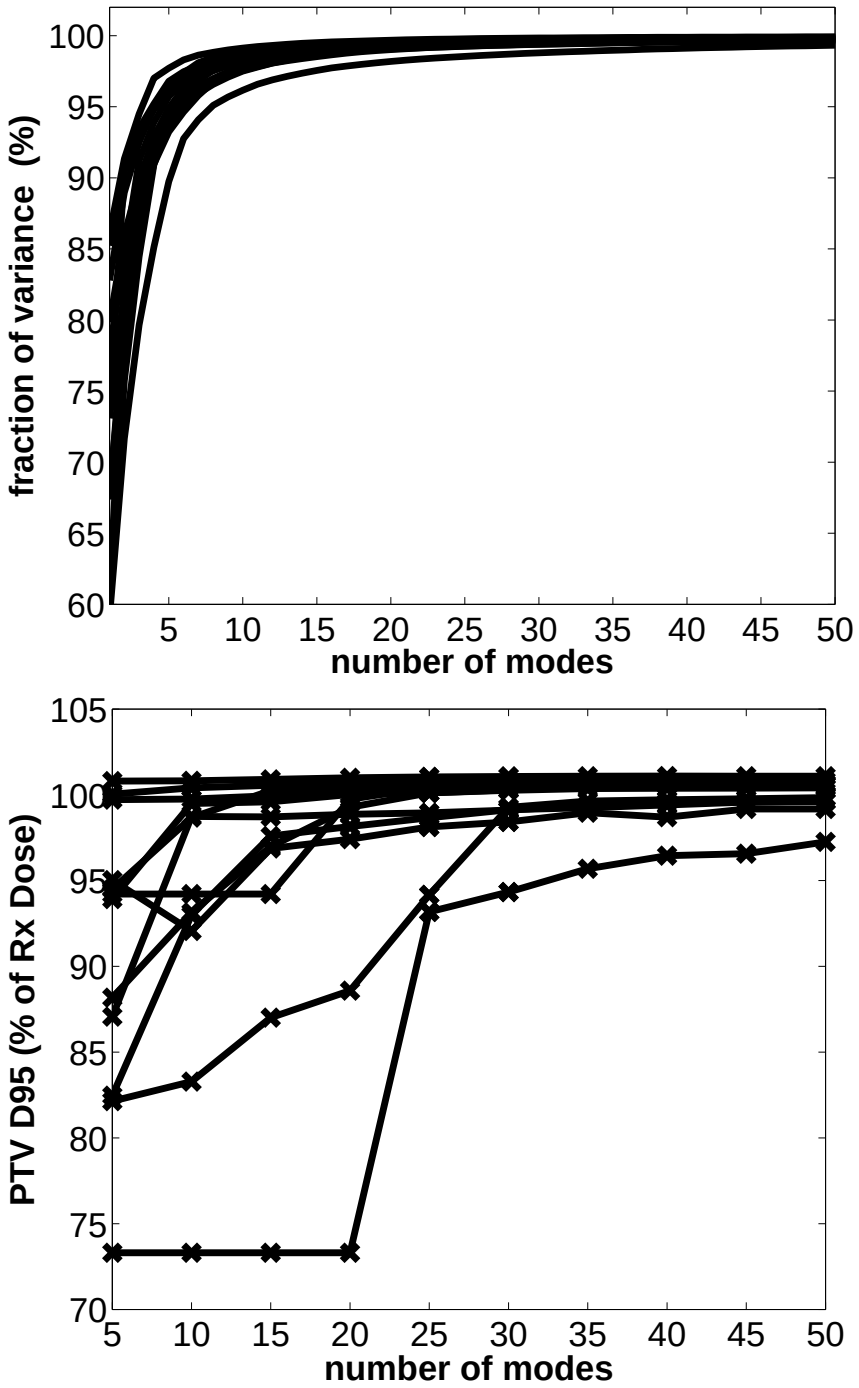

FIG. 4. Top panel: Cumulative fraction of variance in intensity profiles of sampled plans recovered, plotted against N_{modes} used. 25 modes are sufficient to capture 98% of the variance of the samples for all the patients in our study. Bottom panel: target coverage vs N_{modes} used; for these patients, $N_{\text{modes}} = 25$ was sufficient to achieve adequate coverage.

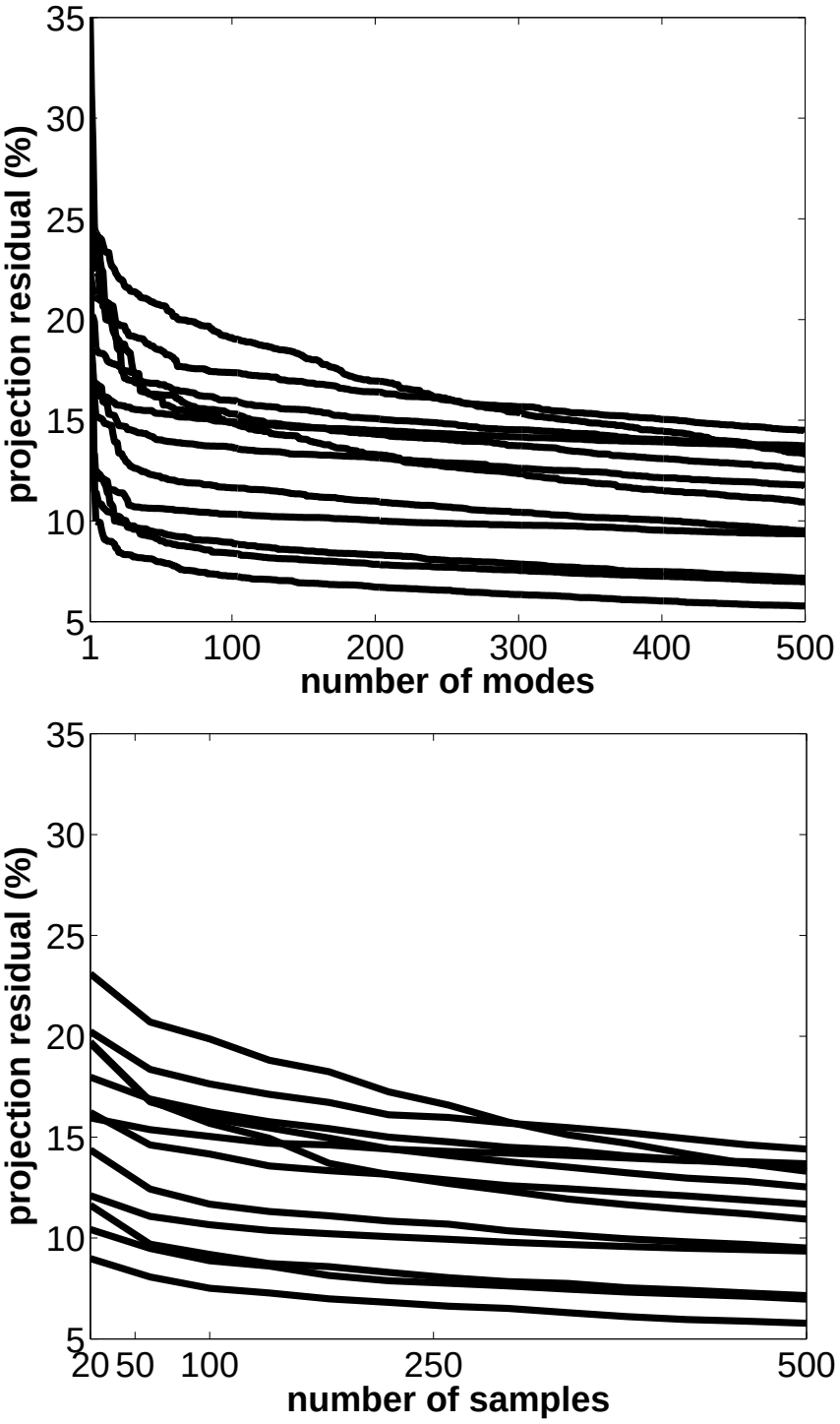

FIG. 5. Projection residual R (see Eq. (8)) plotted against the number of modes used (top panel; using 500 samples) and number of samples used (bottom panel; using $N_{\text{modes}} = N_{\text{samples}}$).

B. Clinical plan comparison

In this section, we compare the ROCO plans to the plans that were used to treat these 12 patients in the clinic. In order to prepare for this comparison, the ROCO plans were evaluated by an experienced treatment planner at MSKCC to ensure clinical acceptability, i.e., compliance with the criteria mentioned in Sec. II A. For the purposes of this comparison, the ROCO plan was normalized to have the same D_{95} value as the plan used for treatment. All ROCO plans were inspected to confirm that the intensity profiles were sufficiently smooth, i.e., there were no large peaks in the intensity profiles, and that the plans did not require an excessive number of monitor units (MU) to deliver (the ROCO and clinical plan MU were similar).

In all cases, ROCO achieved a plan satisfying the given input constraints, which is the primary goal of using hard-constrained optimization. However, after the plan was normalized to have the same D_{95} as the treatment planner's plan, it was not unusual for the constraints to be violated; for example, if the D_{95} achieved by the ROCO plan was smaller than the treatment planner's D_{95}, then after normalization, the lung mean dose constraint might be violated. This is not a failure of ROCO but rather a consequence of the difficulty of directly comparing two plans, a difficult and well-known problem[26–29]. As a result, it was sometimes necessary to reoptimize patients using lower organ dose constraints or a lower PTV max dose constraint. When ROCO is used as a standalone planning tool, without the intention of comparing to a reference plan, this step is not necessary.

The plots in Fig. 6 show that for each case in our study, ROCO plans are competitive with the treatment planner's plans. The required clinical constraints for the spinal cord maximum dose and lung mean dose are satisfied in each case; the esophagus mean dose constraint was satisfied when it was clinically possible to do so without sacrificing coverage. In Table III, we summarize ROCO's performance with a figure of merit D^*: for the PTV $D_{\min}$, $D^* = (D_{\mathrm{ROCO}} - D_{\mathrm{planner}})/D_{\mathrm{Rx}}$, while for all other measures, $D^* = (D_{\mathrm{planner}} - D_{\mathrm{ROCO}})/D_{\mathrm{constraint}}$. D^* is therefore a fractional measure of target coverage or sparing, normalized to the prescription dose or clinical constraint; positive numbers are better for ROCO. Table III shows the median D^* for each structure: we tested the differences in the medians for statistical significance using the Wilcoxon signed-rank test[30]. We have found that differences in doses to the OARs (i.e., differences between the median values of D*) for the clinical plans and for ROCO are

Structure	**median D^*(%)**	**p < 0.1?**
PTV $D_{\max}$	-0.9	no
PTV $D_{\min}$	0.2	no
Lung $\bar{D}$	-0.9	no
Cord $D_{\max}$	-1.5	no
Esophagus $\bar{D}$	0.0	no

TABLE III. Median differences between ROCO plans and plans produced by treatment planners. D* is a target or organ dose normalized by the prescription dose (for PTVs) or clinical constraints (for organs; see text). Positive values of D* indicate either improved dose homogeneity in the PTVs, or better sparing of the OARs in ROCO plans, when compared to the treatment planner's plans. Significance was evaluated with the Wilcoxon signed-rank test; for ROCO, the median difference in doses to targets and organs was not significantly different from the treatment planner's.

not statistically significant.

These plans were generated in a short time, requiring around 30 minutes of CPU time. In contrast, treatment planners using conventional IMRT optimization required around 3 hours for the same task, which is an important amount of time in a busy clinic. ROCO can thus save a great deal of planner time: assuming that each readjustment-reoptimization cycle requires 10 minutes for a treatment planner to complete, Fig. 7 shows that ROCO saves a median of 1.75 hours.

ROCO plans have spared the OARs as well as the treatment planner's plans have, but as we can see from Fig. 5, the PCA modes from which the ROCO plans are constructed cannot be used to completely reconstruct the clinical plan: 10–20% of the intensity profile that the treatment planners use lies outside of the space spanned by these modes. However, the clinical plan should not be viewed as a "ground truth" correct answer; several authors have noted a high degree of degeneracy in IMRT plans, which result in similar objective function values but different clinical tradeoffs[23]. We conjecture that the degree of this degeneracy is greater for lung patients than for prostate patients, because our PCA modes are not able to represent the clinical plan as well: the value of the projection residual R (see Eq. (8) and Fig. 5) is much larger in the current study than it was found to be for prostate patients, where fewer than 25 modes sufficed to bring the projection residual to less than 1%.

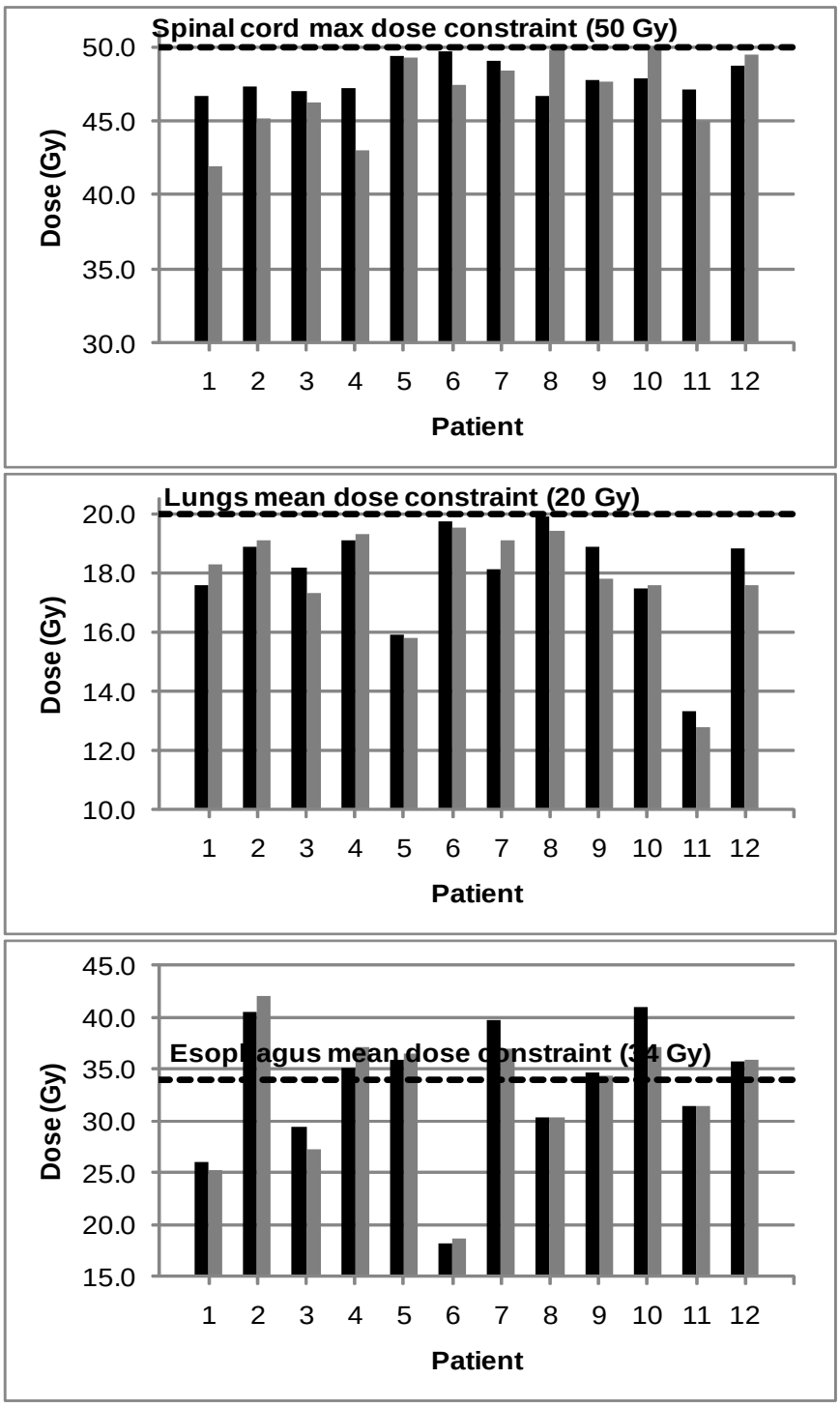

FIG. 6. Comparison of ROCO and treatment planner plans. Black bars show doses to organs for ROCO plans, and gray bars show the doses for the planner's plans. ROCO plans are normalized to have the same D_{95} as the planner's plan for each patient. The black dashed lines indicate the relevant clinical constraint. ROCO plans satisfy the same clinical constraints as the planner's plan in all cases.

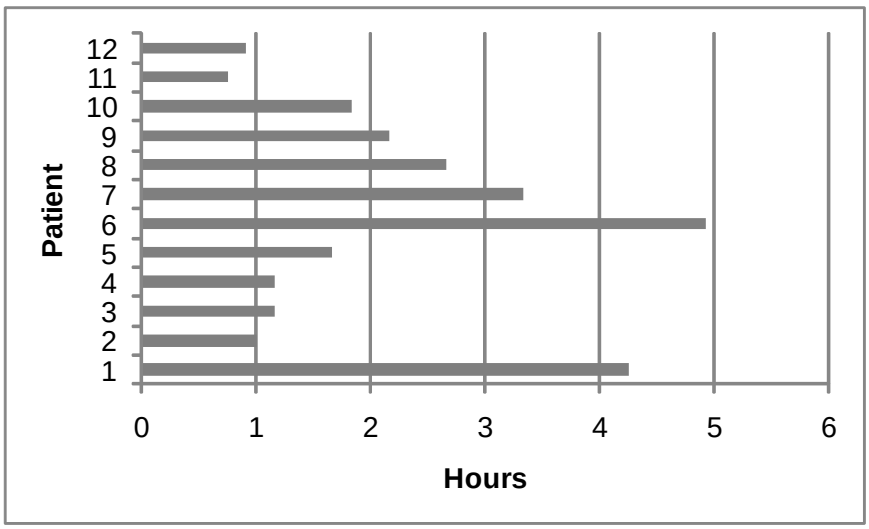

FIG. 7. Time saved by ROCO. ROCO requires about 30 minutes (largely unsupervised) to compute a treatment plan. This plot shows the time saved assuming that each adjustment-reoptimization cycle undertaken by the planner takes 10 minutes. The median time saved is 105 minutes.

C. Rind structures

An important aspect of current techniques for lung cancer treatment planning is the creation of non-specific normal tissue structures (referred to here as "rinds") that help avoid regions with undesirable high doses surrounding the PTV ("hot spots"). To make treatment planning less labor-intensive, it would be an advantage to use a standardized rind structure if it could be effective at avoiding hot spots, or if suitable plans could be created without such structures altogether. More generally, we would like to know whether adding constraints on new structures, such as a planner might impose after seeing an initial treatment plan, must be incorporated from the very beginning of the ROCO process, or if the lower-dimensional space generated by the sampling phase may already contain feasible solutions for the new constraints.

We set aside the rinds which had been previously created for each patient by the treatment planners, and created a standardized rind structure for each patient by leaving a 4 mm margin outside of the PTV, and then selecting a 3 cm annulus of tissue outside of this margin. We then used ROCO to plan patients, first without the rind structures present, and then with the standardized rinds in place during both the sampling and constrained optimization, and finally leaving them out during the sampling and including them in constrained optimization

Patient no.	1	2	3	4	5	6	7	8	9	10	11	12
Rinds:	No rinds present.											
PTV D_{95}	100.6	100.2	100.9	100.4	95.4	101.0	85.6	99.2	100.6	99.3	100.6	99.0
Rind $D_{\max}$	109.6	107.4	106.1	106.2	**119.9**	104.5	100.7	**110.7**	108.5	**114.3**	107.5	109.9
Rinds:	Rinds present in both sampling and constrained optimization.											
PTV D_{95}	100.6	100.2	100.9	99.6	95.1	101.0	91.6	99.2	100.6	98.9	100.6	99.0
Rind $D_{\max}$	109.6	107.4	106.1	105.6	107.0	104.5	110.0	110.0	108.5	110.0	107.5	109.9
Rinds:	Rinds in constrained optimization only.											
PTV D_{95}	100.1	100.1	100.7	100.4	98.1	101.0	93.2	98.9	100.5	98.0	100.5	98.7
Rind $D_{\max}$	109.0	108.4	105.7	106.2	110.0	104.5	106.8	106.2	106.6	110.0	106.1	107.0

TABLE IV. ROCO target coverage and rind maximum dose for three different cases: in the first row, no rinds are included in the optimization; in the second row, rind constraints are included in both the sampling and hard constraints; and in the third row, rind constraints are present in the hard constraint optimization step only. Patients 5, 8, and 10 had unacceptable hot spots outside the PTV when rind structures were not included in the ROCO constraints. Rind constraints reduced these doses to acceptable levels (110% of PTV max dose), and the optimizer had sufficient freedom to do this even without the presence of rinds during sampling.

only. Table IV summarizes the results from this experiment. Without the rind structures in place, hot spots were found outside of the PTV in 3 of the patients. After we ran ROCO with the standardized rind structures in place, we found that the hot spots in the 3 cm region outside of the PTV were successfully suppressed. Further, we found that it was not necessary to include the rind structures during the sampling stage: the optimizer had sufficient freedom to honor the RIND constraints even if they had not been included when the sampling for the PCA basis was performed.

In 3 of the patients, hot spots persisted in regions further outside of the PTV than were covered by the standardized rind. For patients 9, 10, and 12, in order to get a clinically acceptable plan, we had to use the treatment planner's rind, which covered a larger volume; we conclude that creation of a standard rind structure based only on the PTV geometry is not a successful strategy for these kinds of lung cases. We observed that for these 3 cases, the hot spot appeared near the intersection of two beam edges, which suggests a strategy

that might be used to generate rinds to deal with this problem.

IV. DISCUSSION AND CONCLUSIONS

In this paper, we extended our previous work on ROCO in several important ways. First, we applied ROCO to a more complicated treatment site: the lung rather than the prostate, and showed that the same general algorithmic strategy produced clinically acceptable plans. We analyzed tradeoffs in sampling and dimensionality reduction and showed that acceptable plans could be obtained in about 30 minutes, a major time savings over the manual trial-and-error process of unconstrained optimization. ROCO plans satisfy all of the clinical constraints that were satisfied by the planner's plans; with the same PTV D_{95}, there were no significant differences between the OAR sparing achieved by ROCO and the organ sparing achieved by the clinical plans. From these results, we are confident that ROCO will be flexible enough for general external beam radiation therapy planning, and is not confined to simpler treatments such as prostate cancer.

A major improvement we made to ROCO in our current work is our incorporation of ROCO into MSKCC's clinical treatment planning system. ROCO is now capable of reading and writing beam and dose information directly to/from the treatment planning system. Most importantly, ROCO uses the clinical full dose calculation to evaluate the dose distributions corresponding to each PCA mode. Using an approximate truncated dose kernel resulted in an inaccurate dose calculation, which proved to be a major difficulty in our previous work.

Ideally, ROCO would return a solution satisfying the specified hard constraints if any such feasible solution exists, and a satisfactory plan would result. In clinical practice, some iterative modification of parameters is inevitable: the notion of clinical acceptability — which varies from clinic to clinic or even planner to planner — is extremely difficult to pose either as an objective function or a hard constraint.In the future, we need to develop new constraints (e.g., rind-type structures to suppress hot spots in normal tissue) or objective function terms (e.g., to try and bias the solution towards more uniform PTV coverage). The key advantage of ROCO with respect to the trial-and-error loop typical of conventional soft-constrained IMRT is that such constraints can be posed and a solution found within a few minutes. This is true because the time-consuming parameter-sampling step to generate the

PCA vectors is only done once, independent of the constraints; the constrained optimization is performed quickly in the low-dimensional space, and the new solution, if one exists, is guaranteed to satisfy the constraints. This makes any trial-and-error much less tedious and the control over the solution much more direct.

Improving planner time savings is one of the primary goals of our future work with ROCO. We plan to apply ROCO to head and neck cancer, which remains a challenging site for current IMRT planning techniques: because of the complexity of dose-painting and the large number of OARs in treatment fields, head and neck plans can require days of planner time, and even then the space of clinical tradeoffs between OAR sparing and target coverage may not have been fully explored. ROCO will be able to improve these limitations by reducing the time it takes to obtain a plan that satisfies clinical constraints.

ACKNOWLEDGMENTS

This publication was supported in part by Grant Number 1R01CA148876-02 from the National Cancer Institute (NCI), a grant from Varian Medical Systems, and by a private donor to Rensselaer Polytechnic Institute. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute, National Institutes of Health. We would like to thank Gig Mageras, Perry Zhang, Joseph McNamara, Howard Amols, Margie Hunt, and Chen Chui for helpful discussions.

Conflict of Interest: Research partially supported by Varian Corporation.

---

* stabenah@mskcc.org

† riverl2@rpi.edu

‡ yorkee@mskcc.org

§ yangj@mskcc.org

¶ renzhilu@gmail.com

** rjradke@ecse.rpi.edu

†† jacksona@mskcc.org

[1] Michael J. Zelefsky, Victor E. Reuter, Zvi Fuks, Peter Scardino, and Alison Shippy, "Influence

of local tumor control on distant metastases and cancer related mortality after external beam radiotherapy for prostate cancer," The Journal of Urology, **179**, 1368 – 1373 (2008).

[2] Deborah A. Kuban, Susan L. Tucker, Lei Dong, George Starkschall, Eugene H. Huang, M. Rex Cheung, Andrew K. Lee, and Alan Pollack, "Long-term results of the M. D. Anderson randomized dose-escalation trial for prostate cancer," International Journal of Radiation Oncology*Biology*Physics, **70**, 67 – 74 (2008).

[3] Sue S. Yom, Zhongxing Liao, H. Helen Liu, Susan L. Tucker, Chao-Su Hu, Xiong Wei, Xuanming Wang, Shulian Wang, Radhe Mohan, James D. Cox, and Ritsuko Komaki, "Initial evaluation of treatment-related pneumonitis in advanced-stage non-small-cell lung cancer patients treated with concurrent chemotherapy and intensity-modulated radiotherapy," International Journal of Radiation Oncology*Biology*Physics, **68**, 94 – 102 (2007).

[4] R Lu, R Radke, L Happersett, J Yang, C Chui, E Yorke, and A Jackson, "Reduced-order constrained optimization in IMRT planning," Physics in Medicine and Biology, **53**, 6749–6766 (2008).

[5] American Cancer Society, "Cancer facts & figures 2010," http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-026238.pdf (2010).

[6] C. C. Ling *et al.*, *A Practical Guide to Intensity-Modulated Radiation Therapy* (Medical Physics Publishing, 2004).

[7] E. Yorke, A. Jackson, L. Braban, K. Rosenzweig, and C. Ling, "Advantages of IMRT for dose escalation in radiation therapy for lung cancer [abstract]," Medical Physics, **28**, 1291 (2001).

[8] Marc L. Kessler, Daniel L. Mcshan, Marina A. Epelman, Karen A. Vineberg, Avraham Eisbruch, Theodore S. Lawrence, and Benedick A. Fraass, "Costlets: A generalized approach to cost functions for automated optimization of IMRT treatment plans," Optimization and Engineering, **6**, 421–448 (2005).

[9] Q. Wu, R. Mohan, A. Niemierko, and R. Schmidt-Ullrich, "Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose," Int J Radiat Oncol Biol Phys., **52**, 224–235 (2002).

[10] S.V. Spirou and C.S. Chui, "A gradient inverse planning algorithm with dose-volume constraints," Medical Physics, **25**, 321–333 (1998).

[11] J.A. Purdy *et al.*, "Intensity-modulated radiotherapy: current status and issues of interest," Int

446 J Radiat Oncol Biol Phys., **51**, 880–914 (2001).

447 [12] Steve Webb, *Intensity-Modulated Radiation Therapy* (Taylor and Francis, 2001).

448 [13] James L. Bedford and Steve Webb, "Elimination of importance factors for clinically accurate
449 selection of beam orientations, beam weights, and wedge angles in conformal radiation therapy,"
450 Medical Physics, **30**, 1788–1804 (2003).

451 [14] A. Samuelsson and K.A. Johansson, "Intensity modulated radiotherapy treatment planning for
452 dynamic multileaf collimator delivery: influence of different parameters on dose distributions,"
453 Radiother Oncol., **66**, 19–28 (2003).

454 [15] G. Starkschall, A. Pollack, and C.W. Stevens, "Treatment planning using a dose-volume feasi-
455 bility search algorithm," Int J Radiat Oncol Biol Phys., **49**, 1419–1427 (2001).

456 [16] T. Bortfeld, K.-H. Küfer, M. Monz, A. Trofimov, and A. Niemierko, "Problems with current
457 IMRT prescription practices and planning systems [abstract]," Medical Physics, **31**, 1761 (2004).

458 [17] Renzhi Lu, Richard J Radke, Laura Happersett, Jie Yang, Chen-Shou Chui, Ellen Yorke, and
459 Andrew Jackson, "Reduced-order parameter optimization for simplifying prostate IMRT plan-
460 ning," Physics in Medicine and Biology, **52**, 849 (2007).

461 [18] Eva K. Lee, Tim Fox, and Ian Crocker, "Integer programming applied to intensity-modulated
462 radiation therapy treatment planning," Annals of Operations Research, **119**, 165–181 (2003),
463 10.1023/A:1022938707934.

464 [19] Tarek Halabi, David Craft, and Thomas Bortfeld, "Dose-volume objectives in multi-criteria
465 optimization," Physics in Medicine and Biology, **51**, 3809 (2006).

466 [20] Tobias Spalke, David Craft, and Thomas Bortfeld, "Analyzing the main trade-offs in multiob-
467 jective radiation therapy treatment planning databases," Physics in Medicine and Biology, **54**,
468 3741 (2009).

469 [21] J.T. Lyman, "Complication probability as assessed from dose-volume histograms," Radiation
470 Research, **104**, S13–S19 (1985), cited By (since 1996) 330.

471 [22] C. Burman, G.J. Kutcher, B. Emami, and M. Goitein, "Fitting of normal tissue tolerance data
472 to an analytic function," International Journal of Radiation Oncology*Biology*Physics, **21**, 123
473 – 135 (1991).

474 [23] M. Alber, G. Meedt, F. Nüsslin, and R. Reemtsen, "On the degeneracy of the IMRT optimiza-
475 tion problem," Medical Physics, **29**, 2584–2589 (2002).

476 [24] Andrea Saltelli, Stefano Tarantola, Francesca Campolongo, and Marco Ratto, *Sensitivity Anal-*

ysis in Practice: A Guide to Assessing Scientific Models (Wiley, 2004).

[25] John Shawe-Taylor, *Kernel Methods for Pattern Analysis* (Cambridge University Press, 2004).

[26] J Menhel, D Levin, D Alezra, Z Symon, and R Pfeffer, "Assessing the quality of conformal treatment planning: a new tool for quantitative comparison," Physics in Medicine and Biology, **51**, 5363 (2006).

[27] Loc Feuvret, Georges Nol, Jean-Jacques Mazeron, and Pierre Bey, "Conformity index: A review," International Journal of Radiation Oncology*Biology*Physics, **64**, 333 – 342 (2006).

[28] Lucullus Hing Tong Leung, Monica Wai Kwan Kan, Ashley Chi Kin Cheng, Wicger King Hang Wong, and Chun Chung Yau, "A new dose-volume-based plan quality index for IMRT plan comparison," Radiotherapy and Oncology, **85**, 407 – 417 (2007).

[29] Grace Tang, Matthew A. Earl, Shuang Luan, Chao Wang, Majid M. Mohiuddin, and Cedric X. Yu, "Comparing radiation treatments using intensity-modulated beams, multiple arcs, and single arcs," International Journal of Radiation Oncology*Biology*Physics, **76**, 1554 – 1562 (2010).

[30] Lloyd Fisher and Gerald van Belle, *Biostatistics: a methodology for the health sciences* (Wiley-Interscience, 1939).