A critical evaluation of worst case optimization methods for robust intensity-modulated proton therapy planning

Albin Fredriksson[a)] and Rasmus Bokrantz
RaySearch Laboratories, Sveavägen 25, SE-111 34 Stockholm, Sweden

(Received 12 March 2014; revised 7 May 2014; accepted for publication 29 May 2014; published 3 July 2014)

Purpose: To critically evaluate and compare three worst case optimization methods that have been previously employed to generate intensity-modulated proton therapy treatment plans that are robust against systematic errors. The goal of the evaluation is to identify circumstances when the methods behave differently and to describe the mechanism behind the differences when they occur.
Methods: The worst case methods optimize plans to perform as well as possible under the worst case scenario that can physically occur (composite worst case), the combination of the worst case scenarios for each objective constituent considered independently (objectivewise worst case), and the combination of the worst case scenarios for each voxel considered independently (voxelwise worst case). These three methods were assessed with respect to treatment planning for prostate under systematic setup uncertainty. An equivalence with probabilistic optimization was used to identify the scenarios that determine the outcome of the optimization.
Results: If the conflict between target coverage and normal tissue sparing is small and no dose-volume histogram (DVH) constraints are present, then all three methods yield robust plans. Otherwise, they all have their shortcomings: Composite worst case led to unnecessarily low plan quality in boundary scenarios that were less difficult than the worst case ones. Objectivewise worst case generally led to nonrobust plans. Voxelwise worst case led to overly conservative plans with respect to DVH constraints, which resulted in excessive dose to normal tissue, and less sharp dose fall-off than the other two methods.
Conclusions: The three worst case methods have clearly different behaviors. These behaviors can be understood from which scenarios that are active in the optimization. No particular method is superior to the others under all circumstances: composite worst case is suitable if the conflicts are not very severe or there are DVH constraints whereas voxelwise worst case is advantageous if there are severe conflicts but no DVH constraints. The advantages of composite and voxelwise worst case outweigh those of objectivewise worst case.
[http://dx.doi.org/10.1118/1.4883837]

Key words: robust optimization, IMPT, uncertainty, worst case

1. INTRODUCTION

The dosimetric benefits of intensity-modulated proton therapy (IMPT) come at the cost of a high sensitivity to errors. In particular, patient misalignments and errors in the conversion from computed tomography densities to proton stopping power can severely compromise the quality of IMPT plans.[1,2] The risk of errors has spurred the development of optimization methods that explicitly take uncertainties into account in order to create robust plans.[3–10]

Methods that explicitly consider uncertainties have been extensively compared to planning that does not, or planning that takes uncertainties into account implicitly by using planning margins (as per ICRU recommendations[11]), and have been found to provide plans that deteriorate less if errors occur—see Stuschke *et al.*[12] in addition to the previous citations. The literature on comparative studies between different methods for handling uncertainties is, however, relatively scarce. Fredriksson[4] compared optimization of the expectation of the objective value to optimization of the worst case objective value, and found that the latter has the advantage that it leads to a sharp dose fall-off outside the treated volume while the former leads to an extended fall-off that contributes little to target coverage but still increases the dose to healthy tissues. Casiraghi *et al.*[13] compared evaluation of treatment plan robustness using two methods that both aim to hedge against the worst possible error, but have distinct interpretations of "worst." They found that a conservative interpretation of "worst" can result in overly pessimistic predictions of plan quality. Nevertheless, the effects of the degree of conservatism employed during treatment plan optimization remain largely unexplored.

In this paper, we juxtapose three worst case methods with regard to how well they perform during treatment plan optimization, including the methods that Casiraghi *et al.*[13] considered with regard to robustness evaluation. The contribution of the present paper is that we illustrate how the worst case methods differ from each other, and explain how come. To answer the latter, we develop a technique for analyzing robust treatment plan optimization methods that is based on the observation that the methods are equivalent to expectation optimization problems, if the expectation is conditioned on specific probability distributions that are well-defined when the optimal solutions to the worst case formulations have been found.

The methods that are evaluated in this paper are:

- **Composite worst case:** The worst objective function value over the possible errors is minimized. All voxel doses are assumed to come from the same error realization. The method was introduced in Fredriksson *et al.*[5]
- **Objectivewise worst case:** Each objective constituent (e.g., a target minimum dose function or a rectum maximum dose function) is considered to be independently affected by the uncertainty. All voxel doses penalized by a given objective function are assumed to result from the same error realization, but different objectives can be affected by distinct error realizations. The method was introduced by Chen *et al.*[3] in a multicriteria optimization setting.
- **Voxelwise worst case:** Each voxel is assumed to be independently affected by the uncertainty, and the penalty to each voxel depends on the worst dose that the voxel can receive under the considered errors. The method was introduced as a linear program by Unkelbach *et al.*[10] and Chan,[14] and as a nonlinear program by Pflugfelder *et al.*[8]

2. METHODS

We consider robust optimization for IMPT on the form of minimization of the sum of functions $f_1, \ldots, f_n$, each penalizing some undesirable characteristic in the planned dose to some anatomical structure. The functions have associated non-negative weights $w_1, \ldots, w_n$ reflecting their relative importance, and evaluate with respect to the dose distribution d. The vector d, in turn, is a function of the non-negative spot weights x and the scenario s from the set $\mathcal{S}$ of considered error scenarios. The set $\mathcal{S}$ can be used to model any source of uncertainty, such as setup and range errors and organ motion, and probabilities need not be attached to the scenarios. It is imperative to bound the magnitudes of errors that are represented by $\mathcal{S}$ because worst case methods do not discriminate between the relative importance of the scenarios. In this paper, we use functions $f_1, \ldots, f_n$ that quadratically penalize deviations from the irradiated structures' prescribed minimum or maximum doses or dose-volume histogram (DVH) points, see, e.g., Oelfke and Bortfeld[15] for a mathematical definition.

2.A. Composite worst case optimization

Composite worst case minimizes the weighted sum of the objective constituents evaluated in the worst scenario, i.e.,

$$\underset{x\geq 0}{\text{minimize}} \quad \max_{s\in\mathcal{S}} \sum_{i=1}^{n} w_i f_i(d(x;s)). \tag{1}$$

Note that all constituents $f_1, \ldots, f_n$ are penalized in the same scenario s. This fact means that composite worst case retains the correlation between voxels.

2.B. Objectivewise worst case optimization

Objectivewise worst case considers the worst case scenario for each objective constituent independently, and is formulated according to

$$\underset{x\geq 0}{\text{minimize}} \quad \sum_{i=1}^{n} w_i \max_{s\in\mathcal{S}} f_i(d(x;s)). \tag{2}$$

Different worst case scenarios can be active for different functions $f_1, \ldots, f_n$, meaning that the formulation protects against both physical and unphysical combinations of scenarios. In total, it considers $|\mathcal{S}|^n$ combinations, where $|\mathcal{S}|$ denotes the number of elements in $\mathcal{S}$.

2.C. Voxelwise worst case optimization

Voxelwise worst case considers each voxel to be independently affected by the uncertainty. The basis for this method is the worst case dose distributions $d^{\min}$ and $d^{\max}$ introduced by Lomax *et al.*[16] for robustness evaluation, defined according to

$$d_v^{\min}(x) = \min_{s\in\mathcal{S}} d_v(x;s) \quad \text{or} \quad d_v^{\max}(x) = \max_{s\in\mathcal{S}} d_v(x;s)$$

for each voxel v, with $d^{\min}$ being used for dose-promoting functions and $d^{\max}$ for dose-limiting function. With the dose-promoting functions indexed by the set $\mathcal{T}$ and the dose-limiting functions indexed by the set $\mathcal{O}$, the voxelwise worst case optimization is formulated as

$$\underset{x\geq 0}{\text{minimize}} \quad \sum_{i\in\mathcal{T}} w_i f_i(d^{\min}(x)) + \sum_{i\in\mathcal{O}} w_i f_i(d^{\max}(x)). \tag{3}$$

For convexity to be preserved, only dose-promoting functions (e.g., minimum dose functions) can be applied to $d^{\min}$ and only dose-limiting functions (e.g., maximum dose functions) to $d^{\max}$. Uniform dose functions should be split into minimum and maximum dose functions applied to, respectively, $d^{\min}$ and $d^{\max}$. For details on compositions that preserve the convexity of functions, see, e.g., Boyd and Vandenberghe (Sec. 3.2.4 in Ref. 18). Because the uncertainty is allowed to affect each voxel individually, the number of scenarios considered by the voxelwise method is in the order of $|\mathcal{S}|^{|d|}$, where $|d|$ denotes the number of elements in d. Only $|\mathcal{S}|$ of these scenarios can occur in practice.

2.D. Equivalence of the methods when there are no conflicts

If there are no conflicting planning criteria and the optimization is performed with respect to minimum and maximum dose functions only, then the worst case methods (1)–(3) behave identically. This is so because the necessary and sufficient conditions for a minimum or maximum dose objective to evaluate to zero are the same for all methods. We exemplify by a minimum dose objective for the target: If there are no conflicts, then this function evaluates to zero for optimal solutions of the composite method. This implies that all target voxels are above the reference dose level under all scenarios. The function therefore also evaluates to zero for the objectivewise method. Moreover, $d_v^{\min}$ is above the reference dose level for all target voxels, so the function also evaluates to

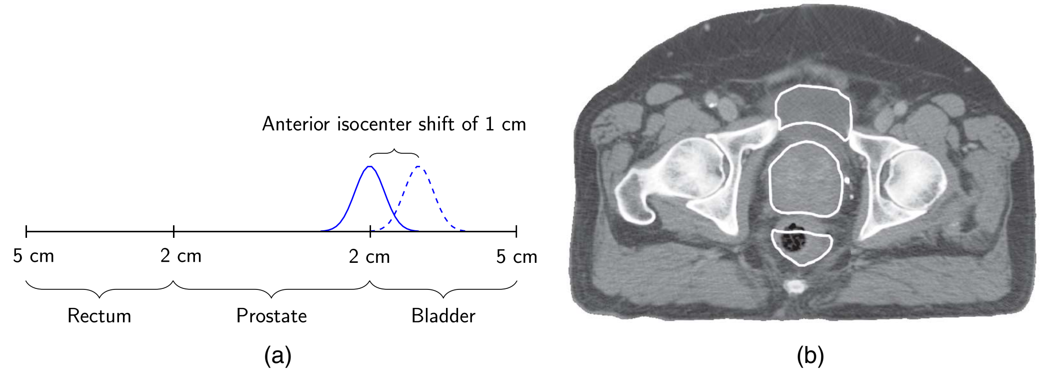

FIG. 1. Prostate cancer cases considered in the experimental study. (a) 1D geometry, (b) 3D geometry.

zero also for the voxelwise method. Taking these steps backwards shows that when the function evaluates to zero for the voxelwise method and there are no conflicts, then it evaluates to zero for the other two methods as well. The reasoning for maximum dose functions is similar.

2.E. Experimental study

The worst case methods were evaluated with respect to treatment planning for prostate subject to systematic setup uncertainty. Note that another important source of error in treatment planning for IMPT is range uncertainty,[1] and that it might be desirable to avoid sharp beam dose gradients in order to protect against misalignment of different beams. These aspects of robustness were not considered in the present study. Moreover, the methods might behave differently for disease sites and beam orientations other than the ones studied.

2.E.1. Patient geometries

Two prostate geometries were studied: a one-dimensional (1D) phantom and a three-dimensional (3D) dataset of clinical size, see Fig. 1. The 1D geometry schematically represents the intersection between a transversal and a sagittal cut through the 3D geometry's isocenter. The 1D geometry was planned with respect to one perpendicularly oriented field while the 3D geometry was planned with respect to two parallel-opposed fields at 90° and 270°. The prostate was prescribed with 60 Gy and the bladder and rectum treated as critical structures. Note that the two organs at risk (OARs) are directly adjacent to the prostate in the 1D geometry, whereas the posteriormost part of the bladder and the anteriormost part of the rectum are not part of the OARs for the 3D geometry.

2.E.2. Optimization formulations

The following optimization formulations were considered for each geometry:

- A symmetric formulation:
  - **Objective:** Uniform dose at 60 Gy to the prostate.
  - **Objective:** Maximum dose of 0 Gy to the OARs, with bladder and rectal sparing considered to be equally important.
- An asymmetric formulation:
  - **Objective:** Uniform dose at 60 Gy to the prostate.
  - **Objective:** Maximum dose of 0 Gy to the OARs, with rectal sparing considered to be five times more important than sparing of the bladder.
- A DVH-constrained formulation:
  - **Constraint:** Minimum dose of 60 Gy to 90% of the prostate.
  - **Objective:** Maximum dose of 0 Gy to the OARs, with bladder and rectal sparing considered to be equally important.

For the 3D case, an objective on maximum dose of 0 Gy to the unclassified tissue was included in all formulations.

2.E.3. Uncertainties

Uncertainties were considered on the form of systematic setup errors modeled as shifts of the beam isocenter. The uncertainty was assumed to be isotropic and errors up to 1 cm were included in the uncertainty set $\mathcal{S}$. For the 1D geometry, this set was discretized into scenarios that were 1.2 mm apart, yielding 19 scenarios. Such fine discretization is too computationally expensive to be practical in 3D. For the 3D geometry, the set $\mathcal{S}$ was therefore constituted of a total of 27 scenarios that, in addition to the nominal scenario, represented shifts of 1 cm in the positive and negative unit directions (six scenarios), all pairwise combinations of unit directions (12 scenarios), and all triplets of unit directions (eight scenarios). The use of scenarios that, except for the nominal scenario, all lie on the boundary of the uncertainty region is supported by the fact that the boundary scenarios generally constitute the worst case ones[4,17] and that the DVHs of the boundary scenarios provide good approximations of the worst case DVHs.[13]

2.E.4. Dose calculation and optimization

The absorbed proton dose was for the 1D geometry modeled by Gaussian spot kernels at a spacing of 1.2 mm and with a standard deviation of 3 mm. The dose distribution was discretized into 1.2 mm voxels and the optimizations performed using the nonlinear optimization solver SNOPT v7.2 (Stanford Business Software, Palo Alto, CA). The numerical

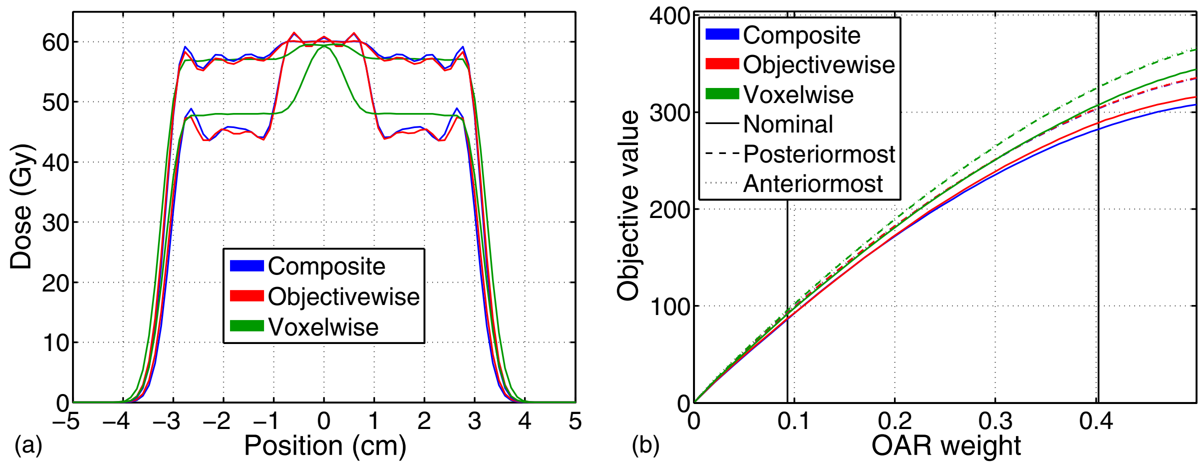

FIG. 2. Numerical results for the symmetric formulation applied to the 1D geometry. (a) Optimal dose in the nominal scenario with respect to the OAR weights that are indicated by vertical lines (slightly below 0.1 and above 0.4) in (b). (b) Objective values in three scenarios for the optimal dose as functions of weight on the OAR constituents (the target and OAR weights sum to unity). In the anteriormost shift scenario, the beam isocenter position is shifted from 0 to 1 cm. In the posteriormost shift scenario, it is shifted from 0 to −1 cm.

experiments for the 3D geometry were performed RayStation v4.0 (RaySearch Laboratories, Stockholm, Sweden), which we augmented with the capability of performing the three types of worst case optimization. We achieved continuous differentiability of the worst case formulations by substituting smooth and conservative approximations for maximum and minimum operators, see Appendix A. The dose was represented in a grid of $3 \times 3 \times 3$ mm^3 voxels and the proton fields were represented by spots at a line spacing and energy layer separation of 5 mm in water. RayStation's pencil beam dose algorithm was used for the dose calculations and the system's nonlinear programming solver used for the optimizations.

3. RESULTS

3.A. One-dimensional geometry

3.A.1. Symmetric formulation

The symmetric formulation for the 1D geometry was assessed by the following experiment: For each method, we first generated a representation of all treatment plans that the method can find. This we did by optimization over a fine discretization of the set of possible weights, i.e., the set $\{w \in \mathbb{R}^2_+ : w_1 + w_2 = 1\}$, where w_1 is the weight for the target and w_2 the weight for the rectum and bladder (these objectives are prioritized equally and can therefore be grouped into a single function). For each method, we then generated plans by calculating the convex combination of plans that had minimal composite objective value in the worst case. These calculations were performed with the priority of the OAR objective varied continuously from a weight of zero to being weighted equally with the target objective.

The composite method and the objectivewise method behaved nearly identically in this experiment. Figure 2(a) shows that both methods gave a dose below the prescription to target regions where the dose can become displaced to an OAR if a setup error occurs, while uniform doses at the prescription were given to the central target region. High-dose peaks are present at the boundary of the treated volume as well as at the boundary of the central region. The voxelwise method behaved differently: its optimal dose lacks the inner high-dose plateau and instead only meets the prescription at the origin. It also lacks the high-dose peaks, and its objective values are worse than those of the other two methods, increasingly so with increasing emphasis on OAR sparing, see Fig. 2(b).

3.A.2. Asymmetric formulation

The asymmetric formulation was analyzed in an analogous fashion as the symmetric formulation, except that the bladder and rectum objectives were treated separately. A separation is necessary for a fair comparison of the methods because the different methods can interpret the five times higher priority of rectal sparing differently.

Figure 3(b) shows that the composite and voxelwise methods adapted to the asymmetric priority between the bladder and rectum, while the objectivewise method did not. The anterior dose reduction of the objectivewise method's optimal dose leads to a lower objective value than for the other methods in the anteriormost shift scenario, see Fig. 3(a). The opposite effect occurs in the posteriormost shift scenario: the high-dose region is shifted from the target into the rectum and the low-dose region that nominally is placed within the bladder is shifted into the target region, causing a severe underdosage. The lack of target coverage in the posteriormost shift scenario is reflected by high objective values.

The composite method delivered a relatively high dose to the bladder that leads to worse objective values in the anteriormost shift scenario compared to the solutions produced by objectivewise worst case. In the posteriormost shift scenario, however, the nominally high bladder dose becomes displaced to the prostate where it contributes to coverage. Composite worst case therefore shows markedly better objective values than objectivewise if the posteriormost shift occurs. The objective values for composite worst case are also better in the nominal scenario.

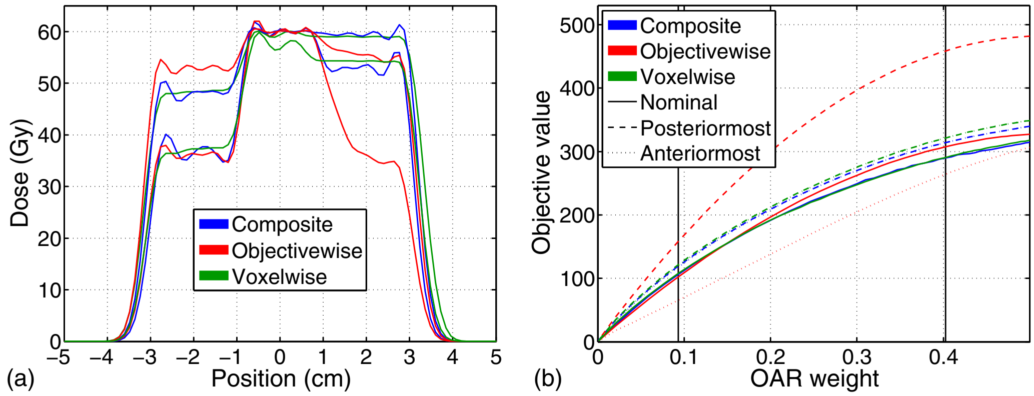

FIG. 3. Numerical results for the asymmetric formulation applied to the 1D geometry. (a) Optimal dose in the nominal scenario with respect to the OAR weights that are indicated by vertical lines (slightly below 0.1 and above 0.4) in (b). (b) Objective values in three scenarios for the optimal dose as a function of the weight on the OAR constituents (the target and OAR weights sum to unity). In the anteriormost shift scenario, the beam isocenter position is shifted from 0 to 1 cm. In the posteriormost shift scenario, it is shifted from 0 to −1 cm.

The voxelwise method behaved qualitatively similar to the composite method: a relatively high dose was placed within the bladder whereas the rectum was spared to a large extent. A difference, however, is that no high-dose peaks are present and that the fall-off from prescription to a lower dose occurs in the center of the target instead of around the 1 cm point. The objective values for voxelwise worst case are slightly worse than for composite worst case but better than for objectivewise worst case.

3.A.3. DVH-constrained formulation

The DVH-constrained formulation was evaluated by a single solve of each of the constrained counterparts of the formulations (1), (2), or (3). The solves led to nearly identical results for the composite and objectivewise methods: both methods gave plans that satisfied the DVH criterion of 90% worst case coverage at the prescription level and underdosed the target volume beyond the 90% volume level in order to maximize OAR sparing, see Fig. 4. The voxelwise method interpreted the DVH constraint more conservatively and accommodated close to 95% worst case coverage at the prescription level. This conservatism led to higher absorbed dose in the OARs and therefore worse objective function value than for the other two methods (the voxelwise method had 21% higher objective value than the composite method, while the objectivewise method had 1% higher objective value than the composite method).

3.B. Analysis of worst case probability distributions

To analyze the worst case methods, we use that optimization of the worst case is equivalent to optimization of the expected value conditioned on a specific probability distribution that we refer to as the "worst case probability distribution." The mathematical details of worst case probability distributions are given in Appendix B.

Worst case probability distributions associated with the optimal solutions to the symmetric and asymmetric formulations were calculated in order to permit visualization of which scenarios that predominantly determine the outcome of the optimization. Figure 5(a) shows that for the symmetric formulation, the composite method considers the two extreme scenarios to an equal extent, and neglects all other scenarios. The objectivewise method considers rectal sparing almost exclusively with respect to the posteriormost shift scenario, bladder sparing almost exclusively with respect to the anteriormost shift scenario, and target coverage with respect to the two extreme scenarios at equal weighting, with the other scenarios neglected. The worst case probability distribution for voxelwise worst case is similar to that for objectivewise worst case for the OAR voxels, although the penultimate and antepenultimate extreme scenarios are considered for the voxels close to the target. The target is, however, considered in a different fashion: here, the anteriormost shift scenario is considered for the posterior part of the target, whereas the posteriormost shift scenario is considered for the anterior part of the target.

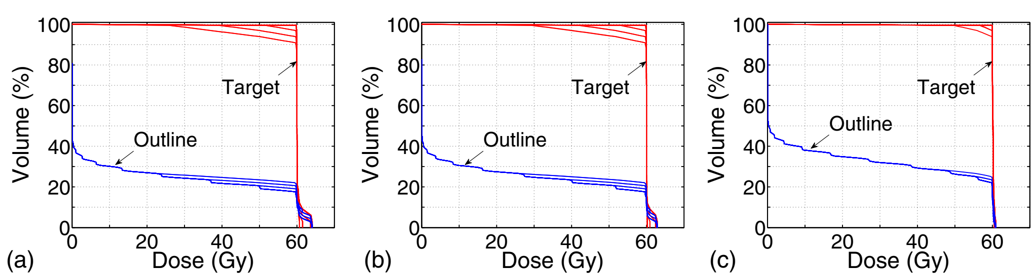

FIG. 4. DVHs for the optimal plans with respect to the DVH-constrained formulation for the 1D geometry in each of the 19 shift scenarios: (a) composite, (b) objectivewise, (c) voxelwise. The outline is the union of the rectum and bladder.

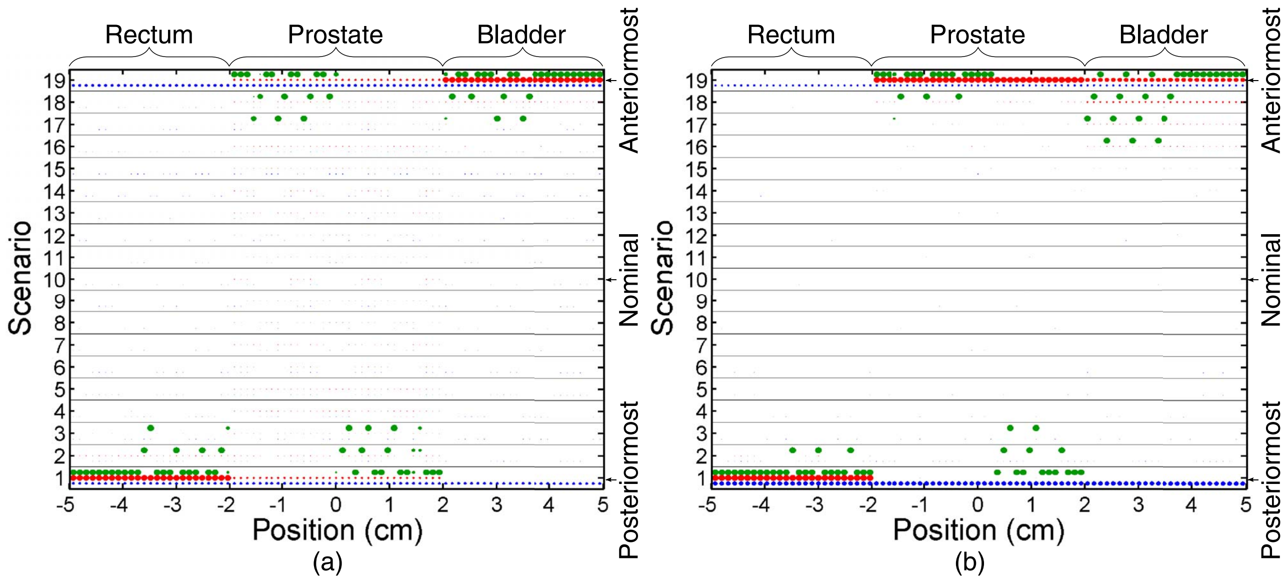

FIG. 5. Voxelwise probabilities for the 19 shift scenarios of the 1D case as optimized by the composite method, the objectivewise method, and the voxelwise method. For each scenario, the bottom row of dots corresponds to the composite method, the center row to the objectivewise method, and the upper row to the voxelwise method. For each voxel, the radii of the circles represent the probabilities under which the different scenarios are considered. The probabilities sum to unity over the scenarios for each voxel. Scenario 1 corresponds to the posteriormost shift of the dose and scenario 19 to the anteriormost shift. (a) Symmetric formulation, (b) asymmetric formulation.

For the asymmetric formulation, Fig. 5(b) shows that the composite method neglects all scenarios but the extreme ones, similar to its behavior for the symmetric formulation. Unlike in the symmetric case, the posteriormost shift scenario is considered to a higher extent than the anteriormost shift scenario. The objectivewise method considers the OARs in the same way as for the symmetric case, while the target is considered differently: the anteriormost shift scenario is considered almost exclusively. The voxelwise method has a similar emphasis on the scenarios as for the symmetric case, but the rectum and posterior part of the target use the penultimate and antepenultimate extreme scenarios to a lesser degree than in the symmetric case, and the bladder uses the nonextreme scenarios to a higher degree.

3.C. Three-dimensional geometry

The worst case methods' behaviors with respect to the 3D geometry were evaluated by a single solve of Eqs. (1)–(3) for each optimization formulation, performed using identical objective weights for all three methods. Recall from Sec. 3.A.2 that identical weights need not produce perfectly comparable results between the methods. The results should therefore be interpreted qualitatively.

3.C.1. Symmetric formulation

The most distinct difference between the methods in the results for the symmetric formulation applied to the 3D geometry is that the voxelwise method displays a protrusion of the high-dose region in the inferior and anteroinferior directions, see Fig. 6. Consequently, it has lower target objective values than the other methods in the superior and posterosuperior shift scenario, see Fig. 7. Although less noticeable in the dose plots, the same holds true for the inferior and anteroinferior shift scenarios. The composite and objectivewise methods differ to a larger extent than for the 1D case. While the composite method extends the high-dose region somewhat in the anterior, posterior, and inferior directions, the objectivewise method does not, and therefore has higher target—and total—objective values than the other methods in most scenarios except the left, right, and nominal. The posterior and posteroinferior scenarios are limiting for the composite method. While the composite method has lower objective value in

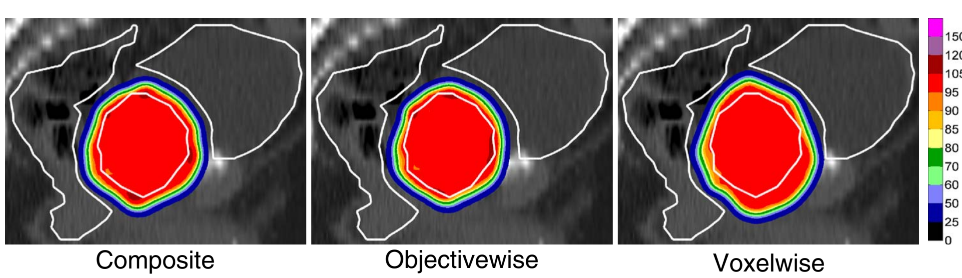

FIG. 6. Optimal dose with respect to the symmetric formulation applied to the 3D geometry. The depicted dose distributions are sagittal cuts through the isocenter of the dose in the nominal scenario. The color table is defined in percentages of the prescribed target dose.

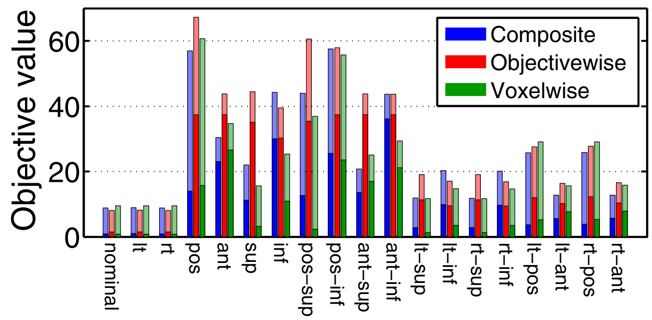

FIG. 7. Objective value in various scenarios for the optimal dose to the symmetric formulation for the 3D geometry (lt=left, rt=right, pos=posterior, ant=anterior, sup=superior, inf=inferior). The dark-colored parts of the bars indicate target constituent's contribution to the total objective value.

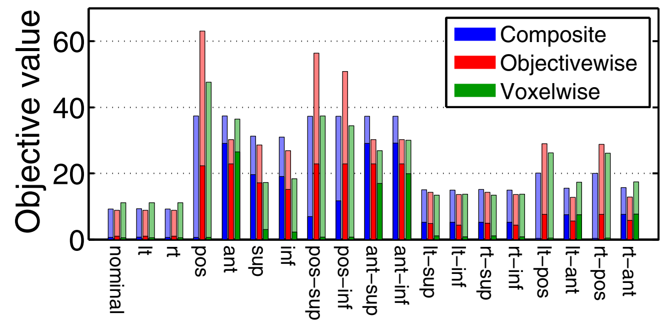

FIG. 9. Objective value in various scenarios for the optimal dose to the symmetric formulation for the 3D geometry (lt=left, rt=right, pos=posterior, ant=anterior, sup=superior, inf=inferior). The dark-colored parts of the bars indicate target constituent's contribution to the total objective value.

the worst case scenario, which is the posterior scenario, this comes at the cost of higher objective values than the voxelwise method in many other scenarios.

3.C.2. Asymmetric formulation

For the asymmetric formulation applied to the 3D geometry, the voxelwise method extended the high-dose region in the anterior, superior, and inferior directions, while the composite method restricted the extension to the anterior direction, and the objectivewise method gave a close to symmetric, nonextended, dose, see Fig. 8. This causes the objectivewise method to underdose the target if a posterior shift occurs, as reflected by its high target objective values for posterior, posterosuperior, and posteroinferior shifts, see Fig. 9. In contrast, the composite and voxelwise methods have low target objective values in these scenarios. Because the voxelwise method extended the high-dose region in the inferior and superior directions, it has lower objective values than the other methods in these scenarios. The enlarged high-dose region also explains how come its total objective value in the posterior scenario is higher than that of the composite method.

3.C.3. DVH-constrained formulation

The results for the DVH-constrained formulation applied to the 3D geometry were highly consistent with those for the 1D geometry. These results are summarized in Fig. 10 for completeness.

4. DISCUSSION

The goal of this work was to identify when the three worst case optimization methods behave differently, and to describe the mechanism behind the differences when they occur. With regard to the first question, we showed that there is no difference between the methods when there are no conflicts between target coverage and OAR sparing and no DVH constraints. Experimentally, we showed that when these conditions do not hold, the methods have notable differences. Below, we summarize the methods' characteristics and discuss the underlying mechanisms.

4.A. Composite worst case

A clear difference between the composite and voxelwise methods is how they handle "easy" scenarios. In particular, the composite method sometimes disregards easier scenarios, which is something that the voxelwise method does not. Examples of this behavior are the posterosuperior shift scenario in Fig. 7 and the superior and inferior shift scenarios in Fig. 9, where the composite method fails to maintain target coverage. This tendency of disregarding easier scenarios occurs because the composite method puts all effort on the worst case scenarios. In other words, the composite method favors

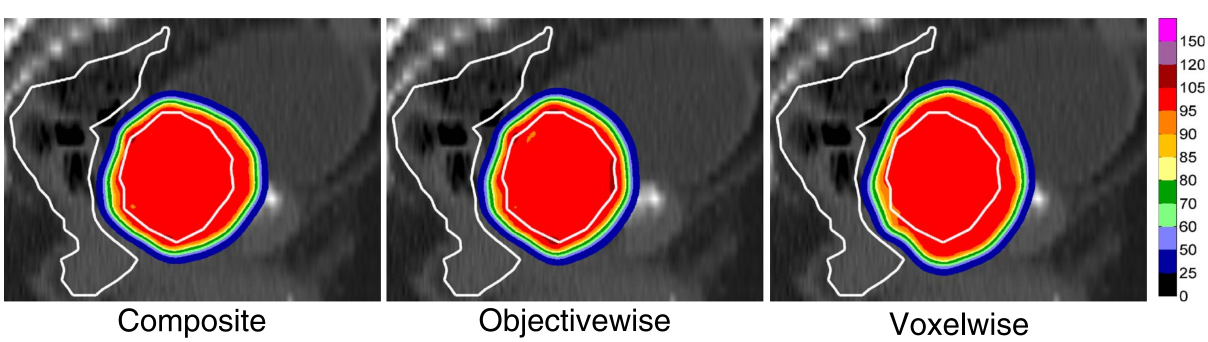

FIG. 8. Optimal dose with respect to the asymmetric formulation applied to the 3D geometry. The depicted dose distributions are sagittal cuts through the isocenter of the dose in the nominal scenario. The color table is defined in percentages of the prescribed target dose.

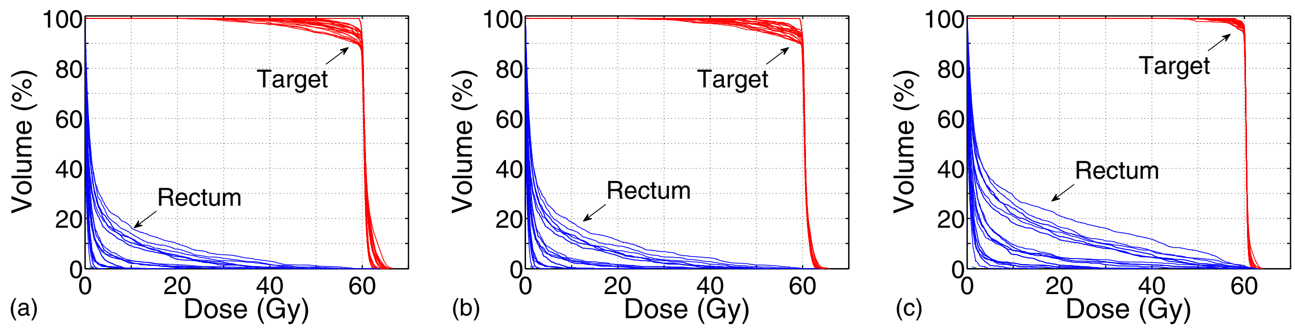

FIG. 10. DVHs for the optimal plans with respect to the DVH-constrained formulation for the 3D geometry in each of the 27 shift scenarios: (a) composite, (b) objectivewise, (c) voxelwise.

lower dose to the healthy tissue in the worst case scenarios over improved coverage in the "easy" scenarios, see Figs. 7 and 9. Note, however, that the composite method's focus on the worst case scenario also contributes to its main merit: it gives better worst case objective function values than the other two methods, as reflected by the results for both the 1D and 3D geometry.

4.B. Objectivewise worst case

Objectivewise worst case leads to symmetric dose distributions even if the optimization formulation is asymmetric, in contrast to the other methods, see Fig. 3. Figure 5(b) shows that the symmetry arises because for the target, the objectivewise method only takes the anteriormost shift into account, whereas the other two methods also take target coverage in the posteriormost shift scenario into account (the composite method assigns less probability to the anteriormost shift than the posteriormost shift because bladder sparing is considered less important than rectal sparing). The objectivewise method thus has no incentive to improve target coverage in the posteriormost shift scenario and consequently gives a symmetric dose where the dose level in the conflict regions is governed by the tradeoff between the target in the anteriormost shift scenario and the rectum in the posteriormost shift scenario.

For the symmetric 1D case, the objectivewise method behaved nearly identically to the composite method, and outperformed the voxelwise method. The similarity between composite and objectivewise worst case follows from that their worst case probability distributions are identical with respect to the target voxels if the formulation is symmetric (the extreme scenarios are the most difficult ones, and they are equally difficult), see Fig. 5(a). The distributions differ in the OARs, but this difference only results in a different scaling of the target objective relative to the OARs. However, for the symmetric 3D case, the objectivewise method performed notably worse than the other methods: fully symmetric cases are unlikely to come up in practice.

4.C. Voxelwise worst case

The optimal dose distributions of the voxelwise method show less spatial variability than those of the two other methods. In particular, the voxelwise method does not produce high-dose peaks at the boundary of the high-dose region. The reason that composite and objectivewise worst case create such peaks is that they contribute to a sharper dose fall-off (the relative decrease of a Gaussian function is larger farther from its maximum), and thereby better target conformance. An explanation to why the voxelwise method does not create peaks is given in Fig. 11(a), which depicts the worst case dose distributions $d^{\min}$ and $d^{\max}$ associated with the results in Fig. 2(b). When peaks are present, they affect $d^{\min}$ and $d^{\max}$ detrimentally over a large region, thus yielding a large increase in the

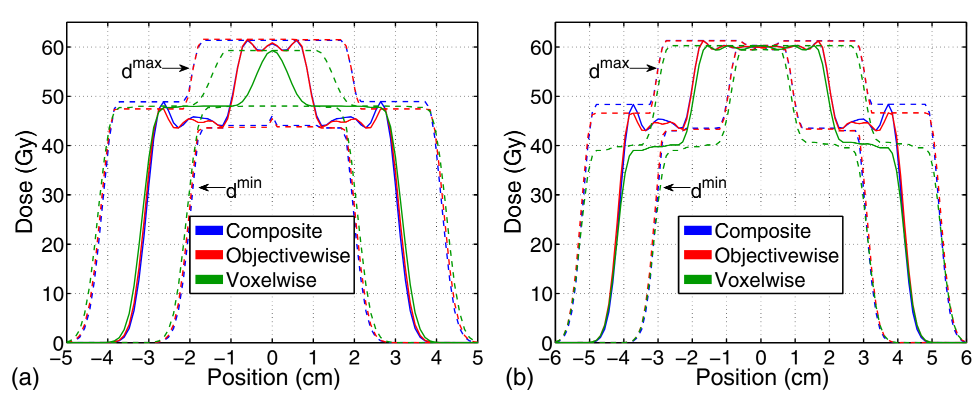

FIG. 11. (a) Worst case dose distributions associated with the plans depicted in Fig. 2(b). (b) Corresponding worst cases doses with respect to a 6 cm wide target.

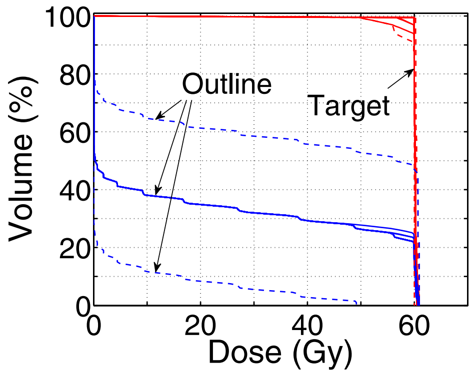

FIG. 12. DVHs for the worst case dose distributions $d^{\min}$ and $d^{\max}$ associated with the plan depicted in Fig. 2(b) (dashed). DVHs for the scenario dose distributions are shown for reference (solid). The outline is the union of the rectum and bladder.

objective value (the plateaus of $d^{\min}$ and $d^{\max}$ for the composite and objectivewise methods are at different levels, whereas those for the voxelwise method are on the same level).

Figure 11(a) also explains why the voxelwise method fails to meet the prescription even for the central region of the target. In particular, $d^{\max}$ is always below the target prescription, and so is nowhere penalized for the target, whereas $d^{\min}$ is constant at a level slightly below 50 Gy within most of the target. The central target voxels thus have a worst case minimum dose of about 50 Gy, and are not positively affected by an increase of the central dose. For their worst case minimum dose to increase, the doses at ± 1 cm would have to be increased, but this would affect the maximum worst case doses to the OARs detrimentally. This behavior shows that the high conservatism of the method can result in plans that take precaution against things that cannot happen. In a follow-up experiment, we however found that if the target is larger, the voxelwise method no longer fails to meet the prescription, see Fig. 11(b) where the target has been extended to encompass the region -3 to 3 cm. This result is in accordance with the results for the 3D geometry, where the target has a diameter of about 5.5 cm and the voxelwise method met the prescription.

Another important characteristic of voxelwise worst case is that it interprets DVH constraints in an overly conservative manner, which can result in higher target and OAR doses than necessary, see Fig. 4(c). This conservatism occurs because voxelwise optimization considers the DVH curves of $d^{\min}$ and $d^{\max}$, which are very conservative bounds on the true scenario DVHs, see Fig. 12. Because the composite and objectivewise methods consider the true scenario DVHs, these methods do not become overly conservative when applied to DVH constraints.

A positive result for the voxelwise method is that it gave the best objective value on average with respect to the 3D geometry, see Figs. 7 and 9. This result is linked to the fact that any extreme scenario is likely to be the worst case scenario in some voxels (the posteriormost shift scenario is, e.g., likely to be the worst case scenario for voxels in OARs that are on the posterior side of the target, and for target voxels on the anterior periphery of the target), which gives voxelwise worst case incentive to increase the target dose in all extreme scenarios.

5. CONCLUSION

The three worst case methods were found to have clearly different behaviors in general, and no particular method is without doubt superior to the others. A sound recommendation is therefore to make the choice of robust method on a case-by-case basis, with the choice guided by the different methods' pitfalls. To practitioners, we recommend that they be aware of which robust optimization method that is implemented in the treatment planning system of use, and of the method's advantages and disadvantages. The advantages and disadvantages of the studied methods are summarized below:

- Composite worst case gives the best worst case objective values, a sharp dose fall-off, and is well applicable to DVH constraints. Its main disadvantage is that it sometimes disregards easy scenarios. The method is therefore a good choice when the conflicts are not so severe that a lot of easy scenarios are disregarded, or when planning with respect to DVH constraints.
- In all cases, objectivewise worst case was outperformed by either the composite or voxelwise worst case. Objectivewise worst case is designed for permitting multicriteria optimization, something that the composite method does not trivially extend to. Based on our empirical results, the voxelwise method, which is equally well-suited for multicriteria optimization, appears to be a better option unless DVH constraints are used.
- Voxelwise worst case has the benefit that it does not disregard easy scenarios and can therefore be a better option than composite worst case when there are severe conflicts. Its conservatism makes the method ill-suited for planning with respect to DVH constraints, and renders the method unable to exploit spatial variability in the dose distribution such as high-dose peaks that promote a sharp dose fall-off.

In view of these characteristics, a method that combines the composite method's sharp dose fall-off and ability to handle DVH constraints with the voxelwise method's ability to handle severe conflicts is desirable.

APPENDIX A: CONTINUOUS DIFFERENTIABILITY

The objective functions of Eqs. (1)–(3) have discontinuous gradients due to the maximum and minimum operations. This discontinuity often leads to slow convergence during optimization if a gradient-based algorithm is used. To avoid convergence problems, we approximate maximum and minimum operators by smooth log-sum-exp functions [see, e.g., Boyd and Vandenberghe (p. 72 of Ref. 18)]. For a given $\epsilon > 0$, the log-sum-exp maximum function $\mathrm{lse}_{\epsilon}^{\max}$ is a conservative bound on the exact maximum $y_{\max} = \max\{y_1, \ldots, y_k\}$

of some real numbers $y_1, \ldots, y_k$ according to

$$\mathrm{lse}_\epsilon^{\max}(y_1, \ldots, y_k) = \epsilon \log\left(\sum_{i=1}^{k} e^{y_i/\epsilon}\right) = y_{\max} + \epsilon \log\left(\sum_{i=1}^{k} e^{(y_i - y_{\max})/\epsilon}\right),$$

where the rightmost expression provides a numerically stable implementation that avoids overflow. The log-sum-exp minimum is directly analogous to the maximum function. The error of the log-sum-exp approximation is at most $\epsilon \log k$. In the present paper, we use $\epsilon = 10^{-3}$.

Voxelwise worst case has been applied in a nonlinear programming framework without concern to differentiability.[7,8] Also in our experience, discontinuous gradients do not lead to practical problems for the voxelwise method, whereas they do for the other two methods. A plausible explanation is that when the penalty for a given voxel has discontinuous gradient in the voxelwise method, those for many other voxels do not, which reduces the effects of the discontinuities. Also, note that a rigorous, but more computationally expensive, reformulation of maximum operations is to use epigraph formulations, as in Refs. 3, 5, 10, and 14. Each $\max\{y_1, \ldots, y_n\}$ is then replaced by an auxiliary variable t and constraints $t \geq y_1, \ldots, t \geq y_n$, and analogously for minimum operations.

APPENDIX B: WORST CASE PROBABILITY DISTRIBUTIONS

Let x^* be an optimal solution to composite worst case according to Eq. (1). Then, a probability distribution π is a worst case probability distribution to Eq. (1) if x^* is optimal to the expected value minimization

$$\underset{x \geq 0}{\text{minimize}} \quad \mathbb{E}_\pi\left[\sum_{i=1}^{n} w_i f_i(d(x;S))\right], \tag{B1}$$

where S is a random variable picking the scenario s from $\mathcal{S}$ with the probability $\pi(s)$. Similarly, if x^* is optimal to Eq. (2), then a sequence $\{\pi^{(i)}\}_{i=1}^{n}$ of probability distributions is a worst case probability distribution to Eq. (2) if x^* is optimal to

$$\underset{x \geq 0}{\text{minimize}} \quad \sum_{i=1}^{n} w_i \mathbb{E}_{\pi^{(i)}}[f_i(d(x;S))]. \tag{B2}$$

Further, if x^* is optimal to Eq. (3), then a sequence $\{(\pi_v^{\min}, \pi_v^{\max})\}_{v \in \mathcal{V}}$ of pairs of probability distributions is a worst case probability distribution to Eq. (3) if x^* is optimal to

$$\underset{x \geq 0}{\text{minimize}} \quad \sum_{i \in \mathcal{T}} w_i f_i\left(\mathbb{E}_{\pi_v^{\min}}[d_v(x;S)]\right) + \sum_{i \in \mathcal{O}} w_i f_i\left(\mathbb{E}_{\pi_v^{\max}}[d_v(x;S)]\right). \tag{B3}$$

Closed-form expressions for these worst case probability distributions can be established by the Karush-Kuhn-Tucker (KKT) conditions, which are first-order necessary conditions for optimality, cf., e.g., Boyd and Vandenberghe (Sec. 5.5.3 of Ref. 18). The equations that constitute the KKT conditions are identical for the worst case formulations (1)–(3) and their expected value counterparts (B1)–(B3), except for a term that is given by the gradient of the objective function. The worst case formulations are consequently equivalent to their expected value counterparts if the expectation is taken with respect to a probability distribution such that the gradients of the two formulations are equal. The gradient of composite worst case's objective function is under the log-sum-exp approximation given by

$$\left(\sum_{s \in \mathcal{S}} e^{f(d(x;s))/\epsilon}\right)^{-1} \sum_{s \in \mathcal{S}} e^{f(d(x;s))/\epsilon} \nabla_x f(d(x;s)),$$

where $f = \sum_{i=1}^{n} w_i f_i$, and that of its expected value counterpart, problem (B1), by

$$\sum_{s \in \mathcal{S}} \pi(s) \nabla_x f(d(x;s)).$$

These gradients are identical if

$$\pi(s) = \left(\sum_{s \in \mathcal{S}} e^{f(d(x;s))/\epsilon}\right)^{-1} e^{f(d(x;s))/\epsilon}, \tag{B4}$$

which therefore defines the worst case probability distribution to Eq. (1) if x is optimal to the log-sum-exp approximated version of this formulation. By completely analogous analysis, the worst case probability distributions $\{\pi^{(i)}\}_{i=1}^{n}$ and $\{(\pi_v^{\min}, \pi_v^{\max}\}_{v \in \mathcal{V}}$ associated with, respectively, objectivewise and voxelwise worst case are given by expressions that are identical to Eq. (B4) except that $f(d(x;s))$ is replaced by $f_i(d(x;s))$ and $-d_v(x;s)$ or $d_v(x;s)$, respectively.

Observe that π according to Eq. (B4) is non-negative, but need not sum to unity. This distribution therefore requires normalization to define a probability. Such normalization corresponds to multiplying the objective function by a positive scalar and does not affect the optimal solution.

[a]Electronic mail: albin.fredriksson@raysearchlabs.com

[1]A. Lomax, "Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: The potential effects of calculational uncertainties," Phys. Med. Biol. **53**(4), 1027–1042 (2008).
[2]A. Lomax, "Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: The potential effects of inter-fraction and inter-field motions," Phys. Med. Biol. **53**(4), 1043–1056 (2008).
[3]W. Chen, J. Unkelbach, A. Trofimov, T. Madden, H. Kooy, T. Bortfeld, and D. Craft, "Including robustness in multi-criteria optimization for intensity-modulated proton therapy," Phys. Med. Biol. **57**(3), 591–608 (2012).
[4]A. Fredriksson, "A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization," Med. Phys. **39**(8), 5169–5181 (2012).
[5]A. Fredriksson, A. Forsgren, and B. Hårdemark, "Minimax optimization for handling range and setup uncertainties in proton therapy," Med. Phys. **38**(3), 1672–1684 (2011).
[6]T. Inaniwa, N. Kanematsu, T. Furukawa, and A. Hasegawa, "A robust algorithm of intensity modulated proton therapy for critical tissue sparing and target coverage," Phys. Med. Biol. **56**(15), 4749–4770 (2011).
[7]W. Liu, X. Zhang, Y. Li, and R. Mohan, "Robust optimization of intensity modulated proton therapy," Med. Phys. **39**(2), 1079–1091 (2012).
[8]D. Pflugfelder, J. Wilkens, and U. Oelfke, "Worst case optimization: A method to account for uncertainties in the optimization of intensity modulated proton therapy," Phys. Med. Biol. **53**(6), 1689–1700 (2008).

[9]J. Unkelbach, T. Bortfeld, B. Martin, and M. Soukup, "Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning," Med. Phys. **36**(1), 149–163 (2009).
[10]J. Unkelbach, T. Chan, and T. Bortfeld, "Accounting for range uncertainties in the optimization of intensity modulated proton therapy," Phys. Med. Biol. **52**(10), 2755–2773 (2007).
[11]ICRU, "Geometric terms, and dose and dose-volume definitions," *ICRU Report 78: Prescribing, Recording, and Reporting Proton-Beam Therapy* (Oxford University Press, Oxford, UK, 2007), pp. 83–94.
[12]M. Stuschke, A. Kaiser, C. Pöttgen, W. Lübcke, and J. Farr, "Potentials of robust intensity modulated scanning proton plans for locally advanced lung cancer in comparison to intensity modulated photon plans," Radiother. Oncol. **104**(1), 45–51 (2012).
[13]M. Casiraghi, F. Albertini, and A. Lomax, "Advantages and limitations of the 'worst case scenario' approach in IMPT treatment planning," Phys. Med. Biol. **58**(5), 1323–1339 (2013).
[14]T. Chan, "Optimization under uncertainty in radiation therapy," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2007.
[15]U. Oelfke and T. Bortfeld, "Inverse planning for photon and proton beams," Med. Dosim. **26**(2), 113–124 (2001).
[16]A. Lomax, E. Pedroni, H. Rutz, and G. Goitein, "The clinical potential of intensity modulated proton therapy," Z. Med. Phys. **14**(3), 147–152 (2004).
[17]T. Chan, J. Tsitsiklis, and T. Bortfeld, "Optimal margin and edge-enhanced intensity maps in the presence of motion and uncertainty," Phys. Med. Biol. **55**(2), 515–533 (2010).
[18]S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, UK, 2004).