Scalable Optimization Methods for Incorporating Spatiotemporal Fractionation into Intensity-Modulated Radiotherapy Planning

Ali Adibi and Ehsan Salari
Department of Industrial, Systems, and Manufacturing Engineering, Wichita State University, Wichita, KS , USA

It has been recently shown that an additional therapeutic gain may be achieved if a radiotherapy plan is altered over the treatment course using a new treatment paradigm referred to in the literature as spatiotemporal fractionation. Due to the non-convex and large-scale nature of the corresponding treatment-plan optimization problem, the extent of the potential therapeutic gain that may be achieved from spatiotemporal fractionation has been investigated using stylized cancer cases to circumvent the arising computational challenges. This research aims at developing scalable optimization methods to obtain high-quality spatiotemporally fractionated plans with optimality bounds for clinical cancer cases. In particular, the treatment-planning problem is formulated as a quadratically constrained quadratic program and is solved to local optimality using a constraint-generation approach, where each subproblem is solved using sequential linear/quadratic programming methods. To obtain optimality bounds, cutting-plane and column-generation methods are combined to solve the Lagrangian relaxation of the formulation. The performance of the developed methods are tested on de-identified clinical liver and prostate cancer cases. Results show that the proposed method is capable of achieving local-optimal spatiotemporally fractionated plans with an optimality gap of around 10-12% for cancer cases tested in this study.

Key words: radiotherapy planning; spatiotemporal fractionation; quadratically constrained quadratic programming

1. Introduction

The main goal of radiotherapy is to deliver a therapeutic dose of radiation to the clinical target volume in order to eradicate cancer cells while sparing the surrounding healthy tissue to the largest extent possible. As the radiation beam passes through the patient's body, it damages both cancerous and healthy cells along its path; thus, radiotherapy plans must be carefully designed to maximize the chance of tumor eradication while minimizing the risk of normal-tissue side effects resulting from the treatment (Romeijn and Dempsey 2008). *Intensity-modulated radiotherapy* (IMRT) is one of the most common forms of photon

therapy, which uses a *multi-leaf collimator* (MLC) system to modulate the shape and intensity of radiation beams (Bortfeld 2006). In IMRT, each beam is discretized into smaller beams, called *beamlets*, whose intensities are individually adjustable by using the MLC. The goal of IMRT planning is to find the optimal beamlet intensities at each incident beam, called *fluence map*, in order to yield a dose distribution that delivers the prescribed dose to the target volume and at the same time preserves the functionality of nearby normal tissue and *critical* structures. This is achieved through formulating and solving a mathematical optimization problem known as *fluence-map optimization* (FMO).

IMRT plans are typically delivered in daily fractions over the course of one to several weeks, where a fraction of the prescribed dose is delivered at every treatment session. This is motivated by the *fractionation effect*, which suggests that cells can survive a larger total radiation dose if the radiation is delivered in daily fractions. Though the fractionation effect is beneficial to preserving normal tissue, it may adversely impact the fraction of tumor cells killed. However, because the fractionation effect in a majority of tumors is smaller than the normal tissue, fractionated radiotherapy is routinely used in the clinic to safely deliver larger prescription dose values that would not be possible otherwise. A fractionation scheme is characterized by the number of treatment sessions and the dose delivered per fraction. The concept of *biologically effective dose* (BED) is clinically used to calculate iso-effective fractionation schemes (Fowler 1989, 2010). Specifically, two fractionation schemes with identical BED values are expected to cause the same biological damage in a structure, irrespective of the number of fractions and the dose delivered per fraction.

Traditionally, the fractionation decision and the FMO problem have been handled independently. The solution to the FMO problem determines the set of fluence maps for the incident beams and the resulting spatial dose distribution. The fractionation decision then involves finding a suitable fractionation schedule to deliver the given spatial dose distribution. The radiation oncologist typically determines the choice of the fractionation scheme based on clinical trails. Relatively recently, several studies have shown the dependence of the optimal fractionation scheme on the spatial dose distribution (Mizuta et al. 2012, Unkelbach et al. 2012, Yang and Xing 2005, Saberian et al. 2015). Given a fixed spatial dose distribution, those studies model the fractionation decision as a mathematical optimization problem with the goal of optimizing the BED delivered to the target volume and normal tissue. They have shown that when the spatial dose distribution is fixed and

given, a *uniform* fractionation scheme with equal dosage per fraction is optimal, where the number of fractions depends on the BED parameter values for the target volume and critical structures. Saberian et al. (2017) proposed an integrated planning approach to simultaneously optimize the spatial dose distribution and the fractionation length while assuming a uniform fractionation scheme. Ajdari and Ghate (2016b) extended this integrated planning approach by developing a robust counterpart to address the uncertainty associated with the BED parameter values. More recently, several studies have focused on developing adaptive strategies using stochastic control techniques to re-optimize the spatial dose distribution (Saberian et al. 2016, Ajdari et al. 2020, Ten Eikelder et al. 2020) and the fractionation scheme (Ajdari et al. 2018, 2019) in response to changes in imaging bio-markers tracked during the course of treatment.

In uniform fractionation, the total dose received by a structure is divided into equal number of fractions for both the target volume and the normal tissue. However, it is desired to deliver the dose to the target volume in as few fractions as possible to reduce the fractionation effect, thus killing more tumor cells. On the other hand, the dose received by the normal tissue should be ideally divided into as many fractions as possible to exploit the fractionation effect in the normal tissue to the largest extent possible. *Spatiotemporal fractionation* is a treatment paradigm aiming at selective reduction of the fractionation effect in the target volume by altering the spatial dose distribution across treatment sessions. In spatiotemporal fractionation, fraction-dependent dose distributions are used to irradiate a specific subregion of the target volume at each treatment session while maintaining a relatively uniform dose in the normal tissue across fractions. In that case, a majority of the prescription dose for each target subregion will be delivered in a single or a few fractions, thus reducing the fractionation effect in the target. Meanwhile, critical structures receive a relatively uniform dose from each fraction-dependent dose distribution, thus benefiting from the fractionation effect.

To design spatiotemporally fractionated radiotherapy plans, the fractionation effect is integrated into the FMO problem by directly optimizing for the BED distribution in the target volume and the surrounding critical structures. This is achieved by formulating and solving the FMO problem in terms of fraction-dependent fluence maps using BED-based treatment-plan evaluation criteria. The resulting formulation is non-convex in general and

larger in scale compared to the conventional FMO problem and thus is computationally more challenging to solve.

Several studies by Unkelbach and colleagues have demonstrated proof-of-concept for the potential therapeutic gain achievable from spatiotemporal fractionation in different treatment sites and modalities (Unkelbach and Papp 2015, Unkelbach et al. 2016, 2017). In those studies, the FMO problem was formulated with the goal of minimizing the average BED delivered to a specific subset of critical structures while constraining BED-based evaluation criteria in the target volume and other structures to the values obtained by uniform fractionation. They employed an L-BFGS Quasi-Newton method along with augmented Lagrangian method to locally solve the formulation. The results show a reduction of 10-20% in the mean BED delivered to the specified subset of critical structures while maintaining the same quality of BED distributions in the target volume and other structures compared to uniform fractionation. Ajdari and Ghate (2016a) proposed a formulation aiming at maximizing the BED equivalence of the average dose delivered to the target volume across fractions while constraining the mean or maximum BED in different critical structures. They developed a heuristic solution approach using model predictive control to obtain local solutions. They reported a 2-5% gain in tumor BED compared to uniformly fractionated plans for head-and-neck test cases.

More recent studies on spatiotemporal fractionation address the non-convexity of the problem by proposing methods to obtain optimality bounds on the quality of the solutions found using local optimization (Gaddy et al. 2018, Adibi and Salari 2018). Specifically, the FMO problem was formulated as a non-convex *quadratically constrained quadratic programming* (QCQP) problem, which is locally solved using convex optimization techniques. Optimality bounds on the obtained local solutions were then provided by solving a convex relaxation based on *semi-definite programming* (SDP). However, the non-convex QCQP formulation and its SDP relaxation are large scale for clinical cancer cases and thus are computationally prohibitive to solve using general-purpose solvers. Therefore, those studies use stylized cancer cases composed of two-dimensional CT slices for proof-of-concept illustration.

In this study, we aim to extend prior work in this area by developing scalable solution methods to find local solutions and optimality bounds for the spatiotemporal fractionation problem for clinical cancer cases. Optimality bounds provide valuable insight into the

extent of the potential gain achievable from spatiotemporal fractionation, which may be used to identify cancer cases that may benefit the most from spatiotemporal fractionation. To that end, we reformulate the spatiotemporal fractionation problem as a QCQP problem and develop a constraint-generation method to handle the large number of quadratic constraints that arise in clinical cases. At each iteration, a relaxed problem with only a subset of constraints is solved and the most violated constraint(s) are then identified and added sequentially. Sequential linear/quadratic programming methods are used to solve the relaxed problem at each iteration efficiently. To obtain optimality bounds on the QCQP reformulation, we solve the Lagrangian relaxation of the QCQP problem by reformulating it as an SDP problem and solving it using a cutting-plane method embedded in a column-generation framework. The computational performance of the proposed solution approaches are tested on de-identified clinical cancer cases. The main contributions of the study are as follows:

- An optimization framework is proposed to obtain spatiotemporally fractionated radiotherapy plans for clinical cases, aimed at altering the spatial dose distribution during the course of the treatment.
- A tailored constraint-generation method is developed to find local solutions to large-scale QCQP formulation arising in spatiotemporal fractionation.
- A column-generation method is developed to solve the Lagrangian relaxation of the QCQP problem in order to obtain optimality bounds for spatiotemporal fractionation.

The remainder of this paper is organized as follows: In Section 2, we present the notation and a BED-based formulation for spatiotemporally fractionated IMRT planning. In Sections 3 and 4, we develop tailored solution methods to obtain local optimal solutions along with optimality bounds. Section 5 presents the numerical results of applying the proposed spatiotemporal fractionation approach to de-identified clinical cancer cases. A discussion on the insights obtained from the results is provided in Section 6. Finally, Section 7 summarizes the paper and highlights future research directions.

2. Problem Formulation

In this section, we first introduce the notation and then incorporate spatiotemporal fractionation into the FMO problem. The resulting treatment-plan optimization problem is then reformulated as a QCQP problem, which serves as the basis for our solution method development.

Given a prespecified set of beam angles, each beam is discretized into a set of beamlets indexed by $i \in \mathcal{I}$ the intensities of which are individually adjustable. We let x_i be the intensity of beamlet $i \in \mathcal{I}$. Let $\mathcal{T}$ and $\mathcal{C}$ denote the set of target and critical structures, respectively. Additionally, we let $\mathcal{S}$ denote the set of all relevant structures with $\mathcal{S} = \mathcal{T} \cup \mathcal{C}$. To evaluate the dose delivered to each structure by the IMRT plan, we conceptually discretize the relevant anatomy into a fine grid of three-dimensional cubes, which are called *voxels*. The set of voxels in structure $s \in \mathcal{S}$ are indexed by $v \in \mathcal{V}_s$. Also, we let $\mathcal{V} = \bigcup_{s \in \mathcal{S}} \mathcal{V}_s$ be the set of all voxels in the relevant anatomy. The dose (measured in Gy) delivered to voxel $v \in \mathcal{V}$, denoted by d_v, can be approximated as

$$d_v = \sum_{i \in \mathcal{I}} D_{iv} x_i, \tag{1}$$

where D_{iv} is the *dose-deposition coefficient* measuring the amount of deposited dose in voxel v from beamlet i under unit intensity. The conventional FMO problem aims at optimizing the spatial distribution of the *physical* dose calculated using equation (1).

In spatiotemporal fractionation, we aim at varying the dose received by each voxel $v \in \mathcal{V}$ over fractions. In that case, one has to consider the impact of dose variation on the fractionation effect within each voxel. More specifically, consider a fractionation scheme with a pre-specified number of sessions indexed by $n \in \mathcal{N}$, where a physical dose of d_{nv} is delivered to voxel v in fraction n. To account for the fractionation effect, we calculate the BED in voxel v, denoted by b_v, as

$$b_v = \sum_{n \in \mathcal{N}} d_{nv} \left(1 + \frac{d_{nv}}{[\alpha/\beta]}\right), \tag{2}$$

which is known as the *biological* dose. In the equation above, $[\alpha/\beta]$ (measured in Gy) is a structure-specific parameter and describes the degree of sensitivity of a structure to fractionation (Beyzadeoglu et al. 2010). For the target volume, the BED model may be extended to include an additional term to account for tumor re-population. However, assuming an exponential tumor growth, if the total number of treatment fractions (i.e., $|\mathcal{N}|$) is fixed, then the re-population term will be constant (Fowler 2010). We note that there are several extensions of the BED model to account for other biological effects such as accelerated re-population, re-oxygenation of hypoxic tumors, and drug radio-sensitization (Jones and Dale 2019). In this study, we focus on the basic fractionation effect, which is well established and widely used to guide fractionation decisions in clinical practice.

A collection of structure-based evaluation criteria, denoted by $F_s : \mathbb{R}^{|\mathcal{V}|} \to \mathbb{R}$ $(s \in \mathcal{S})$, are used to measure the quality of the biological dose distribution within each structure. Without the loss of generality, we assume smaller values of the evaluation criteria are preferred over larger ones. We can then formulate the spatiotemporal fractionation problem in terms of the fraction-dependent beamlet intensities as

$$\min \sum_{s \in \mathcal{C}} \gamma_s F_s(b) \tag{3}$$

subject to

$$b_v \geq \tau_s \qquad v \in \mathcal{V}_s, s \in \mathcal{T} \tag{4}$$

$$b_v = \sum_{n \in \mathcal{N}} d_{nv} \left(1 + \frac{d_{nv}}{[\alpha/\beta]_s}\right) \qquad v \in \mathcal{V}_s, s \in \mathcal{S} \tag{5}$$

$$d_{nv} = \sum_{i \in \mathcal{I}} D_{iv} x_{ni} \qquad n \in \mathcal{N}, v \in \mathcal{V} \tag{6}$$

$$x_{ni} \geq 0 \qquad n \in \mathcal{N}, i \in \mathcal{I} \tag{7}$$

In the formulation above, using non-negative weight factors γ_s $(s \in \mathcal{C})$, the objective function minimizes the weighted sum of evaluation criteria over all critical structures. If protection of a target structure from large BED values is desired, then that structure can also be simultaneously considered as a critical structure to participate in the objective function. The constraints in (4) ensure the delivery of at least the prescribed biological dose τ_s to every voxel in the target structure $s \in \mathcal{T}$. Equations (5) and (6) measure the biological dose and the fraction-dependent physical dose at each voxel, respectively. Lastly, the constraints in (7) enforce non-negativity on fraction-dependent beamlet intensities.

The generalized mean of the dose distribution has been traditionally used to evaluate the quality of the physical dose distribution in different structures (Niemierko 1999, Henríquez and Castrillón 2011). An approximation of the generalized mean using the convex combination of the maximum and average dose in a structure, known as *max-mean* approximation, has also been used in conventional IMRT planning (Thieke et al. 2003, Long et al. 2012, Craft et al. 2005). In our prior study, we employed the max-mean approximation to the BED distribution, called *max-mean BED*, as a treatment-plan evaluation criterion for spatiotemporal fractionation (Adibi and Salari 2018). In this study, we use structure-based one-sided linear penalties to penalize any positive deviations of the max-mean BED from

a user-specified threshold. More formally, the plan evaluation criteria $F_s\,(s \in \mathcal{C})$ are defined as

$$F_s(b) = \left(\kappa_s \max_{v \in \mathcal{V}_s} b_v + (1-\kappa_s)\,\frac{1}{|\mathcal{V}_s|} \sum_{v \in \mathcal{V}_s} b_v - \tau_s \right)_+, \tag{8}$$

where $(\cdot)_+$ represents the ramp function, $\kappa_s \in [0,1]$ is structure-specific convex-combination coefficient, and τ_s is structure-specific user-specified threshold. F_s can be represented in a linear form by introducing auxiliary variables u_s and w_s corresponding to the maximum and mean of BED distribution in structure s, respectively. This linear representation also requires the addition of the following auxiliary constraints: $b_v \leq u_s\,(v \in \mathcal{V}_s)$ and $\frac{1}{|\mathcal{V}_s|}\sum_{v \in \mathcal{V}_s} b_v \leq w_s$. The problem can then be reformulated as

$$\min \sum_{s \in \mathcal{C}} \gamma_s p_s \tag{9}$$

subject to

$$(4) - (7)$$

$$b_v \leq u_s \qquad \forall v \in \mathcal{V}_s, s \in \mathcal{C} \tag{10}$$

$$\frac{1}{|\mathcal{V}_s|} \sum_{v \in \mathcal{V}_s} b_v \leq w_s \qquad \forall s \in \mathcal{C} \tag{11}$$

$$p_s \geq (\kappa_s u_s + (1-\kappa_s)\,w_s) - \tau_s \qquad \forall s \in \mathcal{C} \tag{12}$$

$$p_s \geq 0 \qquad \forall s \in \mathcal{C} \tag{13}$$

In the rest of this section, we first reformulate the spatiotemporal fractionation problem in (9)–(13) as a QCQP problem, which serves as the basis for our solution method development. We highlight features of this reformulation that must inform the development of scalable solution methods. We then consider the special case of uniform fractionation and show that in that case the spatiotemporal fractionation problem can be reformulated as a convex QCQP problem. Finally, we consider a variant of the proposed formulation to maximize the potential gain achievable from spatiotemporal fractionation in a single dose-limiting critical structure.

2.1. QCQP Reformulation

Let column vector $D_{\cdot v} = (D_{iv} : i \in \mathcal{I})^\top$ represent the dose-deposition coefficients associated with voxel $v \in \mathcal{V}$. Additionally, let row vector $x_{n\cdot} = (x_{ni} : i \in \mathcal{I})$ represent the beamlet-intensity profile at fraction $n \in \mathcal{N}$. Using equations (5) and (6), we substitute b_v in constraints (4), (10), and (11) with a quadratic expression of the form $\sum_{n \in \mathcal{N}} (x_{n\cdot} D_{\cdot v})^2 / [\alpha/\beta] +$

$x_{n\cdot}D_{\cdot v}$ for voxel $v \in \mathcal{V}$ and thus drop equations (5) and (6) from the formulation. Let column vector y be the stacked vector of all remaining variables in the formulation, namely, fraction-dependent beamlet-intensity vectors $x_{n\cdot}\,(n \in \mathcal{N})$ and auxiliary variables $u_s, w_s, p_s\,(s \in \mathcal{C})$. The spatiotemporal fractionation problem can then be reformulated in terms of vector y as

$$\min\ q_0^\top y \tag{14}$$

subject to

$$y^\top Q_v y + q_v^\top y + \tau_s \leq 0 \qquad \forall v \in \mathcal{V}_s, s \in \mathcal{T} \tag{15}$$

$$y^\top Q_v y + q_v^\top y \leq 0 \qquad \forall v \in \mathcal{V}_s, s \in \mathcal{C} \tag{16}$$

$$y^\top \overline{Q}_s y + \overline{q}_s^\top y \leq 0 \qquad \forall s \in \mathcal{C} \tag{17}$$

$$c_s^\top y - \tau_s \leq 0 \qquad \forall s \in \mathcal{C} \tag{18}$$

$$-y \leq 0. \tag{19}$$

In the reformulation above, q_0 is the vector of objective function coefficients with all elements equal to zero except for those corresponding to auxiliary variables $p_s\,(s \in \mathcal{C})$, which are equal to γ_s. The constraints in (15) are a reformulation of (4), where Q_v is block diagonal with $|\mathcal{N}|$ identical blocks of the form $-D_{\cdot v}D_{\cdot v}^\top/[\alpha/\beta]$ and a block of zeros corresponding to beamlet-intensity profiles and auxiliary variables, respectively. Similarly, q_v is a stacked column vector composed of $|\mathcal{N}|$ identical vectors of the form $-D_{\cdot v}$ and a vector of zeros. The constraints in (16) are a reformulation of (10), where Q_v is block diagonal with $|\mathcal{N}|$ blocks of the form $D_{\cdot v}D_{\cdot v}^\top/[\alpha/\beta]$ and a block of zeros. Additionally, q_v is a stacked column vector consisting of $|\mathcal{N}|$ identical vectors of the form $D_{\cdot v}$ and a unit vector with one non-zero element of -1 corresponding to auxiliary variable u_s. The constraints in (17) are a reformulation of (11), where $\overline{Q}_s = \frac{1}{|\mathcal{V}_s|}\sum_{v\in\mathcal{V}_s} Q_v$ and $\overline{q}_s$ is a stacked column vector consisting of $|\mathcal{N}|$ vectors of the form $\frac{1}{|\mathcal{V}_s|}\sum_{v\in\mathcal{V}_s} D_{\cdot v}$ and a unit vector with one non-zero element of -1 corresponding to auxiliary variable w_s. The constraints in (18) are a reformulation of (12) in which c_s is a column vector with all elements equal to zero except for three elements corresponding to auxiliary variables u_s, w_s, and p_s, which are κ_s, $1-\kappa_s$, and -1, respectively.

Because the objective function in (14) is linear, the formulation above may be considered as belonging to the class of *quadratically constrained linear programming* (QCLP) problems. QCLP is a special class of QCQP where the objective function is linear with a zero

Hessian matrix. On the other hand, any given QCQP problem may be reformulated as a QCLP problem by transforming the objective function into a quadratic constraint using an auxiliary variable. The QCLP reformulation then attempts to minimize the value of the added auxiliary variable as the new linear objective function. Hence, QCQP solution methods are applicable to QCLP and vice versa. Therefore, in this study, we consider the formulation (14)–(19) as belonging to the broader class of QCQP problems.

There are two major challenges in solving the QCQP reformulation above. First, the formulation is non-convex in general due to the negative semi-definiteness of matrices Q_v for target voxels in (15). Note that a block diagonal matrix is negative (positive) semi-definite if and only if all its diagonal blocks are negative (positive) semi-definite. Second, the formulation is large-scale, with the number of variables on the order of number of beamlets times the number of fractions and the number of quadratic constraints on the order of number of voxels. Non-convex QCQP problems are NP-hard and difficult to solve, and usually there is no guarantee to reach global optimality (Park and Boyd 2017, Boyd and Vandenberghe 2004). General-purpose convex-optimization solvers, such as Ipopt (Wächter and Biegler 2006) and Knitro (Byrd et al. 2006), can find local optimal solutions for non-convex QCQP problems. However, our observation in the case of the QCQP problem considered in this study is that those solvers take a long time to reach local solutions and, depending on the number of voxels considered, computational-time and memory requirements may be prohibitive for large clinical cases. In Section 3, we develop customized solution methods to efficiently solve the QCQP reformulation by finding local optimal solutions as well as tight optimality bounds.

2.2. Special Case of Uniform Fractionation

In this section, we consider a special case of the spatiotemporal fractionation problem with the additional constraint that the spatial dose distribution is fraction-invariant, which requires the use of identical beamlet-intensity profiles across fractions. In that case, our spatiotemporal fractionation problem is equivalent to the traditional BED-based planning problem utilizing a uniform fractionation scheme. Here, we discuss the steps to obtain a convex QCQP reformulation of the spatiotemporal fractionation problem in (9)–(13) for this special case.

The non-convexity of the QCQP reformulation presented in Section 2.1 stems from the constraints in (4) that enforce the minimum prescribed BED in target voxels. Under

uniform fractionation, we can rewrite those constraints in terms of physical dose. More specifically, under uniform fractionation, we can replace fraction-dependent dose variables $d_{nv}\,(n \in \mathcal{N})$ in (2) with fraction-invariant variable d_v. Next, by substituting b_v from (2) and bringing the right-hand side to the left, we can write the constraints in (4) as

$$|\mathcal{N}|d_v\left(1+\frac{d_v}{[\alpha/\beta]}\right)-\tau_s \geq 0.$$

The left-hand side of the inequality above is a convex quadratic function of d_v with two real roots, and thus is non-negative if and only if

$$d_v \geq \frac{-|\mathcal{N}|+\sqrt{|\mathcal{N}|^2+4|\mathcal{N}|\tau_s/[\alpha/\beta]}}{2|\mathcal{N}|/[\alpha/\beta]}. \tag{20}$$

By replacing constraint (4) with the equivalent form (20) in the spatiotemporal fractionation problem (9)–(13) and applying the reformulation steps described in Section 2.1, we obtain a convex QCQP reformulation for the special case of uniform fractionation.

2.3. Special Case of Considering a Major Dose-limiting Critical Structure

In the proposed spatiotemporal fractionation formulation (9)–(13), the objective function aims at minimizing the weighted sum of max-mean BED penalties in all critical structures, potentially distributing the spatiotemporal fractionation gain across different critical structures. However, it may be clinically desired to focus only on a major dose-limiting critical structure in a treatment site (e.g., healthy liver in liver cancer and rectum in prostate cancer), and use spatiotemporal fractionation as a means to further reduce the deposited BED in that particular structure as much as possible. To accommodate this case, we consider a variant of the proposed formulation in which the objective function only consists of the max-mean BED penalty associated with the dose-limiting critical structure while the max-mean BED penalty in all other critical structures are constrained to the values obtained from uniform fractionation. This requires imposing additional linear constraints of the form $p_s \leq \hat{p}_s$ for the remaining critical structures, where $\hat{p}_s$ is the penalty value for critical structure s in the uniformly fractionated plan. This problem can also be reformulated as a QCQP problem following the steps in Section 2.1.

3. Solution Approach

One of the major difficulties of solving the QCQP formulation is the large number of quadratic constraints associated with all voxels. In this section, we develop a constraint-generation method that starts by solving a relaxed problem that includes only a subset

of those quadratic constraints. The solution to the relaxed problem is then substituted in the remaining constraints to identify possible violations. A subset of constraints with the largest amount of violation are then identified and added to the relaxed problem. These steps are repeated until a termination condition (i.e., no significant violation in any of the relaxed constraints) is met.

3.1. Constraint-Generation Method

To apply the constraint-generation method to our QCQP formulation, we start by considering a relaxed problem obtained by including only some of the constraints in (16) that correspond to a subset of voxels within each critical structure, denoted by $\bar{\mathcal{V}}_s\,(s \in \mathcal{C})$. All other constraints in (15) and (17)–(19) are included. At each iteration, we solve the relaxed problem to local optimality using a customized solution method presented in Section 3.3. Using the obtained fraction-dependent fluence maps, we calculate the BED values for all voxels in a given structure and, in turn, the amount of violation of the corresponding constraint in (16). In particular, to calculate constraint violations, we use the equivalent constraint (10). We then add a subset of voxels with the largest violations in each critical structure, if any, to the relaxed problem and proceed to the next iteration. Algorithm 1 summarizes the steps taken by the constraint-generation method to find local solutions to the QCQP problem.

Algorithm 1 Constraint-generation method to solve QCQP formulation

1: Initialize subset of voxels $\bar{\mathcal{V}}_s \subseteq \mathcal{V}_s\,(s \in \mathcal{C})$
2: **while** termination condition not met **do**
3: Solve the relaxed QCQP formulation (14)–(19) with $\bar{\mathcal{V}}_s\,(s \in \mathcal{C})$ to obtain $\hat{y}$
4: Calculate BED distribution associated with $\hat{y}$
5: Calculate violation of constraint (10) for voxels in $\mathcal{V}_s \setminus \bar{\mathcal{V}}_s$ for structure $s \in \mathcal{C}$
6: Update $\bar{\mathcal{V}}_s$ by adding voxels with largest violations, if any, for structure $s \in \mathcal{C}$
7: **end while**
8: Return $y^* \leftarrow \hat{y}$

3.2. Sequential Linear Programming

The relaxed QCQP formulation in the constraint-generation framework above is nonconvex in general due to the presence of quadratic constraints associated with target voxels. General-purpose solvers are not able to provide high-quality solutions for the relaxed problem, particularly as the number of voxels considered in the relaxed problem increases.

Hence, we develop a *sequential linear programming* (SLP) method to solve the relaxed problem efficiently. At each iteration, the SLP method approximates the QCQP problem using the first-order Taylor expansion of the nonlinear constraints around the current solution and solves the resulting linear program to update the solution. In the rest of this section, we provide more details on the steps of our SLP implementation, which is based on the *penalty successive linear programming* algorithm described in Bazaraa et al. (2013).

Consider the relaxed QCQP formulation (14)–(19) in which for each critical structure $s \in \mathcal{C}$ in (16), the set of voxels $\mathcal{V}_s$ is substituted with a subset $\bar{\mathcal{V}}_s$. At each iteration, the SLP method constructs a linear approximation of the QCQP formulation by linearizing all quadratic constraints in (15)–(17) around the current solution $y^{(k)}$ and relaxing them by introducing an ℓ_1-norm penalty in the objective function. Additionally, a trust-region constraint is imposed to ensure boundedness of the LP problem as well as convergence of the SLP method. Hence, to obtain an updated solution, we solve an LP problem of the form

$$\min\ q_0^\top y + \mu_1 \sum_{s\in\mathcal{S}} \sum_{v\in\bar{\mathcal{V}}_s} h_v + \mu_2 \sum_{s\in\mathcal{C}} \bar{h}_s$$

subject to

$$\left(2y^{(k)\top} Q_v + q_v^\top\right) y - y^{(k)\top} Q_v y^{(k)} + \tau_s \le h_v \qquad \forall v \in \mathcal{V}_s, s \in \mathcal{T} \tag{21}$$

$$\left(2y^{(k)\top} Q_v + q_v^\top\right) y - y^{(k)\top} Q_v y^{(k)} \le h_v \qquad \forall v \in \bar{\mathcal{V}}_s, s \in \mathcal{C} \tag{22}$$

$$\left(2y^{(k)\top} \overline{Q}_s + \overline{q}_s^\top\right) y - y^{(k)\top} \overline{Q}_s y^{(k)} \le \bar{h}_s \qquad \forall s \in \mathcal{C} \tag{23}$$

$$c_s^\top y - \tau_s \le 0 \qquad \forall s \in \mathcal{C} \tag{24}$$

$$\left\| y - y^{(k)} \right\|_\infty \le \Delta^{(k)} \tag{25}$$

$$y \ge 0, h \ge 0, \bar{h} \ge 0. \tag{26}$$

Following a standard trust-region procedure, the decision whether to accept or reject the updated solution and the adjustment of the trust-region radius $\Delta^{(k)}$ is made based on the ratio of the actual improvement and the improvement predicted by the linearized model. In particular, let y^* be the optimal solution to the LP problem. The predicted improvement is calculated as the difference between the LP objective values at y^* and $y^{(k)}$. Also, the actual improvement is calculated by using the original quadratic constraints (15)–(17), in lieu of the linearized constraints (21)–(23), to evaluate the ℓ_1-norm penalty at y^* and $y^{(k)}$. If the

improvement ratio is smaller than a user-specified threshold, then the update is rejected, the trust-region radius is decreased, and the step is repeated. Otherwise, the update is accepted, the trust-region radius is adjusted, and SLP proceeds to the next iteration (see Bazaraa et al. (2013) for more details on the adjustment of the trust-region radius).

3.3. Sequential Quadratic Programming

We also develop a *sequential quadratic programming* (SQP) method to solve the relaxed QCQP problem. At each iteration of the SQP method, a convex *quadratic programming* (QP) approximation of the QCQP problem is constructed and solved around the current solution to obtain a search direction. A line-search problem is then solved to find the optimal step length along the search direction. Our SQP implementation, summarized below, is an adaptation of the *merit-function sequential quadratic programming* algorithm described in Bazaraa et al. (2013).

To construct the QP approximations, the quadratic constraints in (15)–(17) are linearized using their first-order Taylor series expansions at the current solution. This leads to a first-order approximation of the relaxed QCQP formulation (14)–(19), where only a subset of voxels $\bar{\mathcal{V}}_s$ for each critical structure $s \in \mathcal{C}$ is considered. The obtained approximation problem is then augmented by second-order information through adding the Hessian of the Lagrangian to the objective function in the form of a quadratic term. However, the Hessian of the Lagrangian may not always be positive semi-definite due to the quadratic constraints associated with target voxels in the QCQP problem, potentially rendering the QP approximation non-convex. To overcome this issue, we employ the *damped* BFGS updating method to approximate the Hessian with a positive-definite matrix at each iteration (Nocedal and Wright 2006). The optimal solution to the resulting convex QP approximation provides a search direction. To obtain the optimal step length, the line-search problem aims at minimizing a merit function to ensure the maximum possible improvement in the objective function along the search direction while maintaining feasibility. The above-mentioned steps are repeated until the termination condition of no significant change in the objective function value is satisfied.

4. Obtaining Optimality Bounds

Thus far, we have presented customized solution methods to find local-optimal solutions to the QCQP formulation of the spatiotemporally fractionated radiotherapy planning problem. However, due to the non-convexity of the formulation, they merely provide an upper

bound on the global optimal solution, and there is no guarantee on the global optimality of the found solutions. To examine the quality of the solutions, we also obtain lower bounds on the global optimal solution of the non-convex QCQP formulation. This is typically achieved by solving a convex relaxation of the QCQP formulation, namely, the SDP relaxation or the Lagrangian relaxation. These two relaxations of the QCQP problem are closely related in that the Lagrangian relaxation is the conic dual of the SDP relaxation. Hence, under a constraint qualification condition (e.g., Slater's condition), the strong duality applies and both relaxations generate the same bounds (d'Aspremont and Boyd 2003). Our initial experiments showed that it is computationally more efficient to solve the Lagrangian relaxation, rather than the SDP relaxation, of the QCQP formulation for clinical cases.

In the rest of this section, we present the Lagrangian relaxation of the QCQP problem and develop a column-generation approach to solve it. At each iteration, the proposed column-generation approach solves a restricted Lagrangian relaxation that includes only a subset of dual variables using a cutting-plane method. The found solution is then used to identify and add new promising variables.

4.1. Lagrangian Relaxation

Lagrangian relaxation is commonly used to compute bounds on the optimal objective value of non-convex problems, where the quality of the bound depends on the optimality gap (Boyd and Vandenberghe 2004). To formulate the Lagrangian relaxation of the QCQP formulation in (14)–(19), we first obtain the Lagrangian. Let $\mathcal{J}$ be the index set of all constraints in (15)–(19) and λ_j ($j \in \mathcal{J}$) be the associated dual variables. We represent the left-hand side of each constraint $j \in \mathcal{J}$ as a quadratic function $H_j(y) = y^\top Q'_j y + q_j'^\top y + r'_j$, where Q'_j, q'_j, and r'_j are the corresponding second-order, first-order, and constant coefficients. The Lagrangian can then be written as

$$\mathcal{L}(y, \lambda) = y^\top \widetilde{Q}(\lambda) y + \widetilde{q}(\lambda)^\top y + \widetilde{r}(\lambda), \tag{27}$$

where $\widetilde{Q}(\lambda)$, $\widetilde{q}(\lambda)$, and $\widetilde{r}(\lambda)$ are defined as

$$\widetilde{Q}(\lambda) \equiv \sum_{j \in \mathcal{J}} \lambda_j Q'_j \tag{28}$$

$$\widetilde{q}(\lambda) \equiv q_0 + \sum_{j \in \mathcal{J}} \lambda_j q'_j \tag{29}$$

$$\widetilde{r}(\lambda) \equiv \sum_{j \in \mathcal{J}} \lambda_j r'_j. \tag{30}$$

Next, the Lagrangian dual function is obtained by finding the infimum of the Lagrangian over vector y as follows (Park and Boyd 2017):

$$\inf_{y} \mathcal{L}(y,\lambda) = \begin{cases} \widetilde{r}(\lambda) - \frac{1}{4}\widetilde{q}(\lambda)^{\top}\widetilde{Q}(\lambda)^{\dagger}\widetilde{q}(\lambda) & \widetilde{Q}(\lambda) \succeq 0; \\ -\infty & \text{otherwise,} \end{cases} \tag{31}$$

where † denotes the Moore-Penrose pseudo inverse, and $\succeq 0$ denotes *positive semi-definiteness* (PSD). Lagrangian relaxation is then formulated as the maximization of the Lagrangian dual function subject to non-negativity of the dual variables.

By applying the Schur complement, we can write the Lagrangian relaxation as an SDP problem as (Park and Boyd 2017)

$$\max \quad \alpha \tag{32}$$

subject to

$$\begin{bmatrix} \widetilde{Q}(\lambda) & (1/2)\widetilde{q}(\lambda) \\ (1/2)\widetilde{q}(\lambda)^{\top} & \widetilde{r}(\lambda) - \alpha \end{bmatrix} \succeq 0 \tag{33}$$

$$\lambda_j \geq 0 \qquad\qquad j \in \mathcal{J} \tag{34}$$

The optimal objective value of the SDP problem above provides a lower bound for the QCQP formulation. However, a major challenge in solving this SDP formulation is the size of the vector λ because of the large number of constraints in the QCQP formulation. General-purpose SDP solvers (e.g., SeDuMi) fail to solve the problem for clinical instances due to computational memory and time requirements. Therefore, we develop a customized solution method for this SDP formulation by combining the column-generation and cutting-plane methods.

4.2. Column-generation Method

To solve the Lagrangian relaxation, we develop a column-generation method where at each iteration a restricted version of the SDP formulation (32)–(34) is solved by including only a subset of dual variables λ_j ($j \in \mathcal{J}$). Given the optimal solution to the restricted Lagrangian relaxation, we then solve the *pricing* problem to identify new promising dual variables to be added to the restricted problem. We repeat the above-mentioned steps until the desired termination condition is met.

At each iteration of the column-generation method, we solve a restricted Lagrangian relaxation with a maximization objective; therefore, the optimal objective value α^{*}_{res} is less

than or equal to the optimal objective value of the full Lagrangian relaxation α^*. Hence, the solution to the restricted Lagrangian relaxation at each iteration provides a lower bound for the optimal solution to the QCQP formulation $q_0^\top y^*$, that is, $\alpha^*_{\text{res}} \leq \alpha^* \leq q_0^\top y^*$. Hence, if desired, the column-generation method may be terminated prior to convergence without compromising the validity of the bound computed. In the remainder of this section, we first develop a solution method for the restricted Lagrangian relaxation, and then formulate and solve the pricing problem.

4.2.1. Cutting-plane method. Consider a restricted Lagrangian relaxation where only a subset of dual variables $\overline{\mathcal{J}} \subseteq \mathcal{J}$ is included. Despite considering only a subset of dual variables, the PSD constraint in (33) involves a large-sized matrix, due to the size of the $Q_v\,(v \in \mathcal{V})$ and $\bar{Q}_s\,(s \in \mathcal{S})$ matrices for clinical cases, which renders the restricted problem computationally prohibitive to solve using general-purpose SDP solvers (e.g., SeDuMi). Hence, we propose a cutting-plane method to solve the restricted problem. Such methods are an alternative to interior-point methods for solving large-scale SDP problems efficiently (Furrer 2009). The SDP formulation of the Lagrangian relaxation presented in (32)–(34) can be expressed as a semi-infinite LP if the PSD constraint in (33) is substituted with infinitely many linear constraints in terms of (λ, α) of the form

$$\zeta^\top \begin{bmatrix} \widetilde{Q}(\lambda) & (1/2)\widetilde{q}(\lambda) \\ (1/2)\widetilde{q}(\lambda)^\top & \widetilde{r}(\lambda) - \alpha \end{bmatrix} \zeta \geq 0 \qquad \forall \zeta \in \mathbb{R}^{m+1}, \tag{35}$$

where m is the dimension of vector y. Cutting-plane methods for SDP aim at solving this equivalent semi-infinite LP formulation by solving a sequence of relaxed LPs that only contain a finite number of the constraints in (35). In other words, the relaxed LP is obtained by substituting the constraints in (35) with

$$\zeta_\ell^\top \begin{bmatrix} \widetilde{Q}(\lambda) & (1/2)\widetilde{q}(\lambda) \\ (1/2)\widetilde{q}(\lambda)^\top & \widetilde{r}(\lambda) - \alpha \end{bmatrix} \zeta_\ell \geq 0 \qquad \forall \ell \in \mathcal{E},$$

which are referred to as *linear cuts* and are indexed by $\ell \in \mathcal{E}$. This relaxed LP can be written in terms of the vector of variables $z = \left[\lambda^\top \alpha\right]^\top$ as

$$\min\ f^\top z \tag{36}$$

subject to

$$Az \geq e \tag{37}$$

$$z \geq 0 \tag{38}$$

in which $f = [0 \dots 0 \ -1]^\top$, and the element at the ℓ-th row ($\ell \in \mathcal{E}$) and j-th column ($j \in \bar{\mathcal{J}}$) of the coefficient matrix A is calculated as

$$A_{\ell j} = \left\langle \begin{bmatrix} Q'_j & (1/2)q'_j \\ (1/2)q'^{\top}_j & r'_j \end{bmatrix}, \zeta_\ell \zeta_\ell^\top \right\rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the trace of the product of two matrices. The ℓ-th element in the last column of the coefficient matrix A, corresponding to variable α, is calculated as $-\zeta^2_{\ell,m+1}$. Finally, the right-hand side of the ℓ-th constraint in the relaxed LP is calculated as

$$e_\ell = -\left\langle \begin{bmatrix} O_{m\times m} & (1/2)q_0 \\ (1/2)q_0^\top & 0 \end{bmatrix}, \zeta_\ell \zeta_\ell^\top \right\rangle, \tag{39}$$

where O is the matrix of zeros.

At iteration k of the cutting-plane method, we first obtain the optimal solution $(\lambda^{(k)}, \alpha^{(k)})$ to the relaxed LP and then check for possible violation of the PSD constraint in (33) evaluated at $\left(\lambda^{(k)}, \alpha^{(k)}\right)$. This is done via spectral decomposition of the matrix into eigenvectors and corresponding eigenvalues. If the PSD constraint is not satisfied (i.e., there are negative eigenvalues), then we add a collection of eigenvectors that correspond to the most negative eigenvalues to the set of linear cuts $\mathcal{E}$. The above-mentioned steps are repeated until the PSD constraint is (approximately) restored, that is, all eigenvalues are above a user-specified threshold (i.e., -10^{-6}) (Qualizza et al. 2012).

4.2.2. Pricing problem. Upon termination, the cutting-plane method provides the optimal solution to the restricted Lagrangian relaxation. To check whether the solution to the restricted problem is also optimal for the full Lagrangian relaxation, we investigate whether it satisfies *Karush-Kuhn-Tucker* (KKT) conditions of the SDP formulation (32)–(34). Let Y be the dual matrix associated with the PSD constraint in (33). The KKT conditions can be expressed in terms of the primal-dual solution (λ, α, Y) as follows (Correa and Ramirez C. 2004):

$$\left\langle \begin{bmatrix} Q'_j & (1/2)q'_j \\ (1/2)q'^{\top}_j & r'_j \end{bmatrix}, Y \right\rangle \leq 0 \qquad\qquad j \in \mathcal{J} \tag{40}$$

$$\left\langle \begin{bmatrix} O_{m\times m} & O_{m\times 1} \\ O_{1\times m} & -1 \end{bmatrix}, Y \right\rangle = -1 \tag{41}$$

$$\begin{bmatrix} \widetilde{Q}(\lambda) & (1/2)\widetilde{q}(\lambda) \\ (1/2)\widetilde{q}(\lambda)^\top & \widetilde{r}(\lambda) - \alpha \end{bmatrix} \succeq 0 \tag{42}$$

$$\lambda_j \geq 0 \qquad\qquad j \in \mathcal{J} \tag{43}$$

$$Y \succeq 0 \tag{44}$$

$$\left\langle \begin{bmatrix} \widetilde{Q}(\lambda) & (1/2)\widetilde{q}(\lambda) \\ (1/2)\widetilde{q}(\lambda)^\top & \widetilde{r}(\lambda) - \alpha \end{bmatrix}, Y \right\rangle = 0, \tag{45}$$

where (40)–(41) correspond to the stationarity condition, (42)–(43) correspond to primal feasibility, (44) corresponds to dual feasibility, and (45) corresponds to the complementary slackness condition. We note that in obtaining the KKT conditions above, we use an equivalent representation of the non-negativity constraints in (34) in the form of a PSD constraint on a diagonal matrix, denoted by Λ, composed of $\lambda_j\,(j \in \mathcal{J})$. This equivalent representation is then integrated into the PSD constraint in (33) by appending Λ as an additional diagonal block. KKT conditions (40)–(45) are also applicable to the restricted Lagrangian relaxation if the index set $\mathcal{J}$ is substituted with $\bar{\mathcal{J}}$. Theorem 1 shows that the optimal solution to the relaxed LP in (36)–(38) can be used to construct a primal-dual solution that satisfies the KKT conditions of the restricted Lagrangian relaxation.

THEOREM 1. *Given the optimal solution* $\left(\hat{\lambda}, \hat{\alpha}\right)$ *and the corresponding dual values* $\hat{\pi}$ *obtained from solving the relaxed LP in (36)–(38), the primal-dual solution* $\left(\hat{\lambda}, \hat{\alpha}, \hat{Y}\right)$ *with* $\hat{Y} = \sum_{\ell\in\mathcal{E}} \hat{\pi}_\ell \left(\zeta_\ell \zeta_\ell^\top\right)$ *satisfies the KKT conditions of the restricted Lagrangian relaxation.*

Proof. See the online supplement.

At each iteration of the column-generation method, Theorem 1 provides us with a primal-dual solution $\left(\hat{\lambda}, \hat{\alpha}, \hat{Y}\right)$ that satisfies the KKT conditions of the restricted Lagrangian relaxation. To investigate whether $\left(\hat{\lambda}, \hat{\alpha}, \hat{Y}\right)$ also satisfies the KKT conditions of the full Lagrangian relaxation, all we need to check is whether the condition in (40) is satisfied for all constraints $j \in \mathcal{J}$. To that end, we formulate and solve the pricing problem to find the largest value of the left-hand side of the inequality evaluated at $\hat{Y}$ over all constraints $j \in \mathcal{J}$ as follows:

$$\hat{\varphi} = \max_{j\in\mathcal{J}} \left\langle \begin{bmatrix} Q'_j & (1/2)q'_j \\ (1/2)q_j'^\top & r'_j \end{bmatrix}, \hat{Y} \right\rangle \tag{46}$$

If the optimal objective value of the pricing problem is non-positive, then the primal-dual solution $\left(\hat{\lambda}, \hat{\alpha}, \hat{Y}\right)$ also satisfies the KKT conditions of the full Lagrangian relaxation, and the column-generation method returns $\hat{\alpha}$ as a lower bound on the optimal solution of the QCQP problem. Otherwise, we add a subset of promising dual variables (i.e., dual variables corresponding to constraints with large positive objective values in (46)) to the restricted Lagrangian relaxation and proceed to the next iteration.

Algorithm 2 shows the steps of the proposed column-generation method to solve the Lagrangian relaxation. In particular, we initialize the set $\bar{\mathcal{J}} \subseteq \mathcal{J}$ by including all dual variables associated with constraints (17)–(19). The subset $\bar{\mathcal{J}}$ also includes dual variables associated with some of the constraints in (15)–(16). Moreover, we initialize the set of linear cuts $\mathcal{E}$ using the set of vectors in $\mathbb{R}^{m+1}$ that have exactly two nonzero elements, each equal to ± 1, which are known as *diagonally dominant sum-of-squares* (DSOS) cuts (Ahmadi et al. 2017). Our experimental results show that the use of DSOS cuts for initialization leads to fewer iterations, compared to randomly generated vectors. Finally, to solve the pricing problem, we perform an exhaustive search by evaluating the objective function in (46) for each constraint $j \in \mathcal{J} \setminus \bar{\mathcal{J}}$.

5. Computational Results

In this section, we apply the proposed solution approach to clinical cancer instances in order to evaluate the computational performance and to quantify the potential benefit of spatiotemporal fractionation.

5.1. Cancer Cases

We start by testing the performance of the proposed methods on two de-identified clinical instances selected from the *common optimization for radiation therapy* (CORT) dataset (Craft et al. 2014). The first case is a liver cancer case with one target volume, *planning target volume* (PTV), and five critical structures, namely healthy liver, heart, stomach, skin, and entrance region. A non-coplanar beam arrangement is used for the liver case. The second case is a prostate cancer case with one PTV and three critical structures, namely, rectum, bladder, and body. A coplanar equispaced beam arrangement is used for the prostate case (see the online supplement for more details on specifications of the two test cases). We perform down-sampling by selecting a random subset of voxels from each critical structure to be included in the optimization model in an attempt to reduce the total

Algorithm 2 Column-generation method to solve Lagrangian relaxation

1: Set $\epsilon > 0$ and initialize subset of dual variables $\bar{\mathcal{J}} \subseteq \mathcal{J}$ and linear cuts $\mathcal{E}$
2: **while** termination condition not met **do**

Solve restricted Lagrangian relaxation using cutting-plane method

3: Set $k = 1$
4: Solve relaxed LP in (36)–(38) to obtain $\left(\lambda^{(k)}, \alpha^{(k)}, \pi^{(k)}\right)$
5: Perform spectral decomposition of matrix in (33) evaluated at $\left(\lambda^{(k)}, \alpha^{(k)}\right)$
6: **if** minimum eigenvalue $\geq -\epsilon$ **then**
7: $\left(\hat{\lambda}, \hat{\alpha}, \hat{\pi}\right) \leftarrow \left(\lambda^{(k)}, \alpha^{(k)}, \pi^{(k)}\right)$ and **go to line** 12;
8: **else**
9: Add new linear cuts to set $\mathcal{E}$ using eigenvectors with most negative eigenvalues
10: $k \leftarrow k + 1$ and **go to line** 4.
11: **end if**

Solve pricing problem

12: Calculate dual matrix $\hat{Y}$ using $\hat{\pi}$ as in Theorem 1
13: Solve pricing problem in (46) using $\hat{Y}$ to obtain $\hat{\varphi}$
14: **if** $\hat{\varphi} \leq 0$ **then**
15: $\alpha^* \leftarrow \hat{\alpha}$ and stop;
16: **else**
17: Add promising dual variables with most positive objective values in (46) to $\bar{\mathcal{J}}$
18: **go to line** 3
19: **end if**
20: **end while**

21: Return $\alpha_{\text{res}} \leftarrow \hat{\alpha}$

number of voxels considered in the optimization stage. Specifically, for the liver case, we use a down-sampling ratio of 1:4, 1:4, 1:16, 1:32, and 1:32 in heart, stomach, healthy liver, entrance, and skin, respectively. For the prostate case, we use a down-sampling ratio of 1:2, 1:1, and 1:32 in bladder, rectum, and body, respectively. No down-sampling is performed in PTV. The number of voxels in the original and down-sampled voxel grid are on the order of 10^6 and 10^4, respectively. We note that all evaluations are still performed on the original voxel grid.

For each critical structure, we tune the max-mean BED coefficient κ_s in the plan evaluation criterion (8) to achieve a conformal BED distribution and to avoid any hot spots in that structure. In the liver case, the coefficient is set to 0.8, 0.9, 0.8, 0.8, and 0.9 for the

healthy liver, heart, stomach, skin, and entrance region, respectively. In the prostate case, a coefficient of 0.5, 0.4, and 0.9 is used for the rectum, bladder, and body, respectively. The weighting factor γ_s in the objective function is chosen based on the relative importance of each critical structure as specified by the user. In the liver case, a weighting factor of 0.5 is used for all critical structures while in the prostate case, a weighting factor of 0.5, 0.4, and 0.4 is used for the rectum, bladder, and body, respectively.

A target $[\alpha/\beta]$ value of 10 Gy and 1.5 Gy is used for the liver and prostate case, respectively. Additionally, an $[\alpha/\beta]$ value of 4 Gy is used for all critical structures in the liver case. For the prostate case, an $[\alpha/\beta]$ parameter value of 3, 3, and 5 Gy is used for body, rectum, and bladder, respectively. A BED threshold value of zero is used in the penalty functions for all critical structures. To set the prescribed BED thresholds for PTV, we use the equivalent BED of a prescribed physical dose of 70 and 74 Gy delivered in a standard fractionation scheme with 30 fractions for the liver and prostate case, respectively. We consider a hypo-fractionation scheme to generate both uniformly and spatiotemporally fractionated plans with $|\mathcal{N}| = 5$ treatment fractions.

5.2. Implementation and Results

An important factor in spatiotemporal fractionation is the number of distinct dose distributions used over the treatment course. Based on the results obtained in our previous study for stylized cancer cases (Adibi and Salari 2018), the larger the number of distinct dose distributions allowed, the larger the spatiotemporal fractionation gain observed. However, a larger number of distinct dose distributions may increase the logistical cost associated with clinical workload and plan quality assurance. Hence, there is a trade-off between the additional spatiotemporal gain and the increase in clinical workload. Here, we start by considering spatiotemporal fractionation using only two distinct dose distributions. That means two sets of fluence maps for the incident beams will be used. To deliver the resulting spatiotemporally fractionated plan following a hypo-fractionated scheme with five fractions, one set of fluence maps will be delivered three times while the other set will be delivered twice. Note that the order in which the dose distributions are delivered does not alter the results. For each cancer case, we also obtain a uniformly fractionated reference plan by solving the special case of our QCQP reformulation presented in Section 2.2.

The computational experiments are performed in MATLAB environment using a Windows machine with an Intel Core i5 CPU and 8GB of RAM. The interior-point optimizer

(IPOPT) (Wächter and Biegler 2006) from the MATLAB OPTI toolbox (Currie and Wilson 2012) is used to solve the QP subproblems in the SQP method. To solve the LP subproblems in the SLP method, the Gurobi solver is used (Gurobi Optimization 2016).

In what follows, we first report on the performance of the proposed constraint-generation method in finding local optimal solutions using SQP and SLP, which we refer to as SQP-CG and SLP-CG, respectively. We then report on the performance of the proposed column-generation method in solving the Lagrangian relaxation to obtain optimality bounds. Furthermore, for each cancer case, we obtain both spatiotemporally and uniformly fractionated IMRT plans to quantify the potential therapeutic benefit of spatiotemporal fractionation. Finally, we use two additional test cases to evaluate the generalizability of the results.

5.2.1. Local-optimal spatiotemporally fractionated IMRT plans. To initialize the constraint-generation method (Algorithm 1), we start by including all constraints in the QCQP formulation, except for the constraints in (16) where only 50 voxels are randomly selected in each critical structure, and their corresponding quadratic constraints are added to the relaxed problem. Then, at each iteration, we add up to 50 most-violated relaxed constraints associated with each critical structure. The termination condition (i.e., no significant violation of the relaxed constraints) is satisfied after adding around 15% of all critical-structure voxels. Results of the constraint-generation method are reported in Table 1. We note that for the case of uniform fractionation, SLP-CG and SQP-CG are applied to the special case of the QCQP reformulation presented in Section 2.2. Results

Table 1 Comparison of SLP-CG versus SQP-CG performance

		Liver		Prostate	
Fractionation	**Method**	Obj. fun. val.	CPU time [s]	Obj. fun. val.	CPU time [s]
Uniform	SLP-CG	191.96	651.1	99.93	711.3
	SQP-CG	191.92	2163.1	99.89	2313.6
Spatiotemporal	SLP-CG	185.45	1102.5	94.04	1182.4
	SQP-CG	185.57	2871.5	93.95	2948.1

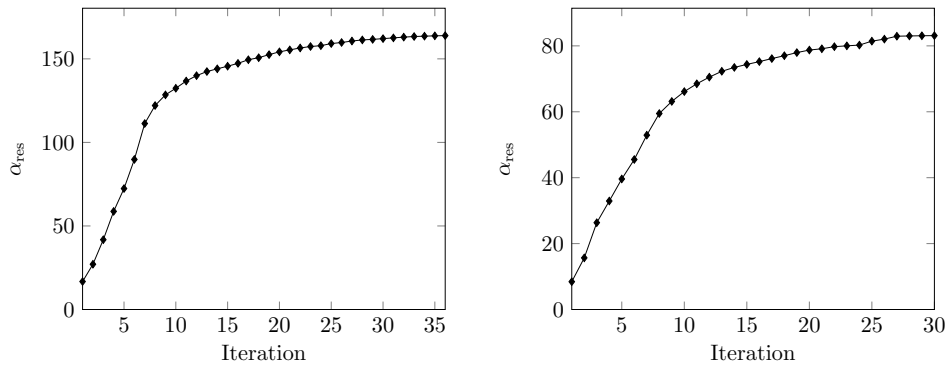
show that the SLP-CG and SQP-CG methods yield locally optimal solutions with almost identical objective values. Nevertheless, the required computational time for SLP-CG is around 50% shorter than that of SQP-CG; therefore, we employ SLP-CG for the remainder of the reported computational experiments. Table 1 also shows that the spatiotemporally fractionated plan yields a smaller objective value than uniform fractionation by 3% and 6% for the liver and prostate case, respectively. Next, we increase the number of distinct dose

distributions used in spatiotemporal fractionation. Table 2 shows the optimal objective values obtained using uniform versus spatiotemporal fractionation with up to five distinct dose distributions. Results show that as the number of distinct dose distributions increases, the optimal objective value decreases, though with a diminishing rate of return.

Table 2 Objective function value of spatiotemporally and uniformly fractionated plans with up to five distinct dose distributions. Percentage gain over uniform fractionation is shown in parentheses

Case	Uniform	Spatiotemporal			
		# of distinct dose distributions			
		2	3	4	5
Liver	191.73	185.45 (3.3%)	184.86 (3.6%)	183.30 (4.4%)	182.79 (4.7%)
Prostate	99.69	94.04 (5.7%)	90.86 (8.9%)	90.57 (9.1%)	90.36 (9.4%)

5.2.2. Optimality bounds. To ensure the quality of locally optimal spatiotemporally fractionated plans, a lower bound on the global optimal solution is obtained by solving the Lagrangian relaxation using the column-generation method described in Section 4 (Algorithm 2). We initialize the set of dual variables in the restricted Lagrangian relaxation by including all variables corresponding to constraints (17)–(19) as well as 500 dual variables associated with randomly selected constraints from (15)–(16). Then, at each iteration, we add up to 100 most promising dual variables identified based on the pricing problem. We use a termination condition of no significant improvement in the objective function value ($< 10^{-3}$) to stop the column-generation method. Furthermore, for the cutting-plane method, we initialize the set of linear cuts by including 1,000 DSOS cuts. At each iteration, we add all eigenvectors associated with the negative eigenvalues obtained from spectral decomposition that are below the user-specified threshold. The computational time required for solving different instances of the Lagrangian relaxation for each cancer case varies between 4 and 5 hours, depending on the number of distinct dose distributions allowed. Figure 1 shows the progression of the lower bound (α_{res}) through iterations of the column-generation method for the two cancer cases. Once the lower bound is obtained, the optimality bound can be calculated as the relative difference between the lower bound and the objective value of the local-optimal solution obtained using the constraint-generation method (i.e., upper bound). Results are reported in Table 3, which show that for the two cancer cases tested, an optimality bound of 10–12% can be achieved by solving the Lagrangian relaxation. The column-generation method terminates after including only 8–10% of all dual variables. The



(a) Liver (b) Prostate

Figure 1 **Progression of provided lower bound through iterations of column-generation method for spatiotemporally fractionated plans with two distinct dose distributions**

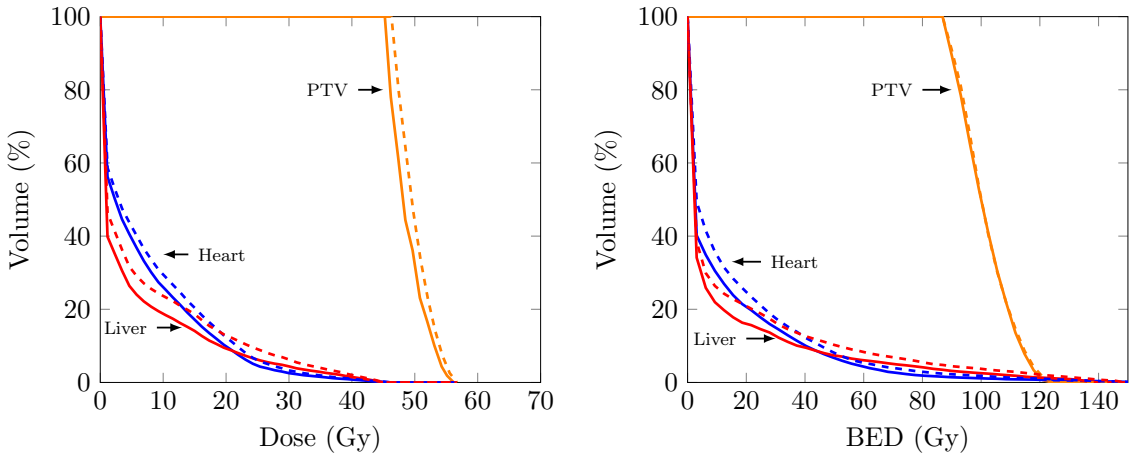
well-known *tailing-off* effect of the column-generation method can be seen in Figure 1. As discussed in Section 4.2, if desired, we can terminate the column-generation method prior to convergence and still obtain a valid lower bound. In particular, an optimality bound of 15.2% and 15.1% can be obtained within a running time of 135 and 119 minutes for the liver and prostate case, respectively.

Table 3 **Lower bounds and relative optimality bounds obtained for the liver and prostate cases**

Case	**Parameter**	**Spatiotemporal**			
		# of distinct dose distributions			
		2	3	4	5
Liver	Lower bound	163.91	163.98	164.02	164.17
	Optim. gap (%)	11.6	11.3	10.5	10.2
Prostate	Lower bound	83.11	82.02	80.59	78.98
	Optim. gap (%)	11.6	9.7	11.0	12.6

5.2.3. Potential therapeutic benefit of spatiotemporal fractionation. Next, we compare uniformly and spatiotemporally fractionated plans. Figures 2(a) and 3(a) compare the *dose-volume histogram* (DVH) for the physical dose distributions delivered by the two plans. The DVH curves show the percentage of the structure volume that receives at least a specific dose. DVH is commonly used to evaluate the clinical quality of radiotherapy plans. Similarly, Figures 2(b) and 3(b) compare the *BED-volume histogram* (BEDVH) for the uniform and spatiotemporal fractionation.

By comparing the DVH and BEDVH curves for the two plans one can notice that spatiotemporal fractionation achieves the same BED coverage in PTV while using less physical dose compared to the uniform fractionation. Additionally, spatiotemporal fractionation

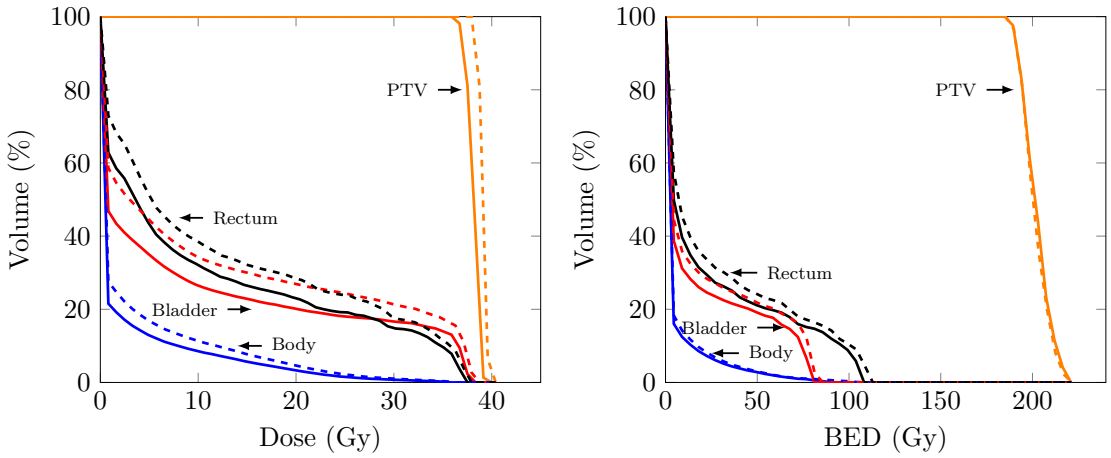


(a) Dose-volume histogram (b) BED-volume histogram

Figure 2 **Results obtained for the liver case: (a) dose-volume histogram and (b) BED-volume histogram of uniformly (dashed) and spatiotemporally (solid) fractionated plans using 5 distinct dose distributions**

achieves a better BED distribution in critical structures without compromising the BED coverage in PTV. The changes in the mean and maximum of the BED distribution in each critical structure when using uniform versus spatiotemporal fractionation are reported in the online supplement. Dose-wash diagrams associated with the uniformly and spatiotemporally fractionated plans are generated for both cancer cases using the *Computational Environment for Radiological Research* (CERR) (Deasy et al. 2003) and are available in the online supplement. They show the division of the target volume into subregions, each irradiated by a distinct dose distribution.

We also evaluate the potential benefit of spatiotemporal fractionation when focusing on max-mean BED reduction in a single dose-limiting critical structure, which is the healthy liver and rectum in the liver and prostate case, respectively. To that end, we solve the special case of the proposed formulation presented in Section 2.3 using both spatiotemporal and uniform fractionation. We then calculate the spatiotemporal gain and compare that against the gain obtained when considering all critical structures in the objective function. The results are reported in Table 4, which show that a higher percentage gain is expected when the emphasis is on a single critical structure. Furthermore, the extent of the spatiotemporal gain in the dose-limiting critical structure depends on the convex-combination coefficient κ used in the max-mean BED criterion in (8). Figure 4 shows the sensitivity of the spatiotemporal gain to the κ parameter value used for the dose-limiting critical



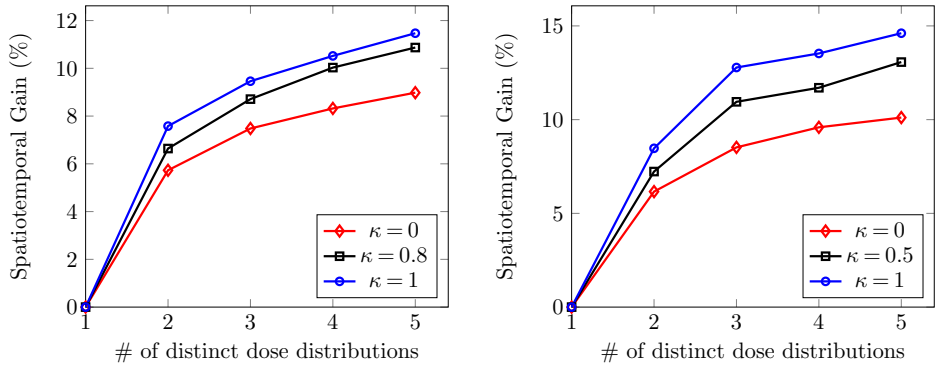
(a) Dose-volume histogram (b) BED-volume histogram

Figure 3 **Results obtained for the prostate case: (a) dose-volume histogram and (b) BED-volume histogram of uniformly (dashed) and spatiotemporally (solid) fractionated plans using 5 distinct dose distributions**

Table 4 **Comparison of spatiotemporal gain in the optimal objective value over uniform fractionation when considering all critical structures versus a single dose-limiting structure in the objective function**

Case	Max-mean BED Penalty	Spatiotemporal Gain (%)			
		# of distinct dose distributions			
		2	3	4	5
Liver	All critical structures	3.27	3.58	4.40	4.66
	Healthy liver only	6.64	8.71	10.03	10.87
Prostate	All critical structures	5.67	8.86	9.14	9.35
	Rectum only	7.23	10.95	11.70	13.07

structure. The spatiotemporal gain increases as more emphasis is put on the maximum of the BED distribution (i.e., $\kappa \to 1$)



(a) Healthy Liver (Liver Case) (b) Rectum (Prostate Case)

Figure 4 **Sensitivity of spatiotemporal gain to max-mean BED coefficient of (a) healthy liver and (b) rectum**

5.2.4. Performance Evaluation Using Additional Test Cases. To examine the generalizability of the reported results, we apply the proposed spatiotemporal fractionation framework to two additional liver cases chosen from *the radiotherapy optimisation test set* (TROTS) (Breedveld and Heijmen 2017). The specifications of these two test cases are provided in the online supplement. To run the additional cases, we use the same model parameters as defined earlier in Section 5.1. The beam arrangement for the liver cases in the TROTS dataset consists of 15 non-coplanar beams with almost twice as many beamlets as in the liver case of the CORT dataset, leading to relatively longer computational times. Depending on the number of distinct dose distributions, the constraint-generation method takes between 20 to 80 minutes to obtain local optimal solutions for the TROTS test cases. The column-generation method takes between 4 to 6 hours to obtain optimality bounds. A summary of the results obtained for these two test cases are reported in Table 5, which shows that the proposed methods yield a similar spatiotemporal gain and optimality gap to those obtained for the CORT liver case.

Table 5 Summary of computational results obtained for TROTS test cases

Case	Performance metric	Uniform	Spatiotemporal			
			# of distinct dose distributions			
			2	3	4	5
Liver 01	Obj. function value	178.31	168.14	167.2	166.92	166.47
	Spatiotemporal gain (%)	-	5.7	6.2	6.4	6.6
	Lower bound	-	147.32	147.67	145.51	145.16
	Optim. gap (%)	-	12.4	11.7	12.8	12.8
Liver 02	Obj. function value	167.12	159.47	158.26	157.31	156.82
	Spatiotemporal gain (%)	-	4.6	5.3	5.9	6.2
	Lower bound	-	143.88	143.01	141.9	141.84
	Optim. gap (%)	-	9.8	9.6	9.8	9.6

6. Discussion

In this section, we discuss the main insights gained from our computational results and highlight the key findings.

As Figures 2(b) and 3(b) suggest, a potential therapeutic gain may be achieved from spatiotemporal fractionation over uniform fractionation due to the reduction of the BED deposited in the critical structures while maintaining the same minimum BED level in the target volume. This gain is attributed to a selective reduction of the fractionation effect in the target volume, which leads to escalated BED values in the target volume. Thus, the spatiotemporally fractionated plan is able to satisfy the minimum BED requirement

in the target volume while delivering smaller BED values to critical structures. A selective reduction of BED values in critical structures may potentially reduce the risk of normal-tissue complications (Liu et al. 2017).

The observed gain in the objective function value due to spatiotemporal fractionation varies across the two cancer cases and is approximately 5% for the liver cases and 9% for the prostate case. This yields a reduction of 15% and 6% in the mean and maximum of the BED distribution in the healthy liver, and a reduction of 11% and 7% in the mean and maximum of the BED distribution in the rectum, respectively. However, this gain may come at the expense of an increased clinical workload to treat a cancer patient. Table 2 suggests that there may be a diminishing return in the potential therapeutic gain as we increase the number of distinct dose distributions for the cancer cases considered in this study. Hence, the trade-off between the potential therapeutic gain and the increased clinical workload may need to be considered when designing spatiotemporally fractionated plans.

Results show that a larger spatiotemporal fractionation gain may be achieved if a single critical structure is considered in the objective function while the quality of the BED distribution in all other critical structures are constrained to the levels achieved in uniform fractionation. This is mostly suited for a clinical scenario in which the reduction of the complication risk in a major dose-limiting structure, beyond what could be achieved with uniform fractionation, is desired. In that case, spatiotemporal fractionation could be used to reduce the BED levels in that particular structure without comprising tumor control or increasing the risk of complications in other critical structures.

Because of the non-convexity of the spatiotemporal fractionation problem, we obtain optimality bounds to evaluate the quality of the local-optimal solutions. As reported in Table 3, the optimality bounds obtained from solving the Lagrangian relaxation for the two cancer cases are 10–12%. The optimality bounds provide valuable insight into the extent of the potential gain that may be achieved from spatiotemporal fractionation.

The proposed constraint- and column-generation methods in this study may be adapted to locally solve and find optimality bounds for non-convex QCQP problems in other application domains. QCQP problems arise in different areas, including graph theory (Boyd and Vandenberghe 2004), machine learning (Zhu et al. 2006), power systems (Low 2014a,b), radar detection (De Maio et al. 2010), and signal processing (Huang and Palomar 2014).

7. Conclusion

In this study, customized models and solution methods were developed to obtain high-quality spatiotemporally fractionated IMRT plans with tight optimality bounds for clinical cancer cases. The proposed BED-based formulation encourages lower BED values in critical structures while ensuring that the target volume receives the prescribed BED value. A QCQP reformulation of the problem was used as the basis for solution method development. The QCQP reformulation is non-convex in general and large-scale. Hence, a constraint-generation method was developed to obtain local-optimal solutions while a column-generation method was employed to solve the Lagrangian relaxation to obtain optimality bounds.

There are limitations with our study that point towards future research. Currently, there is no mechanism to tighten the optimality bound for spatiotemporally fractionated plans beyond what is obtained by solving the Lagrangian relaxation. Further reduction of the optimality gap requires the development of global optimization techniques such as branch-and-bound methods or cutting plane methods (Linderoth 2005, You et al. 2018). There are several sources of uncertainty that may compromise the clinical quality of spatiotemporally fractionated plans. Specifically, it is difficult to accurately estimate radio-biological parameter values, leading to ambiguity in the choice of appropriate values (Van Leeuwen et al. 2018). Additionally, there may be inter- and intra-patient variability in the BED parameter value (Menzies et al. 2014). Finally, spatiotemporally fractionated plans are more susceptible to patient setup errors and anatomical uncertainties due to organ motion compared to conventional IMRT plans (Gaddy et al. 2019). Future research may extend the proposed spatiotemporal fractionation framework in this study by explicitly incorporating those uncertainties.

Acknowledgments

This research was funded in part by the National Science Foundation through Award #1662819.

References

Adibi A, Salari E (2018) Spatiotemporal radiotherapy planning using a global optimization approach. *Physics in Medicine and Biology* 63(3).

Ahmadi AA, Dash S, Hall G (2017) Optimization over structured subsets of positive semidefinite matrices via column generation. *Discrete Optimization* 24:129–151.

Ajdari A, Ghate A (2016a) A model predictive control approach for discovering nonstationary fluence-maps in cancer radiotherapy fractionation. *Proceedings of the 2016 Winter Simulation Conference*, 2065–2075 (IEEE Press).

Ajdari A, Ghate A (2016b) Robust spatiotemporally integrated fractionation in radiotherapy. *Operations Research Letters* 44(4):544–549.

Ajdari A, Ghate A, Kim M (2018) Adaptive treatment-length optimization in spatiobiologically integrated radiotherapy. *Physics in Medicine & Biology* 63(7):075009.

Ajdari A, Niyazi M, Nicolay NH, Thieke C, Jeraj R, Bortfeld T (2019) Towards optimal stopping in radiation therapy. *Radiotherapy and Oncology* 134:96–100.

Ajdari A, Saberian F, Ghate A (2020) A theoretical framework for learning tumor dose-response uncertainty in individualized spatiobiologically integrated radiotherapy. *INFORMS Journal on Computing* 32(4):930–951.

Bazaraa MS, Sherali HD, Shetty CM (2013) *Nonlinear programming: Theory and algorithms* (John Wiley & Sons).

Beyzadeoglu M, Ozyigit G, Ebruli C (2010) *Basic radiation oncology* (Springer Science & Business Media).

Bortfeld T (2006) IMRT: A review and preview. *Physics in Medicine & Biology* 51(13):R363.

Boyd S, Vandenberghe L (2004) *Convex optimization* (Cambridge University Press).

Breedveld S, Heijmen B (2017) Data for TROTS–the radiotherapy optimisation test set. *Data in Brief* 12:143–149, URL `https://www.erasmusmc.nl/en/cancer-institute/research/projects/radiotherapy-trots`.

Byrd RH, Nocedal J, Waltz RA (2006) Knitro: An integrated package for nonlinear optimization. *Large-scale nonlinear optimization*, 35–59 (Springer).

Correa R, Ramirez C H (2004) A global algorithm for nonlinear semidefinite programming. *SIAM Journal on Optimization* 15(1):303–318.

Craft D, Bangert M, Long T, Papp D, Unkelbach J (2014) Shared data for intensity modulated radiation therapy (IMRT) optimization research: The CORT dataset. *GigaScience* 3(37):1–12, URL `http://gigadb.org/dataset/view/id/100110`.

Craft D, Halabi T, Bortfeld T (2005) Exploration of tradeoffs in intensity-modulated radiotherapy. *Physics in Medicine and Biology* 50(24):5857–5868.

Currie J, Wilson DI (2012) OPTI: lowering the barrier between open source optimizers and the industrial MATLAB user. *Foundations of Computer-Aided Process Operations* 8–11.

De Maio A, Huang Y, Palomar DP, Zhang S, Farina A (2010) Fractional QCQP with applications in ML steering direction estimation for radar detection. *IEEE Transactions on Signal Processing* 59(1):172–185.

Deasy JO, Blanco AI, Clark VH (2003) CERR: a computational environment for radiotherapy research. *Medical Physics* 30(5):979–985.

d'Aspremont A, Boyd S (2003) Relaxations and randomized methods for nonconvex QCQPs. Technical report, Lecture Notes of EE392, Stanford University.

Fowler JF (1989) The linear-quadratic formula and progress in fractionated radiotherapy. *British Journal of Radiology* 62(740):679–694.

Fowler JF (2010) 21 years of biologically effective dose. *British Journal of Radiology* 83(991):554–568.

Furrer L (2009) An overview of cutting plane methods for semidefinite programming .

Gaddy MR, Unkelbach J, Papp D (2019) Robust spatiotemporal fractionation schemes in the presence of patient setup uncertainty. *Medical Physics* 46(7):2988–3000.

Gaddy MR, Yıldız S, Unkelbach J, Papp D (2018) Optimization of spatiotemporally fractionated radiotherapy treatments with bounds on the achievable benefit. *Physics in Medicine & Biology* 63(1):015036.

Gurobi Optimization I (2016) Gurobi optimizer reference manual. URL `http://www.gurobi.com`.

Henríquez FC, Castrillón SV (2011) A quality index for equivalent uniform dose. *Medical Physics* 36(3):126–132.

Huang Y, Palomar DP (2014) Randomized algorithms for optimal solutions of double-sided QCQP with applications in signal processing. *IEEE Transactions on Signal Processing* 62(5):1093–1108.

Jones B, Dale R (2019) The evolution of practical radiobiological modelling. *The British Journal of Radiology* 92(1093):20180097.

Linderoth J (2005) A simplicial branch-and-bound algorithm for solving quadratically constrained quadratic programs. *Mathematical Programming* 103(2):251–282.

Liu F, Tai A, Lee P, Biswas T, Ding GX, El Naqa I, Grimm J, Jackson A, LaCouture T, Loo Jr B, et al. (2017) Tumor control probability modeling for stereotactic body radiation therapy of early-stage lung cancer using multiple bio-physical models. *Radiotherapy and Oncology* 122(2):286–294.

Long T, Matuszak M, Feng M, Fraass B, Ten Haken R, Romeijn H (2012) Sensitivity analysis for lexicographic ordering in radiation therapy treatment planning. *Medical Physics* 39(6):3445–3455.

Low SH (2014a) Convex relaxation of optimal power flow—part i: Formulations and equivalence. *IEEE Transactions on Control of Network Systems* 1(1):15–27.

Low SH (2014b) Convex relaxation of optimal power flow—part ii: Exactness. *IEEE Transactions on Control of Network Systems* 1(2):177–189.

Menzies AM, Haydu LE, Carlino MS, Azer MW, Carr PJ, Kefford RF, Long GV (2014) Inter-and intra-patient heterogeneity of response and progression to targeted therapy in metastatic melanoma. *PloS One* 9(1):e85004.

Mizuta M, Takao S, Date H, Kishimoto N, Sutherland KL, Onimaru R, Shirato H (2012) A mathematical study to select fractionation regimen based on physical dose distribution and the linear–quadratic model. *International Journal of Radiation Oncology* Biology* Physics* 84(3):829–833.

Niemierko A (1999) A generalized concept of equivalent uniform dose (EUD). *Medical Physics* 26(6):1100.

Nocedal J, Wright SJ (2006) *Numerical optimization* (Springer).

Park J, Boyd S (2017) General heuristics for nonconvex quadratically constrained quadratic programming. *arXiv preprint arXiv:1703.07870* .

Qualizza A, Belotti P, Margot F (2012) Linear programming relaxations of quadratically constrained quadratic programs. *In Mixed Integer Nonlinear Programming*, 407–426 (Springer).

Romeijn E, Dempsey JF (2008) Intensity modulated radiation therapy treatment plan optimization. *TOP* 16(2):215–243.

Saberian F, Ghate A, Kim M (2015) A two-variable linear program solves the standard linear–quadratic formulation of the fractionation problem in cancer radiotherapy. *Operations Research Letters* 43(3):254–258.

Saberian F, Ghate A, Kim M (2016) A theoretical stochastic control framework for adapting radiotherapy to hypoxia. *Physics in Medicine & Biology* 61(19):7136–7161.

Saberian F, Ghate A, Kim M (2017) Spatiotemporally optimal fractionation in radiotherapy. *INFORMS Journal on Computing* 29(3):422–437.

Ten Eikelder SC, Ferjančič P, Ajdari A, Bortfeld T, den Hertog D, Jeraj R (2020) Optimal treatment plan adaptation using mid-treatment imaging biomarkers. *Physics in Medicine & Biology* 65(24):245011.

Thieke C, Bortfeld T, Niemierko A, Nill S (2003) From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning. *Medical Physics* 30(9):2332–2339.

Unkelbach J, Bussière MR, Chapman PH, Loeffler JS, Shih HA (2016) Spatiotemporal fractionation schemes for irradiating large cerebral arteriovenous malformations. *International Journal of Radiation Oncology* Biology* Physics* 95(3):1067–1074.

Unkelbach J, Craft D, Salari E, Ramakrishnan J, Bortfeld T (2012) The dependence of optimal fractionation schemes on the spatial dose distribution. *Physics in Medicine and Biology* 58(1):159.

Unkelbach J, Papp D (2015) The emergence of nonuniform spatiotemporal fractionation schemes within the standard BED model. *Medical Physics* 42(5):2234–2241.

Unkelbach J, Papp D, Gaddy MR, Andratschke N, Hong T, Guckenberger M (2017) Spatiotemporal fractionation schemes for liver stereotactic body radiotherapy. *Radiotherapy and Oncology* 125(2):357–364.

Van Leeuwen C, Oei A, Crezee J, Bel A, Franken N, Stalpers L, Kok H (2018) The alfa and beta of tumours: a review of parameters of the linear-quadratic model, derived from clinical radiotherapy studies. *Radiation Oncology* 13(96), URL `https://doi.org/10.1186/s13014-018-1040-z`.

Wächter A, Biegler LT (2006) On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming* 106(1):25–57.

Yang Y, Xing L (2005) Optimization of radiotherapy dose-time fractionation with consideration of tumor specific biology. *Medical Physics* 32(12):3666–3677.

You S, Wan C, Dai R (2018) A complementary cutting plane approach for nonconvex quadratically constrained quadratic programs. *2018 IEEE Conference on Decision and Control (CDC)*, 1–7 (IEEE).

Zhu X, Kandola J, Lafferty J, Ghahramani Z (2006) Graph kernels by spectral transforms. *In Semi-supervised learning* 277–291.