UCLA
UCLA Electronic Theses and Dissertations

Title
INTEGRATED BEAM ORIENTATION OPTIMIZATION FOR ROBUST INTENSITY-MODULATED PROTON THERAPY

Permalink
https://escholarship.org/uc/item/16h042rw

Author
Gu, Wenbo

Publication Date
2020

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Integrated Beam Orientation Optimization for Robust Intensity-Modulated Proton Therapy

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Physics and Biology in Medicine

by

Wenbo Gu

2020

© Copyright by

Wenbo Gu

2020

ABSTRACT OF THE DISSERTATION

INTEGRATED BEAM ORIENTATION OPTIMIZATION FOR ROBUST INTENSITY-MODULATED PROTON THERAPY

by

Wenbo Gu

Doctor of Philosophy in Physics and Biology in Medicine

University of California, Los Angeles, 2020

Professor Ke Sheng, Chair

Purpose

Dose conformality and robustness are equally important in Intensity-Modulated Proton Therapy (IMPT). Despite the important role of beam orientation on both dosimetry and robustness, an automated, robust beam orientation optimization algorithm has not been incorporated due to the problem complexity and paramount computational challenge. In this dissertation, we propose a novel IMPT framework that integrates robust beam orientation optimization (BOO) and robust fluence map optimization (FMO) in a unified framework.

Methods

The unified framework is formulated to include a dose fidelity term, a heterogeneity-weighted group sparsity term, and a sensitivity regularization term. The dose fidelity term encourages less physical dose deviation from ideal distribution. The L2,1/2-norm group sparsity is used to reduce the number of active beams from the initial 1162 evenly distributed non-coplanar candidate beams, to between 2 and 4. A heterogeneity index, which evaluates the lateral tissue heterogeneity of a beam, is used to weigh the group sparsity term. With this index, beams more resilient to setup uncertainties are encouraged. There is a symbiotic relationship between the heterogeneity index and the sensitivity regularization; the integrated optimization framework further improves beam robustness against both range and setup uncertainties. This Sensitivity regularization and Heterogeneity weighting based BOO and FMO framework (SHBOO-FMO) was tested on two skull-base tumor (SBT) patients and two bilateral head-and-neck (H&N) patients. The conventional CTV-based optimized plans (Conv) with SHBOO-FMO beams (SHBOO-Conv) and manual beams (MAN-Conv) were compared to investigate the beam robustness of the proposed method. The dosimetry and robustness of SHBOO-FMO plan were compared against the manual beam plan with CTV-based voxel-wise worst-case scenario approach (MAN-WC).

Results

With SHBOO-FMO method, the beams with superior range robustness over manual beams were selected while the setup robustness was maintained or improved. On average, the lowest [D95%, V95%, V100%] of CTV were increased from [93.8%, 91.0%, 70.6%] in

MAN-Conv plans, to [98.6%, 98.6%, 96.1%] in SHBOO-Conv plans with range uncertainties. With setup uncertainties, the average lowest [D98%, D95%, V95%, V100%] of CTV were increased from [92.0%, 94.8%, 94.3%, 78.9%] in MAN-Conv plans, to [93.5%, 96.6%, 97.0%, 91.9%] in SHBOO-Conv plans. Compared with the MAN-WC plans, the final SHBOO-FMO plans achieved comparable plan robustness and better OAR sparing, with an average reduction of [Dmean, Dmax] of [6.3, 6.6] GyRBE for the SBT cases and [1.9, 5.1] GyRBE for the H&N cases from the MAN-WC plans.

Conclusions

A novel robust optimization method was developed for IMPT. It integrates robust BOO and robust FMO into a unified framework, and the resulting optimization problem can be solved efficiently. Compared with the current clinical practice, where beam angles are manually selected and fluence map is optimized by worst-case method, the planning efficiency is improved, and it generates plans with superior dosimetry and good robustness.

The dissertation of Wenbo Gu is approved.

Dan Ruan

Lei Dong

Michael McNitt-Gray

Ke Sheng, Committee Chair

University of California, Los Angeles

2020

To Zhicheng, Wencheng, Dad, and Mom

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF APPENDICES

ACKNOWLEDGEMENTS

I would first like to express my deepest gratitude to my advisor, Dr. Ke Sheng. When I first entered UCLA, I was anxious, unconfident and aimless about my graduate study. It is his countless guidance, encouragements and trust that help me through the difficult times and motivate me to enjoy learning and research. His wisdom, creativity and passion show me the charm of medical physics and give me directions of the research projects and future career. He is the best mentor I can ever ask for, making my journey of research and discovery full of fun and fulfillment.

I would also like to thank Dr. Dan Ruan, who provides so much help to my research work and supports me throughout my graduate study. It has been a true fortune to work with such a knowledgeable, passionate, creative, and compassionate person. I would also like to thank Dr. Wei Zou and Dr. Lei Dong for the valuable insight and knowledge they have provided me.

Many thanks to my lab mates Ryan, Qihui, Elizabeth, Daili, Jiayi, Dan, Daniel, Angelia and Kaley. The friendship and companionship from you make my graduate years fun and unforgettable. Special thanks to Victoria. You are not only a friend, but like an elder sister who takes so much care of me both in and out of the lab. You make Sheng lab like home.

Thanks should also go to Dr. Michael McNitt-Gray, not only for the help as my dissertation committee, but also for being a kind and supportive program director. Thank

you to Reth and Alondra. I feel strong and worry-free knowing you and the program are protecting and supporting me.

I'd like to thank my dad Aifei, and mom Honghua for the love you give throughout my life. Even though you know nothing about medical physics and graduate school, you have been always respecting my decision and trusting me under any conditions. Thank you to my sister Wencheng and brother-in-law Jinxin, for your company and support to my dreams. Thank you to my boyfriend Zhicheng. Having you by my side makes me the happiest and luckiest person. I love you all.

VITA

EDUCATION

M.S.	University of California—Los Angeles, Physics and Biology in Medicine	2018
B.S.	Tsinghua University, Nuclear Engineering	2014

AWARDS

UCLA Physics and Biology in Medicine Colloquium, Best Oral Presentation	2019
AAPM John R. Cameron Young Investigators Symposium Finalist	2019
PTCOG Michael Goitein Best Abstract Award Finalist	2019
Editor's Choice of Medical Physics August 2019	2019

PEER-REVIEWED PUBLICATIONS

1. **Gu W**, Ruan D, Lyu Q, et al. A Novel Energy Layer Optimization Framework for Spot-Scanning Proton Arc Therapy. Medical Physics. 2020. doi:10.1002/mp.14083
2. **Gu W**, Neph R, Ruan D, et al. Robust beam orientation optimization for intensity-modulated proton therapy. Medical Physics. 2019; 46: 3356-3370. doi:10.1002/mp.13641
3. **Gu W**, Ruan D, O'Connor D, et al. Robust optimization for intensity-modulated proton therapy with soft spot sensitivity regularization. Medical Physics. 2019; 46:1408-1425. doi:10.1002/mp.13344.
4. **Gu W**, O'Connor D, Nguyen D, et al. Integrated beam orientation and scanning-spot optimization in intensity-modulated proton therapy for brain and unilateral head and neck tumors. Medical Physics. 2018; 45: 1338-1350. doi:10.1002/mp.12788.
5. Shang D, **Gu W**, Landers A, et al. Technical Note: Robust individual Thermoluminescence dosimeter tracking using optical fingerprinting. Medical Physics. 2020; 47(1):267-271. doi: 10.1002/mp.13895.

SELECTED CONFERENCE PRESENTATIONS

1. **Gu W**, Neph R, Ruan D, et al."Robust Beam Orientation Optimization for Intensity-Modulated Proton Therapy" (2019) Presented at the Annual AAPM Meeting, San Antonio TX
2. **Gu W**, Neph R, Ruan D, et al."Robust Beam Orientation Optimization for Intensity-Modulated Proton Therapy" (2019) Presented at the 58th Annual Conference of the Particle Therapy Co-Operative Group (PTCOG), Manchester, England
3. **Gu W**, Ruan D, Zou W, et al."High Fidelity Robust Optimization for Intensity-Modulated Proton Therapy" (2018) Presented at the Annual AAPM Meeting, Nashville TN
4. **Gu W**, O'Connor D, Nguyen D, et al."Integrated Beam Angle and Scanning Spot Optimization for Intensity Modulated Proton Therapy " (2017) Presented at the Annual ASTRO Meeting, San Diego CA
5. **Gu W**, O'Connor D, Nguyen D, et al."Integrated Beam Angle and Scanning-Spot Optimization in Intensity Modulated Proton Therapy Using Group Sparsity" (2017) Presented at the Annual AAPM Meeting, Denver CO

1 INTRODUCTION

1.1 Intensity-Modulated Proton Therapy

Since the discovery and first medical application of X-rays in 1895[1], radiation therapy has been one of the primary treatments for cancer. The ultimate goal of radiation therapy is to deliver a sufficient dose to eradicate cancerous tissue, while minimizing the irradiation to healthy tissue. Many developments and improvements have been made to reach this goal. Among them, utilizing the unique dose deposition characteristic of proton particles has drawn massive attention. As shown in Figure 1-1, compared with the photon, proton beam has low entrance-dose, a sharp rise near the end of its range, which is called Bragg peak, and negligible exit-dose. The finite range and the characteristic Bragg peak make proton a promising radiotherapy modality to spare healthy tissue while delivering a conformal dose to the target.

Since 1946 when the concept was first brought up by Wilson[2], proton therapy has been gradually translated into clinical treatment. The protons generated by an accelerator,

either a cyclotron or a synchrotron, forms a narrow pencil beam. In order to cover the entire treatment field, the narrow pencil beam needs to be scattered by a foil or scanned by magnets, referred to as passive scattering and active scanning[3], respectively (Figure 1-2). Even though passive scattering was the mainstream due to its simplicity, the freedom of modulation is limited, and the dose conformality fails to outperform than the commonly used Intensity-Modulated X-ray Therapy (referred to as IMXT). The recent decade sees rapid development and increasing adoption of pencil beam scanning (PBS) technique[4] in proton therapy. With PBS, the energy and intensity of the Bragg peaks of well-defined narrow pencil beams can be modulated in 3D space to generate a sophisticated dose distribution for tumor coverage and normal tissue sparing[5,6]. This technique is called Intensity-Modulated Proton Therapy (IMPT). The pencil beam scanning is also called spot-scanning, and the narrow pencil beam is referred to as scanning-spot, or spot, as well.

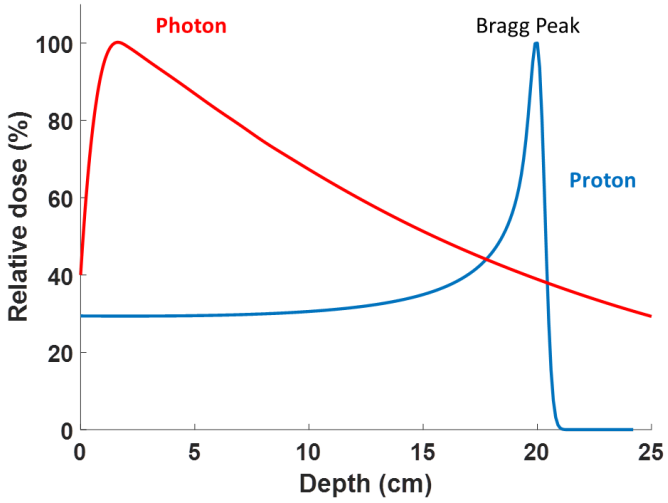

Figure 1-1: Depth-dose curve of photon beam and monoenergetic proton beam.

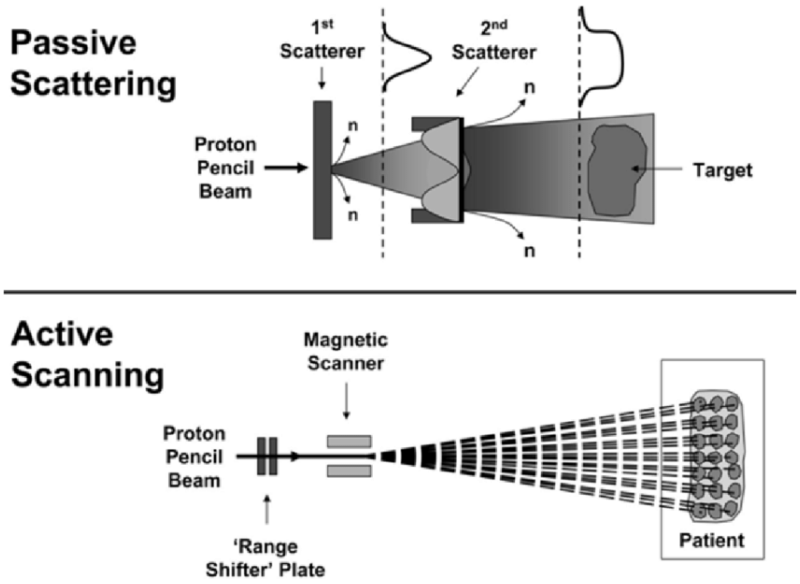

Figure 1-2: The difference of passive scattering and active scanning. The narrow proton beam from a cyclotron or synchrotron is either scattered by a foil or scanned by magnets to cover the treatment field. Reprinted from "Intensity-Modulated Radiation Therapy, Protons, and the Risk of Second Cancers" by E. Hall and D. Phil, Int J Radiat Oncol Biol Phys, 2006 ,65(1):1-7. Copyright 2006 by Elsevier Inc. [3]

There are several variations of intensity modulation, including 2D modulation, distal edge tracking, 2.5D modulation, and 3D modulation[5]. Among them, 3D modulation places the Bragg peaks throughout the target volume in 3 dimensions and optimizes the intensity of each individual pencil beam, therefore fully takes advantage of the freedom of Bragg peaks localizations. 3D IMPT can be further divided into two categories, single-field uniform dose (SFUD) and multi-field optimization (MFO)[5]. In SFUD, each radiation field is optimized independently to achieve a uniform dose distribution. In MFO, all fields are optimized simultaneously. Therefore each field generates a highly heterogeneous dose, but the combination from all fields achieves a highly conformal dose. MFO-IMPT utilizing all available degrees of freedom has the greatest versatility for normal organ sparing. This dissertation focuses on MFO-IMPT, which will be simply termed IMPT in the following sections.

1.2 Uncertainties in IMPT

Due to the sharp drop-off of proton Bragg peak, IMPT is very sensitive to uncertainties occurring during planning and delivery, including setup errors, patient anatomy changes, and range uncertainties[7–11]. The setup uncertainties are from the misalignment between patient anatomy and proton beams as well as the shifting of internal density heterogeneity[8]. The range uncertainties originate from various sources, such as CT (Computed Tomography) image artifacts, uncertainties in CT numbers, and the conversion from CT numbers to stopping powers. Patient anatomy change and tumor shrinkage will also lead to different proton range than expected. The existence of these uncertainties can make the actual dose distribution substantially differ from what is indicated in the treatment plan. Different from X-ray treatment planning, the perturbation of dose distribution in IMPT not only exists in the proximal and distal edge of the target volume but also within it, introducing significant cold or hot spots.

Multi-field optimized IMPT is even more susceptible to positioning errors or range uncertainties due to the high dose heterogeneity in a single field. In MFO, the dose in an individual beam is not homogeneous and only covers a partial volume of the target, with the rest usually compensated by the dose from other fields. Range uncertainties can possibly lead to overshooting (protons shoot deeper than expected) or undershooting (protons shoot shallower than expected) of two compensating beams, leaving cold or hot spots in the target.

If uncertainties are unaccounted for, the effectiveness of IMPT from Bragg peak and 3D modulation may be greatly diminished by its high sensitivity. Therefore, dose

conformality and plan robustness are two equally critical factors to consider in IMPT treatment planning.

1.3 Beam Orientation Optimization

A typical treatment planning process of IMPT involves three steps: first, a number of proton beams are manually selected by the planner. For complicated cases, multiple tries are needed to decide the number and orientations of beams; second, proton pencil beam dose calculation is performed for these beams; finally, inverse planning is carried out to obtain the intensity map and final dose distribution[6]. The intensity map of the scanning-spots is also referred to as the fluence map. The inverse planning process to obtain the fluence map is called fluence map optimization (FMO).

Different from X-ray therapy where equiangular or arc beams are often acceptable, the proton beam orientations are typically asymmetric, and need to be more carefully considered for factors such as the water-equivalent thickness to the target, nearby OAR sparing, heterogeneity of tissues in the beam path, and setup robustness etc[12–14]. Because of the unique proton beam physics, the need to further reduce body dose, the limited beam time, and to relieve the patient-specific QA efforts, 2~4 proton fields are typically utilized in an IMPT session[6,15]. The smaller number of beams makes the selection of proton beam angle particularly important. In the current clinical practice of manual beam selection, planners' experience and skill can heavily influence the final treatment plan quality. For complicated patient cases, tedious trial-and-error attempts may be needed to find better beam configurations. Yet, human operators cannot effectively search the enormous coplanar and non-coplanar beam space, resulting in inconsistent planning results. Beam

orientation optimization (BOO) using a computational model is therefore essential for improving IMPT.

1.4 Biological Effectiveness

Aside from robustness, the biological effect is another prominent issue to consider for IMPT. While conventional radiation treatment is prescribed based on the physical absorbed dose, it is insufficient for proton. The difference in radiation interactions makes the energy deposition pattern differ between photons and protons; thus, the same absorbed dose can lead to various biological outcomes. The concept of relative biological effectiveness (RBE) was developed to account for the biological effect when comparing different modalities under the same physical dose. RBE is defined as the ratio of the physical doses to reach the same endpoint X when comparing a reference radiation source and proton.

$$RBE(\text{EndpointX}) = \frac{Dose_{reference}(\text{EndpointX})}{Dose_{proton}(\text{EndpointX})}$$

Equation 1-1

As recommended by the International Commission on Radiation Units and Measurements (ICRU)[16], the doses in proton therapy are prescribed as the product of RBE and absorbed dose, with the unit of GyRBE.

In the current clinical application, proton therapy is regarded to be 10% more effective than high-energy photons, represented by a constant RBE of 1.1[17–19]. However, the constant 1.1 value is an average of in vivo measurements performed around 1970s[20–23]. Multiple pieces of evidence have shown that actual RBE depends on multiple factors like

physical dose, tissue radiobiological properties (α and β value), LET (linear energy transfer), endpoint, and et al[17,24–26]. In vitro study suggests that, given dose, cell type and biological endpoint, the RBE may increase from values between 1.0 and 1.1 in the entrance region to values around 1.3 at the Bragg peak and 1.6 in the falloff region[27]. Therefore, even though an ideal physical dose distribution is guaranteed, it is possible to induce higher normal tissue complication probabilities when RBE in normal tissue is underestimated. In addition, with the pencil beam scanning technique replacing passive scattering to be the mainstream delivery modality, the biological doses potentially differ from the previous observation on passive scattering[28,29], warranting further investigation in the universal use of RBE=1.1.

There have been concerns that using the generic RBE value in proton therapy can lead to underdosage in the target or underestimation of the normal tissue toxicities. Therefore, it is essential to incorporate the biological dose into clinical treatment planning to achieve a higher therapeutic ratio. Moreover, the ability to optimize biological effectiveness depends on patient geometry and beam arrangement[30]. For example, if an OAR abuts the target in the distal edge of a proton beam, it is difficult to reduce the biological dose in this OAR without compromising physical dose coverage. In clinical practice, a planner can avoid some of the undesirable beam orientations based on experience[31–34], but evaluating all beam angles for their dosimetry, robustness, and the biological effect is a large computational task unsuited for human operators. A beam orientation optimization algorithm for both physical and biological dose optimization is essential for IMPT.

1.5 Overview

Prior to presenting the beam orientation optimization method integrating dosimetry and robustness of IMPT, separate investigations performed related to IMPT BOO problems will first be discussed.

Chapter 2 describes the unified framework integrating BOO and FMO in IMPT based on the group sparsity regularization, to select beams and generate plans with superior dosimetry. It is the foundation for the rest studies in this thesis. It is a version of a manuscript titled "Integrated beam orientation and scanning-spot optimization in Intensity Modulated Proton Therapy for brain and unilateral head and neck tumors" published in Medical Physics[35].

In Chapter 3, works utilizing the group sparsity for different aspects of proton therapy will be discussed. First is a work using group sparsity to select variant beams in different treatment fractions to further improve the dosimetry of IMPT plans. The manuscript titled "Fraction-Variant Beam Orientation Optimization for Intensity-Modulated Proton Therapy" has been submitted to Medical Physics. The second section is a study integrating biological effectiveness into BOO of IMPT, to achieve superior physical and biological dose distribution. The manuscript titled "Biological effectiveness-integrated beam orientation optimization for Intensity-Modulated Proton Therapy" has been submitted to Medical Physics. The third section describes an optimization method for proton arc therapy based on group sparsity, to improve dosimetry and delivery efficiency from the IMPT. It is a version of the manuscript titled "A novel energy layer optimization framework for spot-scanning proton arc therapy" published in Medical Physics[36].

Chapter 4 describes the robust optimization method with sensitivity regularization and combines it with group sparsity to achieve robust integrated BOO and FMO framework in IMPT. The first section is a version of the manuscript “Robust optimization for Intensity-Modulated Proton Therapy with soft spot sensitivity regularization” published in Medical Physics[37], detailing the formulation of sensitivity regularization. The second section integrates sensitivity regularization and group sparsity for the ultimate goal of robust IMPT BOO, which is a version of the manuscript “Robust beam orientation optimization for Intensity-Modulated Proton Therapy” published in Medical Physics[38].

2 Integrating Beam Orientation and Fluence Map Optimization in IMPT Based on Group Sparsity Regularization

2.1 Introduction

Because of the small beam number in IMPT plans (2~4 proton fields), each beam heavily influences the final quality. The importance of beam orientation was highlighted in several studies[12–14]. However, due to the vast search space, it is essentially impossible for a human operator to test all combinations and find the optimal set of beams. Especially non-coplanar space is commonly used in proton therapy because of the modern robotic patient positioning device typically installed in a proton treatment room. Therefore, there arises a need for beam orientation optimization (BOO) for IMPT.

For IMXT, BOO is viewed as a combinatorial problem, which by its nature is mathematically intractable for realistic BOO problems. The challenge is greater with the additional depth dimension in IMPT optimization. While methods for IMXT BOO have been developed using heuristic and stochastic algorithms to overcome the mathematical challenge[9–15], the BOO problem is rarely touched in IMPT. In a previous study by Cao et al.[46], the coplanar BOO problem was treated as a combinatorial problem. To reduce the problem size, starting from a set of initial beams, a local search was performed to identify beams that improve dosimetric quality and robustness. This method is clearly limited in its ability to perfrom a global search in a large solution space, like the 4π steradians.

In this study, we present a novel framework to efficiently integrate BOO and FMO that allows a global search in all feasible beams. The problem is formulated to include a dose fidelity term and a group sparsity regularization to control the number of active beams. Group sparsity, also known as structured sparsity, was originally used in optimizing X-ray beam orientations and showed the potential of reducing the number of beams yet maintaining dense beamlets[47].

In this work, different regularization and dose fidelity terms are introduced as well as a solver that is capable of handling a larger problem than the original IMXT BOO problem owing to the additional depth dimension.

2.2 Methods

The group sparsity based integrated BOO and scanning-spot optimization problem is formulated under two different dose fidelity terms, with either convex or nonconvex group sparsity term. The details are described as follows.

2.2.1 Problem formulation

The simultaneous beam orientation and scanning-spot intensity optimization problem is formulated under the following general framework:

$$\underset{x}{\text{minimize}} \quad \Gamma(A\boldsymbol{x}) + \sum_{b\in\mathcal{B}} \alpha_b \|\boldsymbol{x}_b\|_2^p ,$$

$$\text{subject to} \quad \boldsymbol{x} \geq 0,$$

Equation 2-1

where the optimization variable $\boldsymbol{x}$ is a vector of the intensities of all scanning-spots, which is the fluence map. A is the dose-calculation matrix that transforms the spot intensities $\boldsymbol{x}$ to dose. A single column of matrix A contains the vectorized doses delivered to the voxels in the patient from one unit intensity spot; and A includes columns of all the candidate spots from all candidate beams. $\mathcal{B}$ is the set including all candidate beams, $\boldsymbol{x}_b$ is a vector of spot intensities for the candidate beam b (so $\boldsymbol{x}$ is the concatenation of the vectors $\boldsymbol{x}_b$), and α_b is the regularization parameter for beam b. $\Gamma(A\boldsymbol{x})$ is a dose fidelity term, to penalize dose deviation from prescriptions. $\sum_{b\in B} \alpha_b \|\boldsymbol{x}_b\|_2^p$ is the group sparsity term to control the number of active beams to between 2 and 4. The L2,p-norm ($0 < p \leq 1$) encourages most candidate beams to be zero, resulting in a small number of beams being selected. A common choice for the exponent for the group sparsity term is $p = 1$, which makes the

objective function convex. However, in the BOO problem, the adjacent block columns of the dose-calculation matrix (corresponding to adjacent candidate beams) can be correlated in homogeneous areas. Subsequently, the group restricted isometry property may not be well satisfied and the L2,1-norm may lead to a degenerate solution, i.e. aggregated beams. To avoid degeneracy, the nonconvex group sparsity with $p = 1/2$ is also explored in this study.

We choose the weighting parameter α_b on the individual beam b to be

$$\alpha_b = c\left(\frac{\left\|A_{\mathcal{T}}^{b}\,\vec{\mathbf{1}}\right\|_2}{n_b}\right)^{p/2},$$

Equation 2-2

where $A_{\mathcal{T}}^{b}$ is the dose-calculation matrix of planning target volume (PTV) for beam b, n_b is the number of candidate spots in beam b, and c is a regularization parameter. The term $\left\|A_{\mathrm{PTV}}^{b}\,\vec{\mathbf{1}}\right\|_2$ in the numerator is used to ensure that beams penetrating different depths in the patient are unbiasedly weighted. Without this term, the group sparsity has a tendency towards only selecting beams passing through less tissue. The denominator n_b prevents the group sparsity penalty from having a bias against the beams with more spots. By this weighting method, we are able to tune a single parameter c to control the number of active beams in the solution to Equation 2-1.

In this work, PTV is designated to be the target volume for dose optimization. Although the concept of PTV originally from X-ray therapy is not exactly applicable in IMPT to maintain tumor coverage[48,49], without losing generality, the term PTV here is used to indicate the target volume without the additional implication of plan robustness.

2.2.2 Dose fidelity

The function Γ can take different forms based on the dosimetric goals and whether a compatible solver exists. Two different choices of Γ are implemented and compared in this work.

The first is the quadratic loss function, which is a common choice for dose fidelity. The overall cost is written as:

$$\Gamma(A\boldsymbol{x}) = \sum_{j\in\mathcal{T}} \omega_j \left\|A_j\boldsymbol{x} - p_j\right\|_2^2 + \sum_{j\in\mathcal{O}} \omega_j \left\|\left(A_j\boldsymbol{x} - m_j\right)_+\right\|_2^2,$$

Equation 2-3

where $\mathcal{T}$ is the structure set of the target volumes, with p_j being the prescription dose to jth target, and $\mathcal{O}$ is the dose-limiting structure set which includes the OARs, with m_j being the prescribed maximal allowed dose to the jth structure. A_j is the dose calculation matrix block for structure j. The dose-limiting penalty terms utilize the one-sided quadratic function$\|z_+\|_2^2$, where $z_+ = \max(z, 0)$. This component-wise maximum allows us to consider only the voxels with doses larger than m_j in the jth OAR. m_j can also be set to 0 to penalize any nonzero dose in an OAR. The weights ω_j are the structure-specific weighting parameters to emphasize the different importance of different structures.

The quadratic penalty is mathematically desirable for being convex and differentiable. It heavily penalizes the dose volumes that exceed the constraint. However, in radiation therapy, more controls on the dose volume behavior are often desirable. For serial organs, such as the spinal cord, hot spots need to be avoided. However, for parallel structures, such as the parotid glands, the mean dose can be more important than the

maximum dose. Therefore, the second type of function Γ we use is a linear combination of structure mean dose and maximum dose penalty, which is referred as the linearized equivalent uniform dose (LEUD)[50] cost function in this work. The LEUD penalty function is formulated as:

$$\Gamma(A\boldsymbol{x}) = \sum_{j\in\mathcal{T}} \frac{\omega_j}{\sqrt{N_j}} \left\| A_j\boldsymbol{x} - p_j \right\|_2 + \sum_{j\in\mathcal{O}} \omega_j \left(\gamma_j \text{mean}\left(A_j\boldsymbol{x}\right) + \left(1-\gamma_j\right)\max\left(A_j\boldsymbol{x}\right) \right) + I_{\leq q}(A\boldsymbol{x}),$$

Equation 2-4

where N_j is the number of voxels in jth target, and the weighting factor $\gamma_j \in [0,1]$ balances the mean dose and maximum dose for different organs. The penalty on the target is a voxel-normalized L2-norm. The L2-norm is chosen because it has better coverage than the L1-norm (which is $\text{mean}\left(\left|A_j\boldsymbol{x} - p_j\right|\right)$), and is consistent with the OAR cost with regard to the order (the quadratic term is in the order of 2, while the mean dose is linear). Because the L1-norm on OARs (mean dose) is not as sensitive to outliers, it allows for hot spots in parallel OARs. Therefore, in order to eliminate any dose higher than the upper bound dose q, an upper bound constraint $I_{\leq q}(A\boldsymbol{x})$ is added to the objective function. $I_{\leq q}(z)$ is defined as:

$$\boldsymbol{I_{\leq q}(z)} = \begin{cases} \mathbf{0} & \textbf{if} \quad \boldsymbol{z \leq q}, \\ \infty & \textbf{otherwise}. \end{cases}$$

Equation 2-5

2.2.3 Evaluations

Four cases, which included three unilateral head and neck (H&N) patients and one skull base chordoma (CHDM) patient with simultaneous-integrated boost (SIB), were evaluated in this study. The candidate beams included 1162 non-coplanar beams that were

evenly distributed across the 4π steradians with $6°$ separation. Geometrically undesired beams and beams of infeasible energies, such as those directed through the feet to the head, were manually excluded from the candidate set, resulting in about 700 to 800 candidate beams for each patient. For each candidate beam, the doses of all scanning-spots covering the PTV and a 2.5mm margin were calculated using matRad[51,52], a MATLAB-based 3D treatment planning toolkit. The dose calculation resolution was 2.5 mm with a cut-off of 5×10^{-5} of the maximal dose. The prescriptions, PTVs, and number of scanning-spots are shown in Table 2-1. The magnitude of the optimization problem can be estimated based on the product of the spot number per beam and the total number of candidate beams. The average number of beamlets per beam needed in IMXT for the same patient is also given in Table 2-1 as a comparison of the BOO problem size in IMPT and in IMXT. The IMXT multileaf collimator resolution is 5 mm.

Case		Prescription Dose (GyRBE)	PTV Volume (cc)	Average Spots Number per Beam	Average IMXT Beamlets Number per Beam
H&N #1		40	23.76	906	85
H&N #2		40	32.29	1109	103
H&N #3		66	33.64	1589	111
CHDM	PTV6300	63	86.07	3166	241
	PTV7400	74	26.42		

Table 2-1: Prescription doses, PTV volumes, and average spots number per beam of each tested patient.

A greedy BOO approach, column generation[53,54], was also applied for each patient, as a comparison for our group sparsity method.. The dose fidelity term used in the column generation method is LEUD cost.

The BOO plans were evaluated against plans with manually selected beams. The manual beam orientations were selected to avoid OARs as much as possible. In total, 7

plans were generated for each patient: three plans with quadratic dose fidelity: manual plan (Quad-MAN), L2,1/2-group sparsity (Quad-L2,1/2-GS) and L2,1-group sparsity(Quad-L2,1-GS); and four plans with LEUD dose fidelity: manual plan (LEUD-MAN), L2,1/2-group sparsity(LEUD-L2,1/2-GS), and L2,1-group sparsity (LEUD-L2,1-GS), and column generation (LEUD-CG). All H&N plans were normalized so that 100% of the prescription dose covers 95% of the PTV volume. The CHDM plan with a simultaneous boost volume was normalized to have 100% of the 63Gy prescription dose covering 95% of the PTV6300 volume.

For evaluation, PTV homogeneity, D95, D98, D99, maximum dose and mean dose were evaluated. PTV homogeneity is defined as D95/D5. The maximum dose is defined as the dose to 2% of the structure volume, D2, following the recommendation by IRCU-83[55]. The mean and maximum doses for OARs were also evaluated.

2.3 Results

The matRad-based dose calculation for all candidate beams using an i7 6-core CPU and Matlab parallel computing toolbox took 30 min to 1 h depending on the size of the tumor. The dose matrix size ranged from 20 MB to 70 MB per beam based on target size. The group sparsity based BOO process took 2 - 7 minutes and 3 - 20 minutes for the quadratic and LEUD cost to complete, respectively.

Three beams for each H&N case and four beams for the CHDM case were selected. Figure 2-1 shows the beam arrangement of each plan for the CHDM patient. The beams selected by using the L2,1/2-GS term were spatially well separated. In comparison, the

L2,1-GS term resulted in aggregated beams with both quadratic and LEUD cost terms, indicating a potential degeneracy issue with this group sparsity term. And the CG method tended to choose beams with short pathlengths to target. Interestingly, the beam orientations optimized by L2,1/2-GS were similar to the actual angles selected by an experienced dosimetrist in this CHDM case.

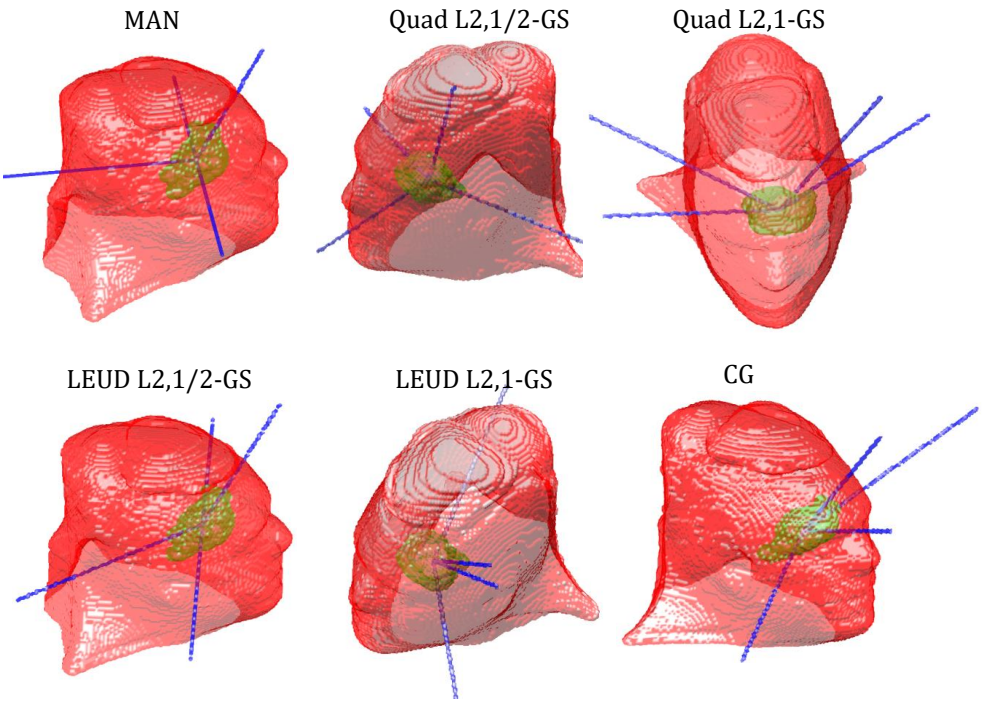

Figure 2-1: The beam arrangement of each plan for the CHDM patient.

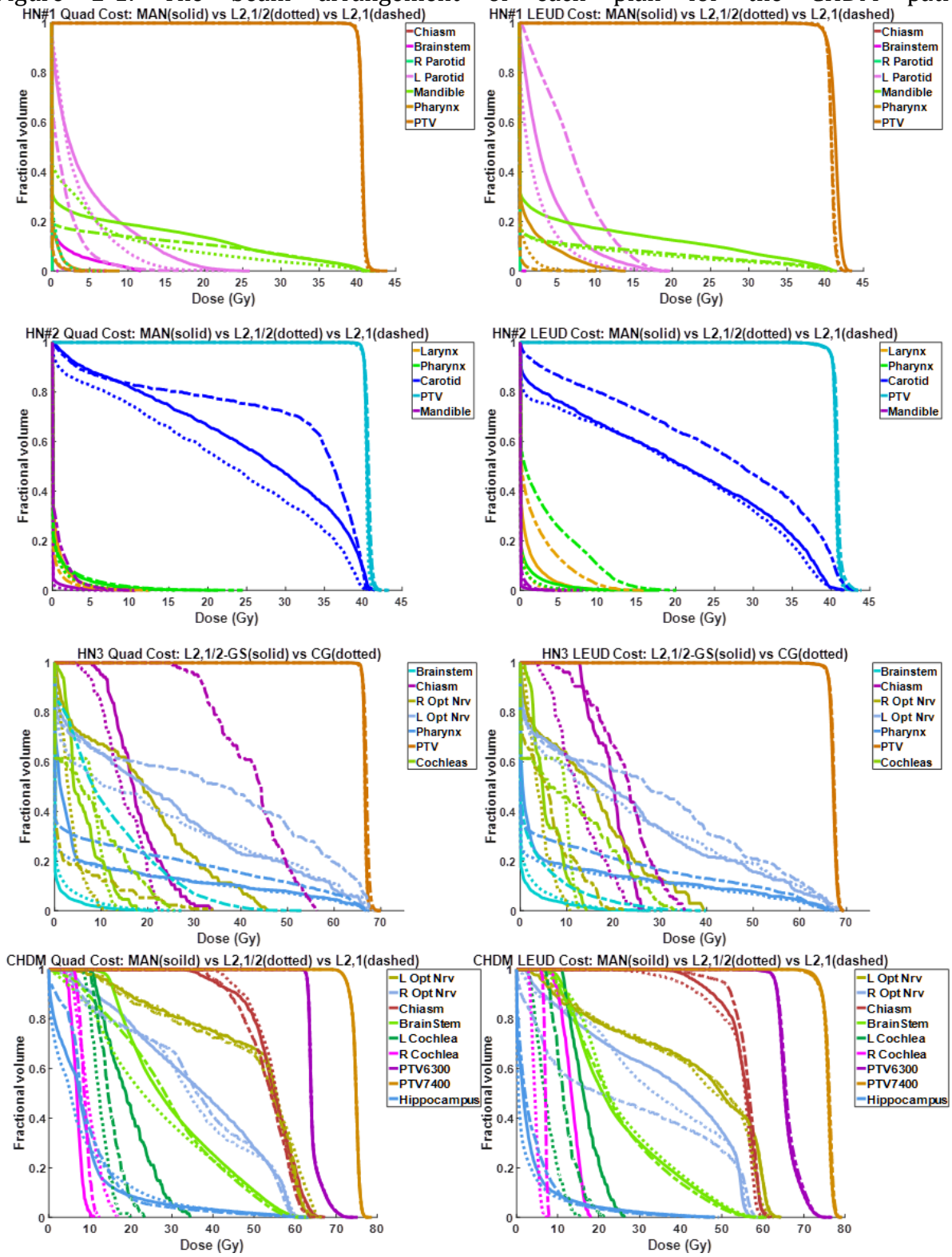

Figure 2-2 shows the DVHs comparison between GS plans and MAN plan for each patient. Table 2-2 and Table 2-3 show the OAR statistics for each patient with quadratic cost and LEUD cost, respectively. On average, the L2,1/2-GS plans reduced the OAR [Dmean, Dmax] from MAN plans by [2.4%, 4.2%] and [2.3%, 3.8%] of the prescription dose for Quad and LEUD cost, respectively, while achieving comparable target coverage.

The L2,1-GS method produced competitive plan for H&N#1 and CHDM case, but led to worse and OAR dose over MAN plan for H&N#2 and H&N#3 case with similar PTV coverage. The average decrease in OAR [Dmean, Dmax] from MAN plans are [-1.0%, 0.1%] and [-1.2%, 1.2%] of the prescription dose for Quad and LEUD cost, respectively.

Case	L2,1/2 - MAN (GyRBE)				L2,1 - MAN (GyRBE)			
	Dmean		Dmax		Dmean		Dmax	
	Largest Value	Average Value	Largest Value	Average Value	Largest Value	Average Value	Largest Value	Average Value
H&N#1	-1.2 L Parotid	-0.5	-8.1 Brainstem	-2.8	-3.1 L Parotid	-0.5	-10.4 L Parotid	-2.6
H&N#2	-3.8 Carotid	-1.0	-2.1 Pharynx	-1.0	-0.2 Larynx	+0.8	-1.5 Larynx	+0.9
H&N#3	-20.1 TMJ	-3.0	-34.6 R Opt Nrv	-3.9	-26.3 TMJ	+1.8	-35.1 TMJ	+3.7
CHDM	-7.0 L Cochlea	-0.7	-16.4 L Cochlea	-1.0	-4.5 L Cochlea	+0.4	-11.2 L Cochlea	-1.2

Table 2-2: The difference of OAR doses between GS and MAN plans for all patients under Quad cost. A negative sign represents a reduction of dose in GS plan from MAN plan.

Cases	L2,1/2 - MAN (GyRBE)				L2,1 - MAN (GyRBE)			
	Dmean		Dmax		Dmean		Dmax	
	Largest Value	Average Value	Largest Value	Average Value	Largest Value	Average Value	Largest Value	Average Value
H&N#1	-1.5	-1.0	-6.75	-2.8	-0.98	-0.1	-8.9	-1.9

	L Parotid		Pharynx		Pharynx		Pharynx	
H&N#2	-0.8 Carotid	-0.3	-3.75 Larynx	-1.1	0.04 Mandible	+2.2	1.0 Mandible	+4.1
H&N#3	-16.2 TMJ	-2.5	-28.83 R Opt Nrv	-2.3	-19.52 TMJ	+0.8	-21.0 R Opt Nrv	+1.0
CHDM	-8.8 R Cochlea	-1.4	-10.99 R Cochlea	-1.0	-8.89 R Opt Nrv	-1.3	-9.6 R cochlea	-1.3

Table 2-3: The difference of OAR doses between GS and MAN plans for all patients under LEUD cost. A negative sign represents a reduction of dose in GS plan from MAN plan.

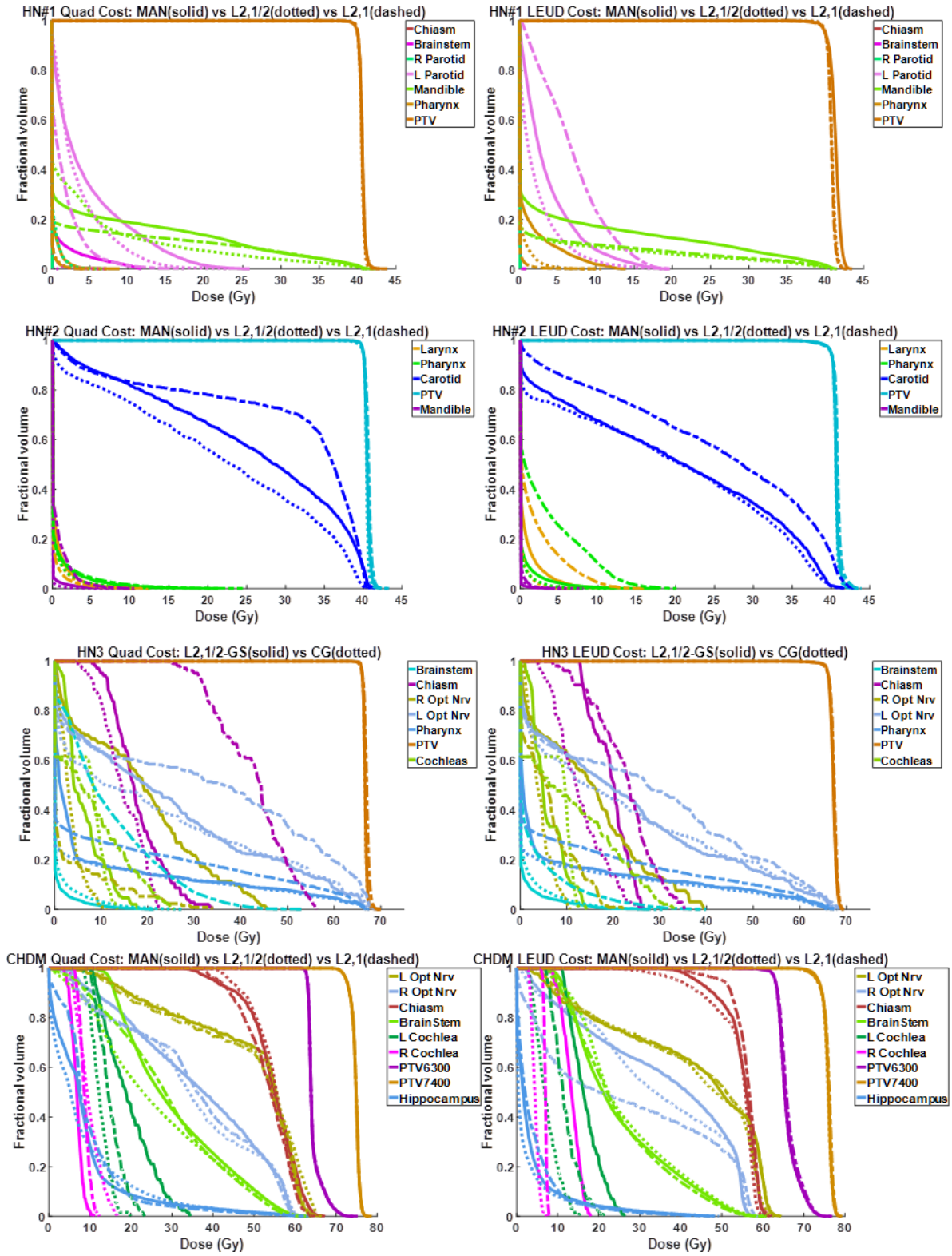

Figure 2-2: DVH comparison of the MAN plan (solid line), L2,1/2-GS plan (dotted line) and L2,1-GS plan (dashed line) for each patient. The plans with quadratic cost are listed on the left column and the plans with LEUD cost are on the right column.

The limitation of using L2,1-norm can be observed from the final value of dose fidelity cost. Figure 2-3 compares the convergence between L2,1/2-norm and L2,1-norm group sparsity for the H&N#3 case. For the quadratic and LEUD cost functions, the L2,1/2-norm group sparsity method converged after about 800 iterations and 1800 iterations respectively. In contrast, with an L2,1-norm group sparsity term, the problem quickly converged within 30 iterations and 80 iterations under the quadratic and LEUD cost functions, respectively. However, the converged dose fidelity function values using the L2,1-norm are substantially greater than the corresponding values of using the L2,1/2-norm ($\sim 10^5$ for L2,1 vs. $\sim 10^2$ for L2,1/2), showing a larger deviation from the prescription dose using L2,1-norm in this case.

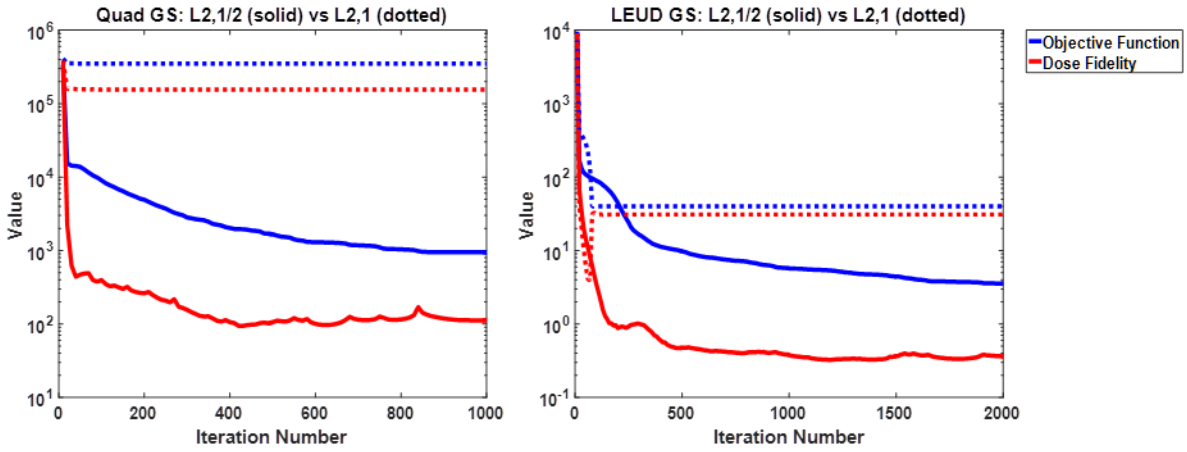

Figure 2-3: The convergence comparison between L2,1/2-norm (solid) and L2,1-norm (dotted) group sparsity for case H&N#3. Left is Quad-GS and right is LEUD-GS. The blue curve shows the value of entire cost function and the red curve shows the value of dose fidelity term.

The DVHs comparison between L2,1/2-GS plans and CG plans is shown in Figure 2-4. In all H&N cases, the CG method produced plans with similar PTV coverage as L2,1/2-GS method, but the OAR doses were inferior to the group sparsity algorithm. In the CHDM case, the CG plan performed comparably with the L2,1/2-GS plan with regard to PTV and

OAR dose. The CG methods reduced the mean dose to brainstem by 9.8Gy over L2,1/2-GS plan by forcing all beams entering from the anterior direction, increasing the risk of exposing the eyes to high dose with slight positioning error.

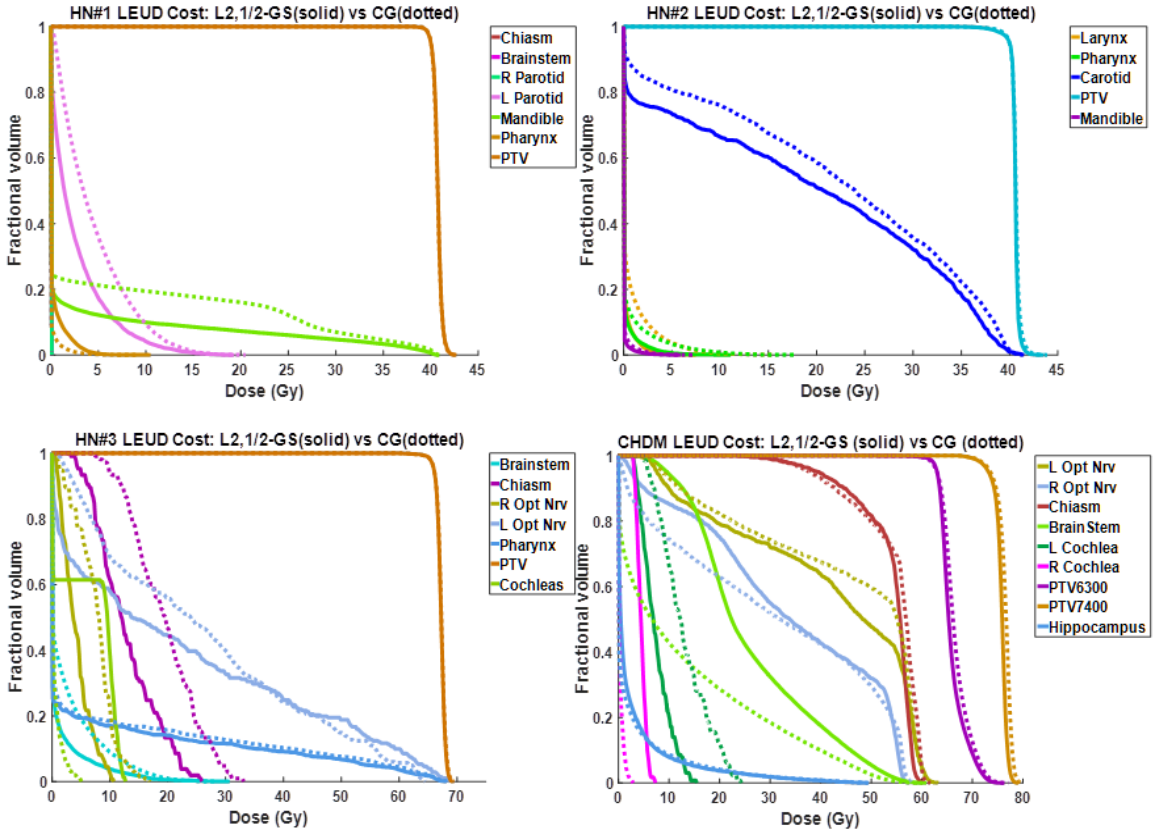

Figure 2-4: DVH comparison between L2,1/2-GS plan (solid) and the CG plan (dotted) for each patient.

2.4 Discussion

This work introduces a group sparsity based IMPT optimization method that simultaneously selects beams and optimizes the fluence map. In addition to the superior dosimetry compared with plans using manually selected beams, the planning method using optimized beams reduces the dependence on the individual operators who select beams based on experience and intuition.

Compared to the earlier coplanar X-ray optimization study using group sparsity[47], the current work expands its scope in several ways. This is a new application to the IMPT problem, which is intrinsically a higher-dimensional optimization problem than IMXT due to the additional modulation in the depth direction. The number of candidate beams in this study is an order of magnitude larger than in the original IMXT BOO study that used only 72 candidate beams. The optimization solver used in the original study was based on ADMM and is not well suited to the current much larger IMPT BOO problem, because the linear systems that ADMM requires to be solved at each iteration would be intractably large. Therefore, we developed an approach based on FISTA, which requires only matrix-vector multiplications involving the dose-calculation matrix. To enable the use of FISTA, we derived several key proximal operators. FISTA has been shown to have an optimal convergence rate of $O(1/k^2)$ among first-order methods[56], and the numerical results demonstrate that FISTA is able to solve the large scale optimization problem in a clinically acceptable time. Although standard convergence results for FISTA assume that both f and g are convex[56], we have found that FISTA converged to a good solution even in the case of using the nonconvex L2,1/2-norm group sparsity penalty.

In this study, two different group sparsity terms, L2,1 and L2,1/2 norms, were compared for the IMPT BOO problem. Although the L2,1-norm is convex and offers certain computational advantages, it leads to suboptimal dosimetry and tends to converge to a worse fidelity value when compared against the nonconvex L2,1/2-norm group sparsity term. Therefore, in this case, the ability to avoid degeneracy and select spatially separate beams appears to be dosimetrically advantageous. This finding shows that while the group

sparsity method is well suited to solve the BOO problem, the selection of the sparsity function is a subtle point that can make a critical difference.

The group sparsity beam orientation optimization method was tested on two different dose fidelity functions: quadratic and LEUD terms. Compared to the quadratic dose fidelity term, the LEUD cost function with a combination of mean and maximum dose constraints is easier to tune to achieve desired DVHs. This is reflected in the process of creating plans for the four cases in this study: the quadratic L2,1/2-GS method took on average 9-10 rounds of parameter tuning, and the LEUD method took 4-5 rounds. The advantage of quadratic dose fidelity function is that while the structure parameters need to be tuned, the group sparsity regularization weighting parameter remains constant for a specific number of beams. In comparison, with LEUD dose fidelity, the group sparsity regularization weight is sensitive to the structure weighting parameter changes, requiring additional adjustment to maintain the desired number of beams. Quadratic dose fidelity also resulted in, on average, more sparse scanning spots.

The group sparsity BOO method was also compared against the greedy column generation BOO algorithm. The results show the GS method produces comparable or superior plans over CG. Specifically, the CG method tends to select aggregated beams with short pathlengths to the target, such as the anterior beams in the CHDM case. The result indicates degeneracy that did not present in the IMXT BOO solution, possibly due to the substantial difference in dose distribution between the first proton beam and photon beam. Additional heuristics, such as minimal separating between selected beams, may be enforced to ameliorate the problem.

The dose matrix size is determined by the target, which is limited to be approximately 110 cm^3 to fit the calculation into a desktop with 64GB memory. For more general IMPT cases with larger tumor targets, either workstations with substantially larger memory or methods to intelligently reduce the dose matrix size without impacting plan quality are needed. These methods include (1) non-uniform sampling resolution with a higher resolution in the target and nearby organs, and lower resolution in the volumes that are considered less critical and faraway from the target, (2) clustering of the proton pencil beam dose matrices, and (3) using heuristics to reduce the number of candidate beams. We will investigate these directions for more general integrated BOO and scanning-spot optimization problems.

2.5 Conclusion

This work shows the first IMPT planning approach that integrates non-coplanar beam orientation and fluence map optimization in a single mathematical framework, which was further formulated to have a computationally efficient solution despite its large problem size. This method resulted in dosimetrically competitive plans compared with the manual planning method and is less operator-dependent. It sets up the framework for the optimization problems that are handled in the subsequent studies.

3 APPLICATIONS OF GROUP SPARSITY REGULARIZATION FOR PROTON THERAPY

3.1 Fraction-Variant Beam Orientation Optimization for IMPT

3.1.1 Introduction

Different from X-ray therapy that found success in using more beams and arcs, proton therapy is conventionally limited to fewer beams due to three reasons. First, because of the unique physics of proton beams, it is feasible to achieve acceptable normal tissue sparing using fewer beams[6,15]. Second, using fewer beams spares larger normal tissue volumes from low-dose radiation exposure, which is one of the main benefit of proton therapy. Third, the proton treatment time is expensive; thus, it is economically desirable to use as few beams as possible to improve throughput. In today's practice, it is typical to use two to four beams in IMPT.

Nevertheless, sparing of organs *adjacent* to the targets can still benefit from using more proton beams. Studies have shown better OARs sparing in three to four-beam plans compared with 2-beam plans[46]. The addition of non-coplanar beams would also improve the plan quality[35]. Further increase in the number of beams to form arc delivery may lead to substantial improvement in dose conformity and adjacent organ sparing[36,38,57–64]. On the other hand, using a larger number of beams in IMPT can be impractical. Spot-scanning proton arc therapy is not clinically deliverable and will likely be limited to selected proton delivery systems. For static beam IMPT, the time to deliver more than four beams in single treatment fraction is increasingly unaffordable, particularly in multi-room configurations when proton beams are shared among different treatment rooms or when increased setup time is needed for non-coplanar beams.

One solution to the apparent conflict between the number of beams and delivery efficiency is to use different groups of beams on different treatment days. For IMXT, approaches have been developed to allow for a large number of beams in the entire treatment while limiting the number of beams in a single fraction. Dink et al[65] proposed a field rotation method, in which multiple sets of coplanar beams were interchanged between fractions using a mixed integer linear program. Results showed an improvement of dose objectives based on this time-varying method. O'Connor et al[66] proposed to use group sparsity regularization to select different beam angles for different fractions out of a large number of non-coplanar candidates. This fraction-variant beam orientation optimization scheme showed that the same dosimetry as conventional plans could be achieved using half as many beams per fraction.

In this study, we aim to investigate the dosimetric and delivery efficiency benefit of fraction-variant beam orientation optimization in IMPT and its potential to improve treatment delivery efficiency.

3.1.2 Methods

The goal of fraction-variant BOO (FVBOO) framework is to select a small number of beams from the candidate set for each fraction in a single step, but allow variant beam angles from fraction-to-fraction. More importantly, the target dose homogeneity is maintained regardless of the number of beams per fraction to avoid undesirable hot or cold spots within each fraction.

3.1.2.1 Problem formation

Let F indicate the number of treatment fractions and assume each fraction sharing the same candidate beam set. The number of candidate beams in each fraction is B. The problem is formulated as the following objective function:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \sum_{i\in\mathcal{T}} \omega_i \left(\sum_{f=1}^{F} \left\| A_i \boldsymbol{x}_f - p_i/F \right\|_2^2 \right) + \sum_{i\in\mathcal{O}} \omega_i \left\| \left(\sum_{f=1}^{F} A_i \boldsymbol{x}_f - m_i \right)_+ \right\|_2^2 + \sum_{f=1}^{F} \sum_{b=1}^{B} \alpha_b \left\| \boldsymbol{x}_{f,b} \right\|_2^{\frac{1}{2}}$$

$$\text{subject to} \quad \boldsymbol{x} \geq 0.$$

Equation 3-1

In Equation 3-1, $\boldsymbol{x}_{f,b}$ is a vector representing the intensities of scanning spots of the bth candidate beam in the fth fraction. Then $\boldsymbol{x}_f$ is the concatenation of the vectors $\boldsymbol{x}_{f,b}$

(for $b = 1, \dots, B$), representing the spot intensities of fraction f, and $\boldsymbol{x}$ is the concatenation of the vectors $\boldsymbol{x}_f$ (for $f = 1, \dots, F$), indicating the entire intensity map.

A_i is the dose–calculation matrix for structure i ($i \in \mathcal{T}$ or $\mathcal{O}$, where $\mathcal{T}$ is the set of target volumes, and $\mathcal{O}$ is the set of OARs). Each column of A_i is the vectorized dose delivered to structure i from one scanning spot of unit intensity. Therefore, the product of A_i and $\boldsymbol{x}_f$ is the dose delivered to structure i in the fth fraction.

Equation 3-1 includes three terms. The first term is a fractional dose fidelity term on target volume. If the prescription dose to structure i ($i \in \mathcal{T}$) of the entire treatment is p_i,the first term intends to penalize any dose deviation in *each fraction* from p_i/F, which is the prescription dose per fraction. The second term captures cumulative dose fidelity on OARs, to encourage the dose to structure i ($i \in \mathcal{O}$) over the *entire treatment* not to exceed a maximum value m_i. m_i can also be set to 0 so that any nonzero doses in the OARs are penalized. ω_i is a structure-specific weighting parameter.

The third term in Equation 3-1 is the non-convex group sparsity term developed in Chapter 2, used for beams selection. It groups the spots in the same candidate beam and same fraction in a L2,1/2 -norm term. With a proper value of weighting hyperparameter α_b of each beam b, most of the candidate beams in each fraction will be turned off, leaving a small active set, consisting of 1-4 active beams. Moreover, the non-convexity of the L2,1/2-norm allows varying beam combinations in different fractions, making the BOO fraction-variant.

3.1.2.2 Evaluations

This FVBOO method was tested on one patient with the base-of-skull tumor (BOS), one bilateral head-and-neck (H&N) patient, and one esophageal cancer (ESG) patient. For all patients, the treatment includes 30 fractions. For the BOS patient, an additional plan with 5-fractions (5f) was generated. This BOS case was not clinically treated with 5 fractions, but the 5-fractions plans were created to test the ability of the FVBOO method for hypofractionated treatments. The initial candidate beams included 400~800 non-coplanar candidate beams. For each candidate beam, dose calculation for the scanning spots covering the PTV and a 5 mm margin was performed using matRad[51,52]. The dose calculation resolution was 2.5×2.5×2.5 mm. The prescription dose, target volume, and the number of fractions for each patient are shown in Table 3-1.

Case		Prescription Dose (GyRBE)	PTV Volume (cc)	Number of fields in the FIBOO plan	Number of fractions
BOS		56	66.8	4	5, 30
H&N	PTV54	54	179.1	3	30
	PTV60	60	204.4		
ESG		50	480.9	3	30

Table 3-1: Prescription doses, PTV volumes, number of fields, and number of fractions for each patient

To determine the effectiveness, the FVBOO plans were compared with the plans with fixed beams throughout the treatment. The fixed beam plans, termed Fraction-Invariant BOO (FIBOO) plans, were created using the group sparsity based BOO algorithm proposed in Chapter 2. Different from FVBOO that penalizes the dose of PTV in each individual fraction, FIBOO penalizes the cumulative PTV dose of all fractions. The number of beams per fraction used in the FIBOO plans is also listed in Table 3-1.

To match the number of beams in the FIBOO plans, the parameter α_b in Equation 3-1 was tuned in the FVBOO plans. Different from the integer beam numbers in FIBOO, the average beam numbers in the matching FVBOO plans are either the same or slightly lower. FVBOO plans with further reduced average beams per fraction were also generated in either the 5f or 30f setting for the BOS patient to determine the feasibility of creating more efficient IMPT plans without compromising dosimetry.

3.1.3 Results

3.1.3.1 Runtime and selected beams

FVBOO planning was performed using an i7 CPU desktop at 4.2 GHz clock. The resultant average number of selected beams and the BOO runtime of each method are listed in Table 3-2.

For the BOS patient, under the 5-fractions (5f) setting, three FVBOO plans were generated, with an average number of beams per fraction (b/f) of 3.8, 3, and 1.8, respectively. The 3.8 b/f plan used a total of 19 beams, among which, 12 and 7 are used in single and multiple fractions, respectively. The 3 b/f plan reduced the number of total beams to 15 with 11 single-use beams. The 1.8 b/f plans further reduced the number of total beams to 9, respectively with 7 single-use beams.

In the 30-fractions (30f) setting, four FVBOO plans were generated for the BOS patient, with an average number of beams per fraction being 3.6, 2.8, 2, and 1.3, respectively. In the 3.6 b/f plan, 109 beams are selected with 22 being single-use. In the 2.8 b/f, 2 b/f and 1.3 b/f plans, the total number of beams reduced to 84, 61 and 40, respectively, with 20 to 13 single-use beams.

For the H&N and ESG patient, the 30f-FVBOO plan selected 90 (24 single-use) beams and 84 (36 single-use) beams, respectively.

As shown in Table 3-2, compared with the FIBOO, the runtime of FVBOO selecting similar number of beams per fraction is approximately 4 time longer for the 5f plans and 9-20 times longer for 30f plans. In the worst case for the 30f ESG FVBOO plan with a PTV volume of 480cc, the runtime is close to 10 hours. The runtime was shortened by up to 40% when the goal was to achieve fewer total and per fraction beams in FBVOO.

Case	Num. fractions	Method	Num. beams selected			BOO runtime (min)
			Total	Unique	Avg. per fraction	
BOS	5	FIBOO	20	4	4	13.4
		FVBOO	19	12	3.8	52.4
			15	11	3	36.5
			9	7	1.8	33.6
	30	FIBOO	120	4	4	13.4
		FVBOO	107	22	3.6	129.9
			84	20	2.8	118.5
			61	20	2	95.2
			40	13	1.3	72.3
HN	30	FIBOO	90	3	3	25.1
		FVBOO	90	24	3	500.8
ESG	30	FIBOO	90	3	3	34.0
		FVBOO	84	36	2.8	590.0

Table 3-2: Number of fractions, total number of beams selected, number of unique beams selected, average number of beams selected per fraction, and BOO runtimes.

3.1.3.2 FIBOO and FVBOO with similar number of beams per fraction

The FIBOO and FVBOO plans with a similar number of beams per fraction are first compared. The cumulative dose distributions for the 5f and 30f FVBOO plans are shown in Figure 3-1. These plans with different BOO methods and different numbers of active beams achieved similar PTV dose coverage. Several OARs are selected and the differences of their

mean and maximum doses between the FVOO plans and the FIBOO plans are presented in Figure 3-2.

For the 5f setting of the BOS patient, the 3.8 b/f plan achieves better OARs sparing compared with the FIBOO plan using four beams. The average reduction of [Dmean, Dmax] of the 3.8 b/f plans from the FIBOO plans were [0.9, 2.1] GyRBE.

For the 30f setting, the FVBOO plans of each patient achieve superior OARs sparing compared with the FIBOO plan with a similar number of beams per fraction, with an average reduction of [Dmean, Dmax] of [1.9, 4.1] GyRBE. The maximal reduction to the Dmax for each patient is right optical nerve (10.4 GyRBE) for the BOS patient, spinal cord (10.9 GyRBE) for the H&N patient and trachea (9.8 GyRBE) for the ESG patient.

The volumes of the patient body irradiated by 2, 5, 10 and 20 GyRBE are listed in Table 3-3. As expected, with more beams being used, the FVBOO plans resulted in a larger volume being irradiated by the 2 GyRBE low dose. However, V5 is actually lower with FVBOO in the H&N and ESG cases.

Case	Plan	fractions	V2	V5	V10	V20
BOS	FIBOO	5,30	1165.9	438	343.2	189
	FVBOO	5	909.5	639.5	370.9	156.5
		30	962.9	683.1	367.7	162.5
HN	FIBOO	30	3819.7	3359.4	2650.4	1672.2
	FVBOO	30	4519.6	3315.6	2440.3	1641
ESG	FIBOO	30	5501.3	4757.5	3702.3	1748.2
	FVBOO	30	9782.9	4525.2	2358.6	1265.8

Table 3-3: The V2, V5, V10, V20 GyRBE to the body in volume (cc).

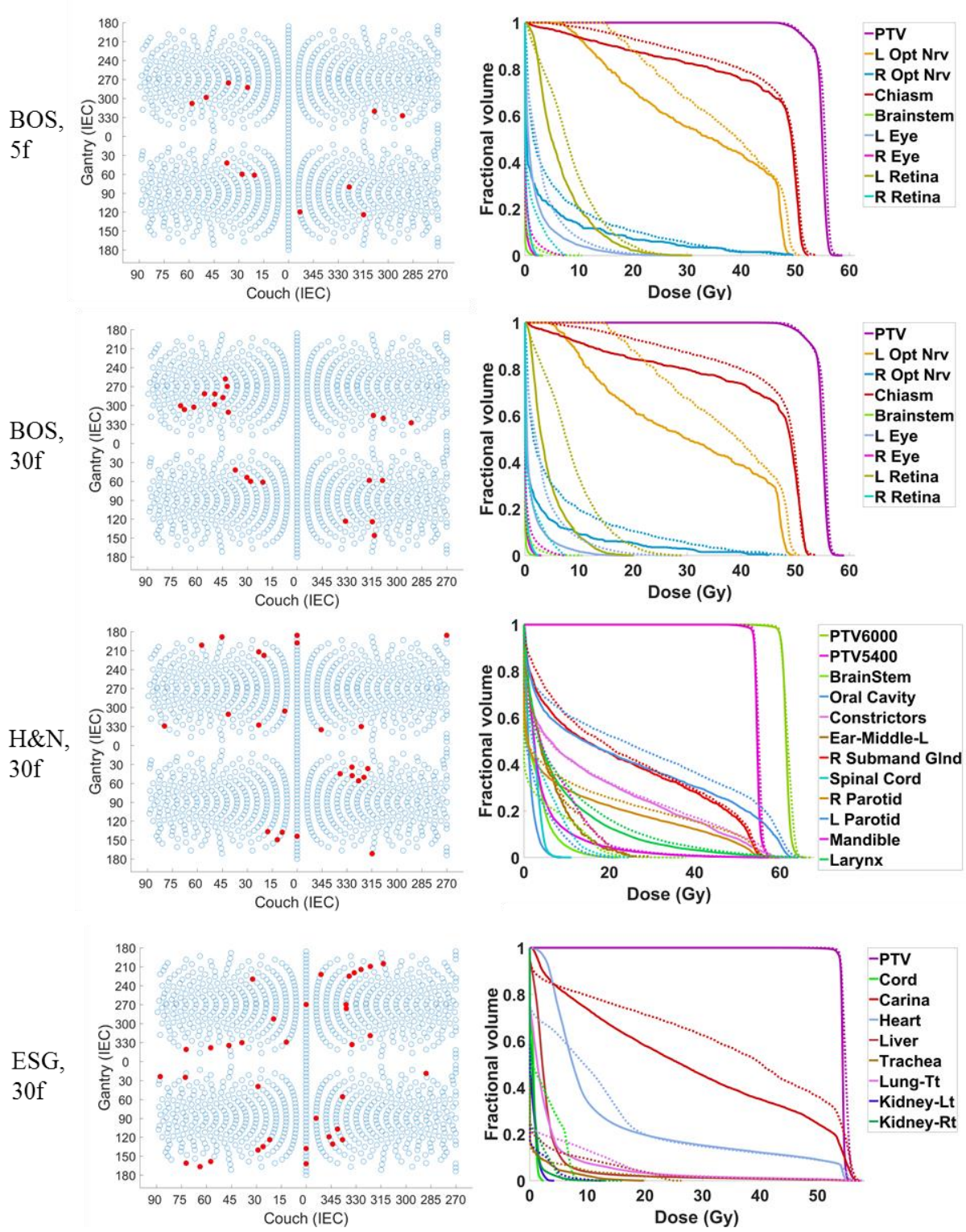

Figure 3-1: The beam angles of the 30f FVBOO plans (left), and the DVH comparison of the 30f FVBOO plans (solid) with the FIBOO plan (dotted) for the three patients. The FVBOO plans have the similar number of beams per fraction as FIBOO.

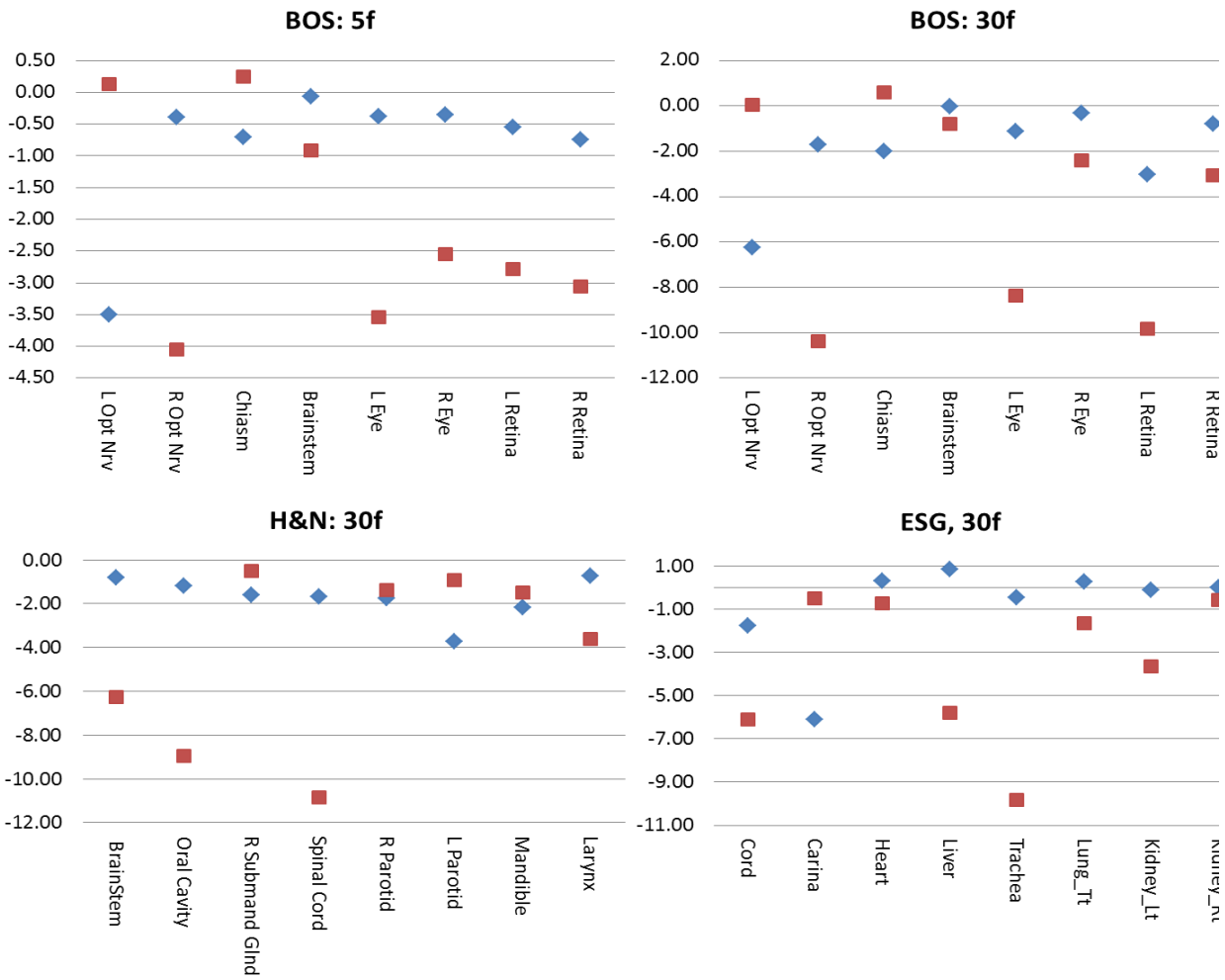

Figure 3-2: The difference of OAR Dmean (blue) and Dmax (red) in GyRBE of the FVBOO plans from the FIBOO plans. A negative value represents a reduction from the FIBOO plan, and a positive value represents an increase.

In addition to the cumulative doses, the dose distributions of individual fractions are presented. The DVH for each fraction of the 5f plans for the BOS patient and the first 5 fractions of the 30f plans for every patient are shown in Figure 3-3. In Figure 3-3, the dotted lines are the fractional dose. The solid lines are the cumulative dose from all fractions re-scaled to the fractional prescription dose (divided by the number of fractions), drawn for comparison. Consistent with our planning goal, despite varying beams, the PTV is covered by a homogeneous dose in individual fractions. While all the cumulative plans

are normalized so D95% = 100% of the prescription dose, the D95% of CTV in all fraction has an average value of 99.6% and a standard deviation of 0.4%.

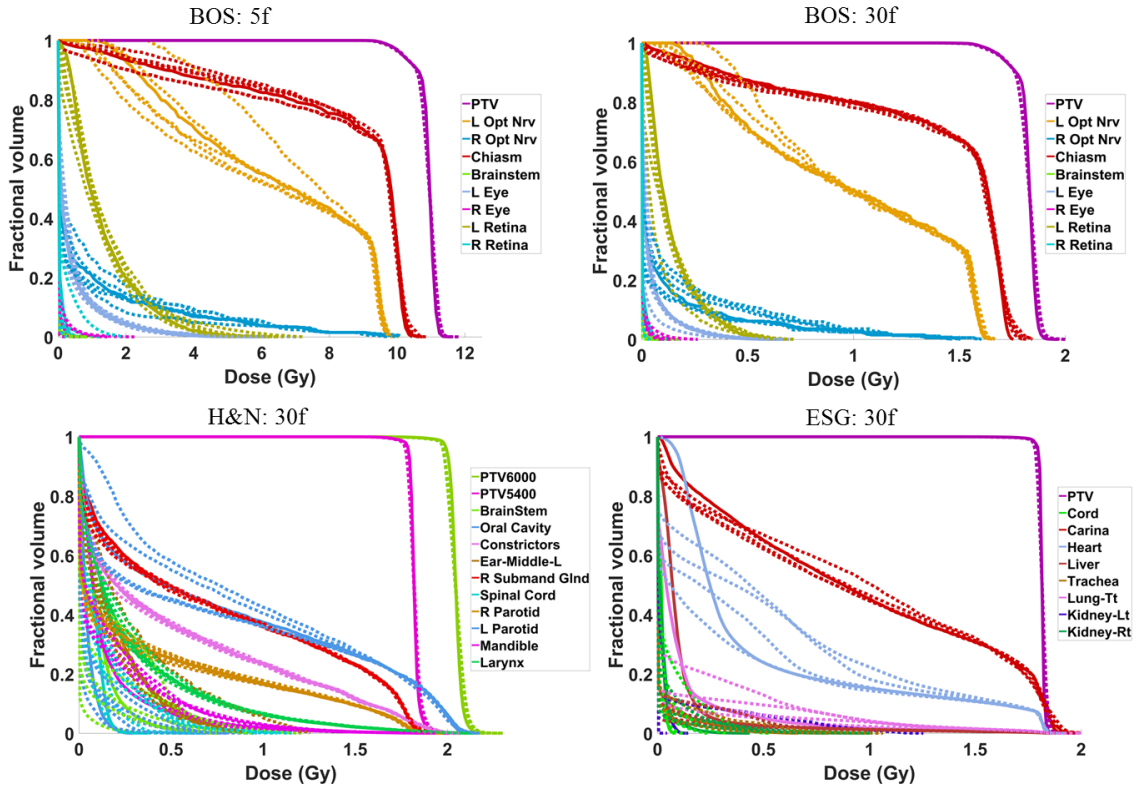

Figure 3-3: DVH of each fraction in the FVBOO 5-fractions plan of the BOS patient and the DVH of the first 5 fraction in the FVBOO 30-fractions plan of each patient. The dotted lines are the fractional dose. The solid lines are the cumulative dose from all fractions re-scaled to the fractional prescription dose.

3.1.3.3 Reduce the number of beams in FIBOO

For the BOS patient, the number of beams per fraction is further reduced from around 4 to around 1. The DVH and dose wash comparison with decreasing number of beams are shown in Figure 3-4. For the 5f setting, while the 3.8 b/f plan achieves the best OARs sparing, with an average reduction of [Dmean, Dmax] from the FIBOO plans of [0.85, 2.08] GyRBE, the FVBOO plan with 3b/f achieves slightly better OARs sparing compared with the FIBOO, with an average reduction of [Dmean, Dmax] of [0.17, 1.45] GyRBE. The 1.8

b/f plan has worse OAR doses than FIBOO plan, except the sparing to the right eye and right retina.

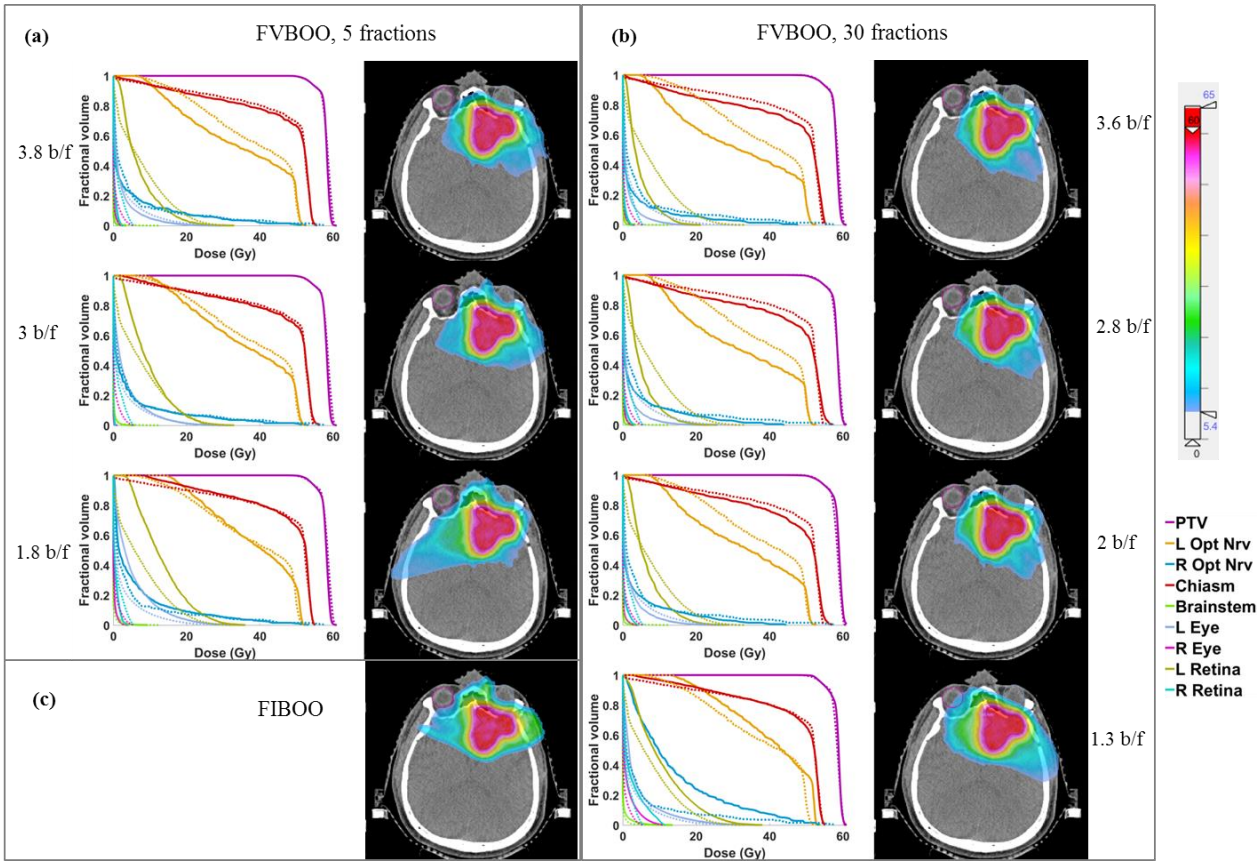

Figure 3-4: The DVH and dose wash using different number of beams for the BOS patient. (a) The DVH comparison of the 5f FVBOO plans (solid) with the FIBOO plan (dotted), and the cumulative dose distribution of the 5f FVBOO plans. (b) The DVH comparison of the 30f FVBOO plans (solid) with the FIBOO plan (dotted), and cumulative dose distribution of the 30f FVBOO plans. (c) The dose wash of the FIBOO plan.

For the 30f setting, when further reducing the beams per fraction, to 2.8 and 2, the FVBOO plan still achieves better OARs sparing compared with the FIBOO plans with more beams, with an average reduction of [Dmean, Dmax] of [1.7, 4.1] GyRBE for the 2.8 b/f plan and [1.3, 3.2] GyRBE for the 2 b/f plan. The average beam number of 1.3 is too low for FVBOO to compete with FIBOO using 4 beams.

3.1.4 Discussion

While traditional fractionated IMPT plans use fixed fields and fluence map to deliver the same dose distribution to the patient during the treatment sessions, this work describes a novel method to treat different fractions with varying beam angles and fluence maps. The fraction-variant beam orientation and fluence map are simultaneously obtained using an optimization framework integrating modified dose fidelity terms and a group sparsity term for beam selection. The FVBOO framework generates fractionated plans that allow the OAR doses to vary for superior cumulative sparing without compromising PTV dose homogeneity in individual fractions. Different from proton arc therapy relying on the proprietary hardware and control system, FVBOO can be delivered on all existing proton systems capable of IMPT.

Besides the ability to improve the cumulative dose distribution using a similar number of beams, FVBOO can reduce the number of beams used per fraction without compromising dosimetry. For example, in the 5-fraction scheme, a 3-beam FVBOO plan performs comparatively with a 4-beam FIBOO plan. In the 30 fraction scheme, a 2-beam FVBOO plan better spares OARs than the 4-beam FIBOO plan. Considering the slow gantry rotating speed and the fact that a beamline is usually shared between multiple gantry rooms, the use of fewer beams in a treatment fraction allows for reducing the patient treatment time and improve the patient throughput without compromising the treatment quality.

Moreover, this fraction-variant scheme can be complementary to FLASH therapy, which utilizes ultra-high dose rate to enhance therapeutic ratio[67–69]. While proton has

potential to achieve the FLASH dose rate for clinical treatment in a single field, the time between two beams cannot be substantially reduced. The proposed method would allow fewer beams to be used to shorten the total treatment time without compromising the physical dose distribution.

In the current problem formulation, the *fractional* dose of PTV and the *cumulative* dose of OARs are considered and penalized. Under this condition, the OARs doses in each fraction are not strictly constrained but naturally limited by the dose to the PTV. To avoid overdosing an OAR in a single fraction, an OAR fractional dose penalization can be similarly incorporated in the FVBOO Equation 3-1.

For the same reason, our FV approach is different from the spatiotemporal modulation proposed by Unkelbach et al[70–72] that introduces a heterogeneous fractional tumor dose for the improved overall biologically equivalent dose (BED). By maintaining the same uniform tumor dose throughout the treatment course, out method does not need to assume a radiobiological model and is more compatible with the current clinical practice.

The ability to select variant beams for the FVBOO method relied on the non-convex L2,1/2-norm group sparsity. We derived the proximal function for the group sparsity term so Equation 3-1 can be solved by FISTA, which is by far more efficient than ADMM. FISTA originally assumes the function convexity[56]. Nevertheless, we found that FISTA converged to a good solution even in the case of using the non-convex L2,1/2-norm group sparsity penalty[35,66]. In this particular case, we observed in the 30 fractions plans, only 20 out of 109 total selected beams are single-use. This is possibly due to the local minima and initialization condition, which currently use random initialization, and that the

optimization is trapped in local minima. Besides the demonstrated dosimetric performance, further improvement may be possible by using a different initialization strategy. On the practical side, re-using the "good" beam angles can reduce the QA workload. The tradeoff between using more beams and using the best beams can be further explored.

FVBOO method would create a challenge in the measurement-based patient-specific qualify assurance (PSQA), which needs to be performed for individual fractions instead of the entire plan. However, the hurdle may be overcome by adopting calculation [73–75] and treatment log file analysis[76–78] based PSQA.

Another challenge of FVBOO is its longer computational time. For the ESG case with a large PTV volume of 480cc, the FVBOO plan took about 10 hours in comparison to the 30 minutes used by FIBOO planning. On the other hand, the extra computational time can be unsupervised after adopting the same planning parameters from the FIBOO plan.

Similar to proton arc therapy, FVBOO improves the dose conformity and adjacent organ sparing at the cost of increased low dose bath to normal tissue, as shown by the increase V2 and V5. Therefore, the use of FVBOO has to consider individual clinical requirements to balance the need for high dose conformality (to spare adjacent OARs) and the need to minimize low-dose bath effects. Nevertheless, FVBOO offers an alternative option to the clinicians.

3.1.5 Conclusion

This work demonstrates a new IMPT optimization approach to vary beam angles during different treatment sessions. It provides a solution to use few beams per fraction

but many more beams throughout the entire treatment to improve either the plan quality or treatment delivery efficiency.

3.2 Integrating Biological Effectiveness into Beam Orientation Optimization for IMPT

3.2.1 Introduction

In current proton therapy clinical practice, a constant relative biological effectiveness (RBE) value of 1.1 is used[17–19], assuming proton therapy is 10% more effective than high-energy photons. However, the generic RBE of 1.1 is an averaged value at the center of a spread-out Bragg peak (SOBP) for 65-250 MeV proton of in-vivo systems[18]. The RBE can vary substantially along treatment fields (ranging from 1.0 to 1.6 in SOBP)[27]. RBE values also depend on several other factors, such as linear energy transfer (LET), tissue radiobiological properties (α and β value), physical dose, and specific biological endpoint[18,27,79]. In addition, with the pencil beam scanning (PBS) technique replacing passive scattering to be the mainstream delivery modality, the biological doses potentially differ from the previous observation on passive scattering[28,29], warranting further investigation in the universal use of RBE=1.1.

There have been concerns that using the generic RBE value in proton therapy can lead to underdosage in the target or underestimation of the normal tissue toxicities. Several empirical RBE calculation models have been proposed[26,79–82] to more accurately predict the RBE values. Efforts have then been made to include the RBE-weighted dose into treatment planning[83–86]. However, the dependence of these models on fitting parameters and tissue

radiobiological properties introduces considerable uncertainties in RBE-weighted dose prediction, making it difficult to incorporate them into clinical treatment planning.

Alternatively, dose-averaged LET has been suggested as a surrogate for indirect biological optimization[87]. The increase of biological effectiveness from the entrance to the distal edge of the Bragg peak is largely due to the increase of LET towards the end of proton range. Although the relationship is nonlinear, RBE increases monotonically with LET[26,27,79–82], making LET a reasonable first order approximation of RBE. Moreover, in contrast to the large RBE estimation uncertainties, LET can be accurately calculated via analytical modeling[88–91] or Monte Carlo simulation[92–95]. The LET values can then be utilized in multi-field optimized Intensity-Modulated Proton Therapy (MFO-IMPT, shorted as IMPT)[30,96].

Studies have been performed to incorporate dose-averaged LET into biological optimization of IMPT. Tseung et al[97] and Fager et al[98] used LET painting to directly optimize the biological dose instead of the physical dose, which was considered impractical. A safer and more acceptable strategy is to simultaneously optimize LET and physical dose. Works have be done to maximize the LET in the target or minimize the LET in the critical organs at risk (OARs), while achieving the physical prescription doses in the target and the OARs[30,31,99–103]. For example, Unkelbach et al[30] suggested to reoptimize the product of LET and physical dose (LET×D) after obtaining an initial IMPT plan based on the physical dose. They showed reduced LET hot spots in the critical structures with little physical dose degradation.

However, existing LET or LET×D optimization is limited to fixed beams despite the significant implication of beam orientations on the LET distribution[30–34]. For example, if an OAR abuts the target in the distal edge of a proton beam, it is difficult to reduce the LET in

this OAR without compromising physical dose coverage. In clinical practice, a planner can avoid some of the undesirable beam orientations based on experience[31–34], but evaluating all beam angles for their dosimetry, robustness, and LET values is a large computational task unsuited for human operators. A beam orientation optimization (BOO) algorithm for both physical and biological dose optimization is essential for IMPT but has not been developed. In this work, we expand the group sparsity BOO framework developed in Chapter 2 for IMPT biological dose optimization.

3.2.2 Methods

This Biological effectiveness-coupled Beam Orientation Optimization (BioBOO) method aims to select proton beams and generate treatment plans with both superior physical dose distribution and biological dose sparing. The dose and LET product (LET×D) is used as a surrogate of biological dose. The optimization function is formulated with a dose fidelity term, a LET×D and a group sparsity term. The details are described as follows.

3.2.2.1 Dose and LET product

As proposed by Unkelbach et al[30], the RBE-weighted dose at voxel i, written as b_i, can be approximated by:

$$b_i = D_i + cLD_i,$$

Equation 3-2

where D_i is the physical dose delivered to voxel i, LD_i is the dose and LET product at voxel i, and c is a scaling factor. The two terms represent the physical and LET weighted doses, respectively. The scaling factor c value is assumed 0.04 μm/keV following the publication[30].

In order to formulate the optimization problem, two matrices, A and L, are first defined. The matrix A is the dose calculation matrix as defined in Section 2.2.1. The element in the ith row and jth column of matrix A, denoted as a_{ij}, representing the physical dose contribution from the pencil beam j of unit intensity to the voxel i. Matrix L is the LET calculation matrix. Similar to A, the ijth element in L, denoted as l_{ij}, is the LET from the pencil beam j to the voxel i of unit intensity. Let $\boldsymbol{x}$ be a vector representing the intensities of all the scanning spots, with x_j indicating the intensity of jth spot, then the physical dose to voxel i from all scanning spots is calculated as

$$D_i = \sum_j a_{ij} x_j.$$

Equation 3-3

The dose-averaged LET to voxel i over all scanning spots is

$$LET_i = \frac{1}{D_i} \sum_j l_{ij} a_{ij} x_j.$$

Equation 3-4

Therefore, the product of dose and LET at voxel i is:

$$LET_i \times D_i = \sum_j l_{ij} a_{ij} x_j.$$

Equation 3-5

In matrix-vector representation, the vectorized physical dose in the patient volume, denoted as D, can be written as

$$D = A\boldsymbol{x}.$$

Equation 3-6

The product of dose and LET in the patient volume, denoted as LD, can be calculated as the following matrix and vector multiplication:

$$LD = (L \circ A)\boldsymbol{x},$$

Equation 3-7

where the symbol '$\circ$' represents element-wise multiplication.

3.2.2.2 Problem formulation

In the BioBOO framework, the LET×D constraint is incorporated into the group sparsity based BOO to encourage selecting proton beams which minimize LET×D in the OARs, while maintaining LET×D to the target as well as achieving superior physical dose distribution. Assume $\mathcal{B}$ is the set containing all the feasible candidate beams. The BioBOO problem is formulated as

$$\begin{aligned}\underset{\boldsymbol{x}}{\text{minimize}} \ & \sum_{j\in\mathcal{T}} \omega_j \left\|A_j\boldsymbol{x} - p_j\right\|_2^2 + \sum_{j\in\mathcal{O}} \omega_j \left\|A_j\boldsymbol{x}\right\|_2^2 \\ & + \sum_{k\in\mathcal{T}} \beta_k \left\|\left((LD)_k^{\text{ref}} - (L\circ A)_k\boldsymbol{x}\right)_+\right\|_2^2 + \sum_{k\in\mathcal{O}} \beta_k \|(L\circ A)_k\boldsymbol{x}\|_2^2 \\ & + \sum_{b\in\mathcal{B}} \alpha_b \|\boldsymbol{x}_b\|_2^{1/2} \\ \text{subject to} \quad & \boldsymbol{x} \geq 0,\end{aligned}$$

Equation 3-8

where $\boldsymbol{x}_b$ is the fluence map candidate beam b, so the optimization variable and $\boldsymbol{x}$ is the concatenation of all the vectors x_b ($b \in \mathcal{B}$). The dose calculation matrix A and LET calculation matrix L include all the candidate beams along the column direction. $\mathcal{T}$ is the set including the target volumes and $\mathcal{O}$ is the set including the OARs.

The first two terms in Equation 3-8 are the conventional physical dose fidelity term similar to Equation 2-3. The first term penalizes the dose deviation of target *j* from prescription dose p_j, to ensure a homogeneous physical dose distribution in the target. The second term penalizes any non-zero doses in the OARs, to reduce the doses delivered to the OARs. The third and fourth terms together are the LET×D conditions. The third term encourages the LET×D values in the target *k* to be greater than $(LD)_k^{\mathrm{ref}}$, to prevent cold spots in the biological dose. The fourth term minimizes the LET×D values in the OARs. ω_j and β_k are the structure weighting hyperparameters for dose and LET×D constraints, respectively. The last term $\sum_{b\in\mathcal{B}} \alpha_b \|\boldsymbol{x}_b\|_2^{1/2}$ is the L2,1/2-norm group sparsity term defined in Section 2.2.1 for selecting beams.

Without the third and fourth terms penalizing LET×D, the Equation 3-8 describes the group sparsity based BOO framework proposed in Chapter 2, only ensuring physical dose sparing. After adding these two terms, proton beam angles and treatment plans are generated simultaneously with optimum physical and biological dose sparing. FISTA[56] is used to solve this non-differentiable problem.

3.2.2.3 Evaluations

Three patients with skull base tumor (SBT) and three patients with bilateral head-and-neck (H&N) cancer were tested. The candidate beam set included 700 to 800 non-coplanar beams for the SBT patients, and approximately 600 beams for the H&N patients. For each candidate beam, dose and LET calculation for the scanning spots covering the PTV (planning target volume) and a 5 mm margin was performed by matRad[51,52]. The dose calculation matrix *A* and LET calculation matrix *L* including all feasible candidate beams

were hence generated. The calculation resolution was 2.5×2.5×2.5 mm^3. Since robust optimization is not considered in this work yet, the PTV was set as the optimization target. The prescription dose, target volume, and average spot count per beam for each patient are shown in Table 3-4.

Case		Prescription Dose (GyRBE)	PTV Volume (cc)	Average Spots Number per Beam
SBT #1		56	66.80	2537
SBT #2		70	70.26	2650
SBT #3	PTV6300	63	128.86	4071
	PTV7400	74	26.58	
H&N #1	PTV5400	54	257.41	10065
	PTV6000	60	274.38	
	PTV6300	63	121.52	
H&N #2	PTV5400	54	205.76	10077
	PTV6000	60	210.43	
H&N #3	PTV5400	54	206.10	9433
	PTV6000	60	173.56	
	PTV6300	63	21.20	

Table 3-4: Prescription doses, PTV volumes and average number of spots per beam for each patient.

For comparison, in addition to the BioBOO plan, the following four plans were also generated for each patient: 1) conventional plan optimizing physical dose with manually selected beams (MAN); 2) the same MAN plan reoptimized with additional cLET×D constraint (BioMAN); 3) the plan generated by group sparsity based BOO with only physical dose constraint (GSBOO); 4) the same GSBOO plan reoptimized with additional cLET×D constraint (GSBOO_BioFMO, with FMO representing fluence map optimization). The differences of these plans are listed in Table 3-5.

Acronym	Initial optimization			Re-optimized with cLET×D constraint ?
	Beam selection method	Physical dose constraint ?	cLET×D constraint ?	
MAN	Manual selection	Yes	No	No
BioMAN	Manual selection	Yes	No	Yes
GSBOO	Group sparsity	Yes	No	No
GSBOO_BioFMO	Group sparsity	Yes	No	Yes
BioBOO	Group sparsity	Yes	Yes	No

Table 3-5: Acronyms of different methods and the comparison.

For all plans, the goal of physical dose optimization is the same as conventional treatment planning. We set the physical dose distribution in the target to be homogeneous and a constant RBE value of 1.1 was used. The plans are normalized so that 95% of the target volume receives the prescribed physical dose, which is $\frac{\text{prescription dose}}{1.1}$. For the biological component, since there is no predefined reference value for cLET × D that can be used, we set the $c(LD)^{\text{ref}}$ of the PTVs to the mean cLET × D value of the PTVs in the MAN plans.

3.2.3 Results

3.2.3.1 Runtime and selected beams

The calculation and optimization were performed on a Xeon 14-core CPU server operating at 2.40 GHz clock. To calculate the dose and LET of all feasible candidate beams, the Matlab Parallel Computing Toolbox was used to accelerate the computation. The times spent on the dose and LET calculation and the BOO runtime for the GSBOO and BioBOO plans are listed in Table 3-6. The couch and gantry angles for the beams from manual selection, GSBOO, and BioBOO, are also listed in Table 3-6. In matRad, the physical dose

calculation and LET calculation share the same ray-tracing procedure, which is also the most time-consuming step. Therefore, the total time for dose and LET calculation is shown. With the analytical calculation model and parallel computing, the total time for dose and LET calculation is between 10 to 60 min depending on the target size. While the GSBOO process with only physical dose constraint took about 20-70 minutes to complete, the BioBOO process with additional cLET×D constraint increased the BOO time by 30-80%.

Case	Dose, LET calculation time (min)	BOO runtime (min)		Selected beam angles (gantry, couch)		
		GSBOO	BioBOO	MAN	GSBOO	BioBOO
SBT #1	11	21	31	(60,275)(270,0) (90,0)	(303,62)(97,324) (42,37)	(288,45)(123,331) (42,37)
SBT #2	16	25	46	(60,275)(270,0) (90,0)(180,0)	(62,339)(341,19) (300,353)(17,46)	(95,281)(60,332) (137,18)(276,342)
SBT #3	21	31	56	(60,275)(270,0) (90,0)(180,0)	(66,0)(84,0) (270,342)(316,314)	(268,84)(276,0) (67,20)(33,66)
H&N #1	56	70	101	(0,0)(160,0) (200,0)	(51,320)(154,293) (330,321)	(188,45)(38,20) (330,321)
H&N #2	58	55	73	(0,0)(160,0) (200,0)	(322,20)(167,296) (212,23)	(167,296)(212,23) (324,301)
H&N #3	55	62	89	(0,0)(160,0) (200,0)	(149,348)(193,27) (41,50)	(149,348)(188,45) (31,76)

Table 3-6: Optimization time and selected beam angles for each patient.

3.2.3.2 SBT cases

BioBOO is compared with MAN and GSBOO. The dose volume histograms and cLEDxD volume histograms for the three SBT patients comparing MAN, BioMAN and

BioBOO are shown in Figure 3-5 and that comparing GSBOO, GSBOO_BioFMO and BioBOO are shown in Figure 3-6. The differences in dose and LET×D of BioMAN, GSBOO_BioFMO and BioBOO plan from the MAN plan for some OARs are shown in Figure 3-7.

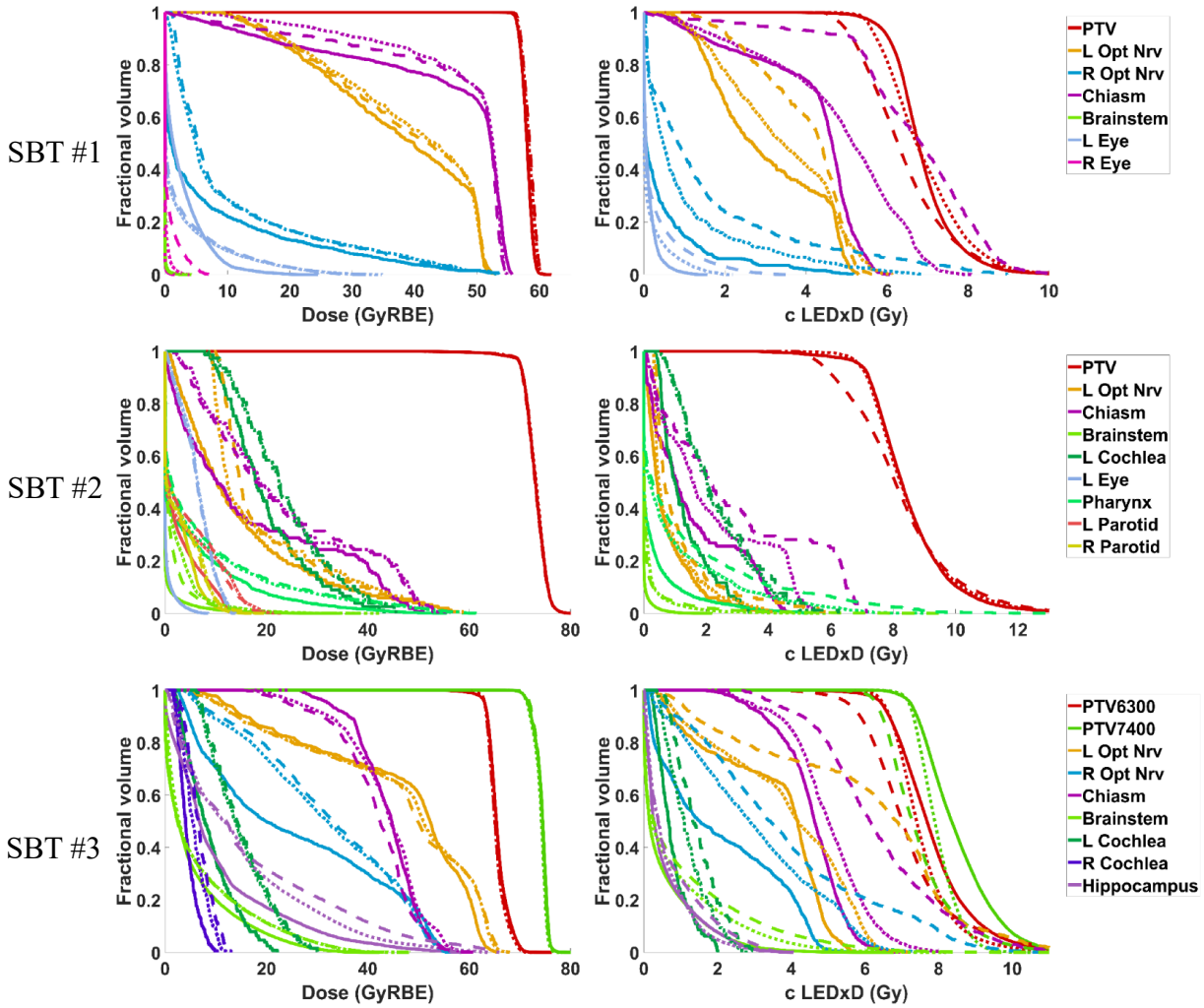

Figure 3-5: Comparison of dose and cLEDxD between BioBOO (solid), BioMAN (dotted) and MAN (dashed) for the SBT patients. Left column is the dose volume histogram and right column is the cLEDxD volume histogram.

Qualitatively, all the five methods achieved similar PTV dose coverage. In the GSBOO plans, the cLET×D values of the PTVs were not guaranteed, which can be higher (SBT #1) or lower (SBT #2) than the MAN plan. With the LET×D constraint in the BioMAN, BioBOO and GSBOO_BioFMO plans, the mean and maximal values of LET×D of the PTVs

were similar compared with the MAN plans, but the minimal values of PTV LET×D were improved relative to the MAN plans.

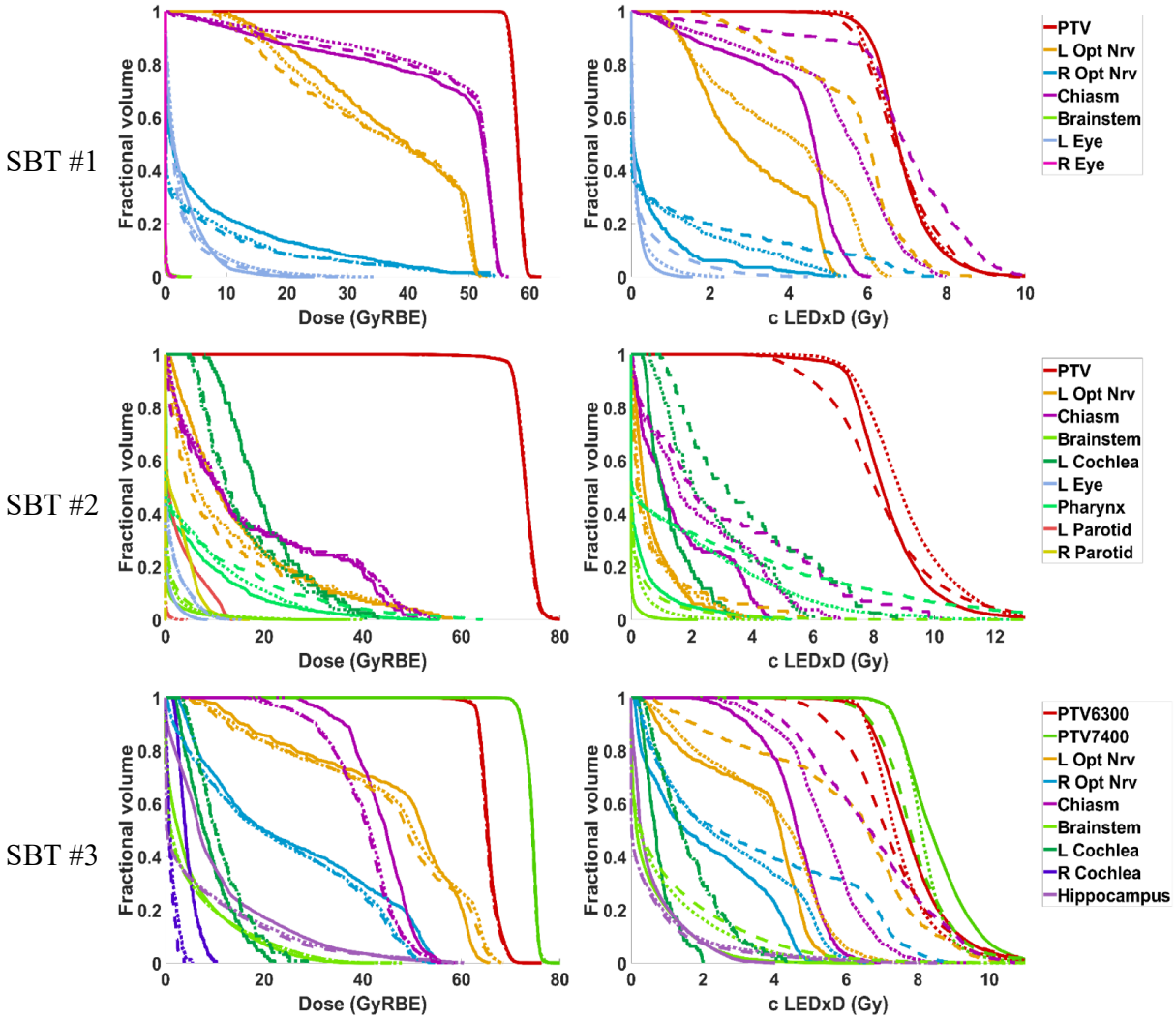

Figure 3-6: Comparison of dose and cLEDxD between BioBOO (solid), GSBOO_BioFMO (dotted) and GSBOO (dashed) for the SBT patients. Left column is the dose volume histogram and right column is the cLEDxD volume histogram.

Reoptimizing the MAN plan based on the proposed LET×D constraint resulted in lower LET×D of the OARs while maintaining similar physical dose distribution. In the GSBOO plans, where only physical dose constraint was considered, the physical dose was improved, but the sparing of LET×D was not guaranteed. For example, the maximal cLET×D

of the left optical nerve for SBT #1 in the GSBOO plan was 2.4 Gy higher than the MAN plan. Reoptimizing LET×D based on the GSBOO plans leads to slightly degraded physical dose distribution and lower LET×D, but the LET×D sparing was not as good as BioMAN. On average, the GSBOO_BioFMO plans reduced [Dmean, Dmax] from the BioMAN plans by [3.8, 4.5] GyRBE on average, but increasing [cLET×Dmean, cLET×Dmax] by [0.1, 0.4] Gy.

The BioBOO plans achieved better OARs dose sparing and further reduced the OARs LET×D. The physical dose was reduced in the BioBOO plans from the MAN plan for most considered OARs except the chiasm in SBT #1 and SBT #3, and the right optical nerve in SBT #2. Even for these structures that were not improved compared with MAN, the difference was smaller than 1 GyRBE. The structure with the largest reduction in the maximal doses from the MAN plans were left eye (9.1 GyRBE) for SBT #1, pharynx (13.8 GyRBE) for SBT #2, and hippocampus (15.7 GyRBE) for SBT #3. The averaged reduction in [Dmean, Dmax] of the BioBOO plans from the MAN plans were [2.85, 4.6] GyRBE, while the averaged reduction of [Dmean, Dmax] of the BioMAN plans from the MAN plans were [0.1, 0.5] GyRBE.

Meanwhile, even though the BioBOO method did not improve the physical dose for certain structures in the SBT cases, it further reduced the LET×D compared with BioMAN. For example, the maximal cLET×D of chiasm in the three cases were 1.5 Gy, 0.9 Gy and 1.0 Gy lower than the BioMAN plans, respectively, while the physical dose were similar. Compared with the BioMAN plans, the structure with the largest reduction of maximal cLET×D by BioBOO in each case was right optical nerve (1.7 Gy) for SBT #1, pharynx (2.9 Gy) for SBT #2, and brainstem (1.6 Gy) for SBT #3. The averaged reduction of [cLET×Dmean, cLET×Dmax] of the BioBOO plans from the MAN plans were [1.1, 2.9] Gy,

while the averaged reduction of [cLET×Dmean, cLET×Dmax] of the BioMAN plans from the MAN plans were [0.7, 1.7] Gy.

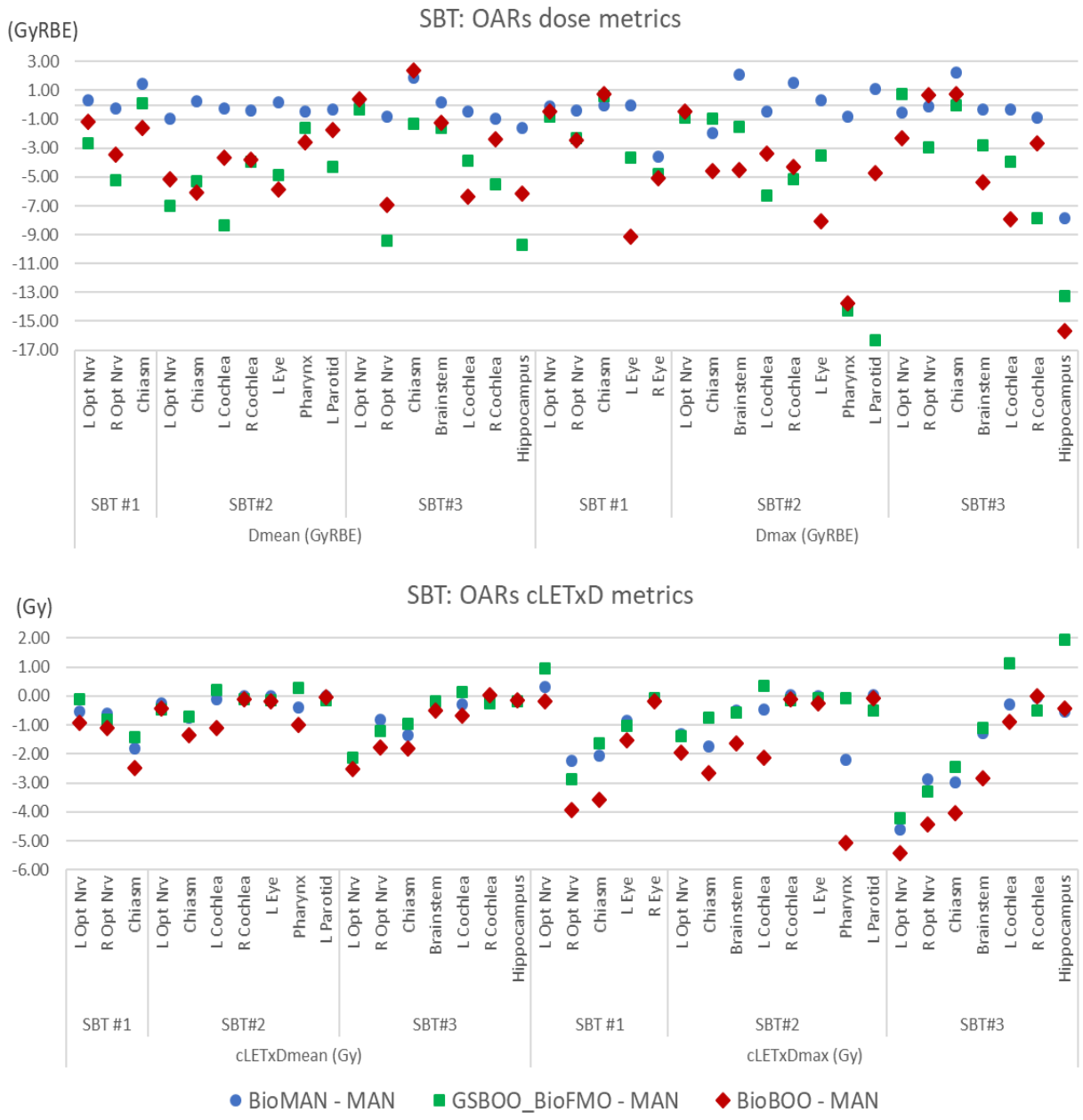

Figure 3-7: The difference of OAR dose and cLET×D metrics of BioMAN, GSBOO_BioFMO and BioBOO from MAN for the three SBT patients. A negative sign represents a reduction from the MAN plan and a positive sign represents an increase.

3.2.3.3 H&N cases

The dose volume histograms and cLEDxD volume histograms for the three H&N patients comparing MAN, BioMAN, and BioBOO are shown in Figure 3-8 and that comparing GSBOO, GSBOO_BioFMO, and BioBOO are shown in Figure 3-9. The differences in dose and LET×D of BioMAN, GSBOO_BioFMO and BioBOO plans relative to the MAN plan for selected OARs are shown in Figure 3-10.

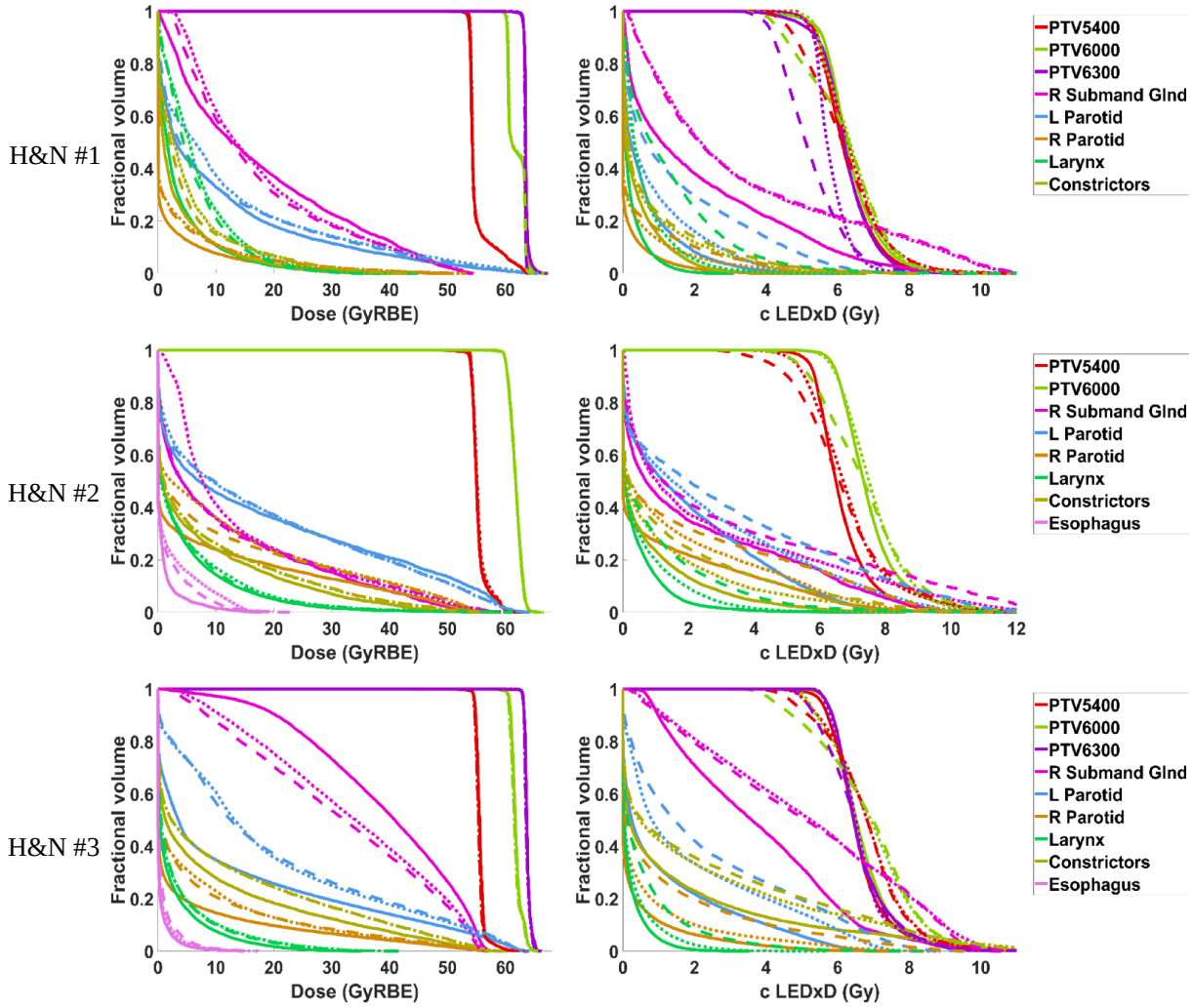

Figure 3-8: Comparison of dose and cLEDxD between BioBOO (solid), BioMAN (dotted) and MAN (dashed) for the H&N patients. Left column is the dose volume histogram and right column is the cLEDxD volume histogram.

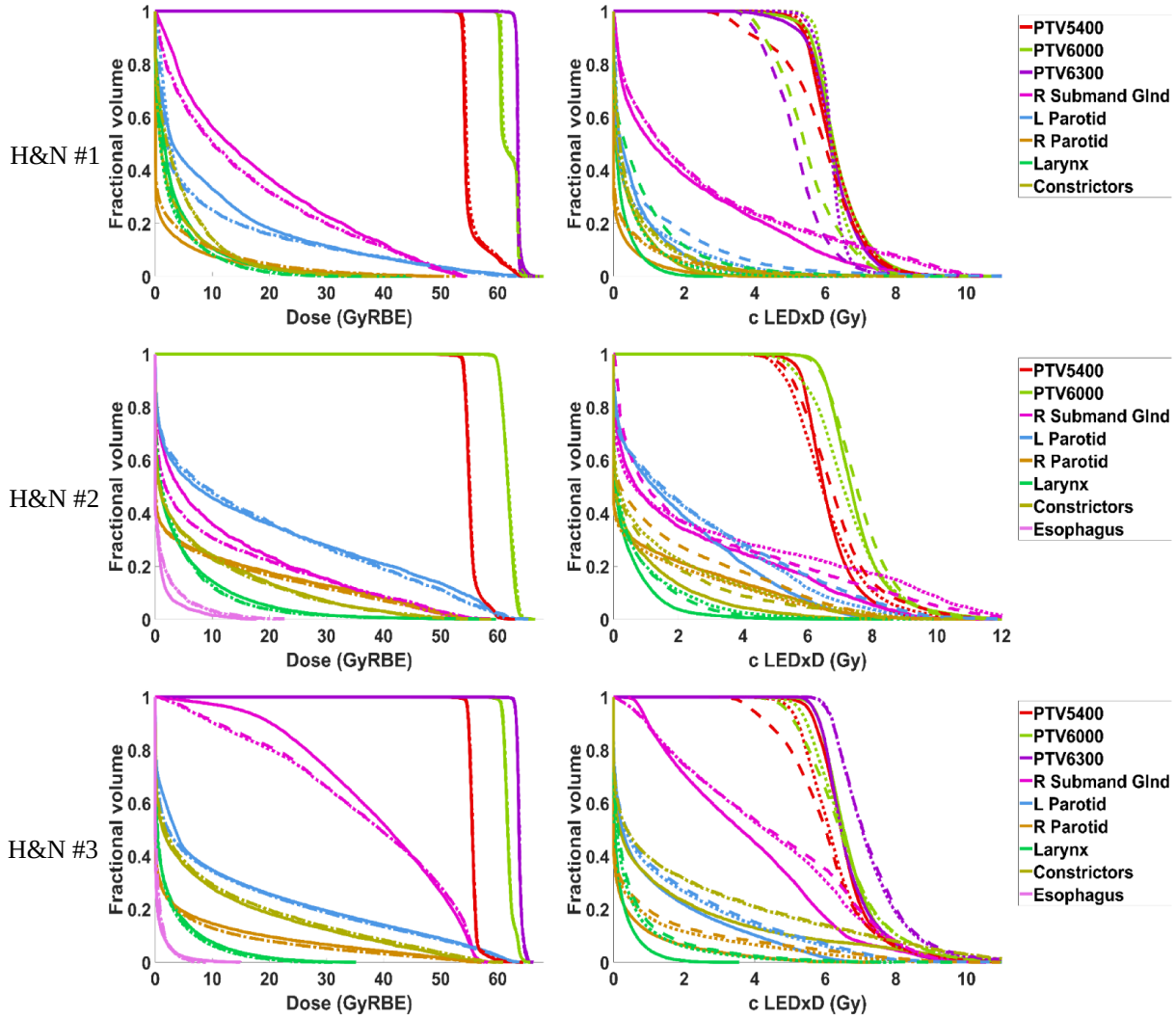

Figure 3-9: Comparison of dose and cLEDxD between BioBOO (solid), GSBOO_BioFMO (dotted) and GSBOO (dashed) for the H&N patients. Left column is the dose volume histogram and right column is the cLEDxD volume histogram.

Similar to the SBT cases, all the five methods achieved similar PTV dose coverage and comparable cLET×D distribution. An improvement of minimal cLET×D is observed in the BioMAN, BioBOO and GSBOO_BioFMO plans from the MAN plans, indicating reduced cLET×D cold spots in the PTVs. Compared with BioMAN, GSBOO_BioFMO reduced [Dmean, Dmax] by [1.7, 2.6] GyRBE on average, with similar LET×D and an average increase of [cLET×Dmean, cLET×Dmax] by [-0.1, 0.1] Gy.

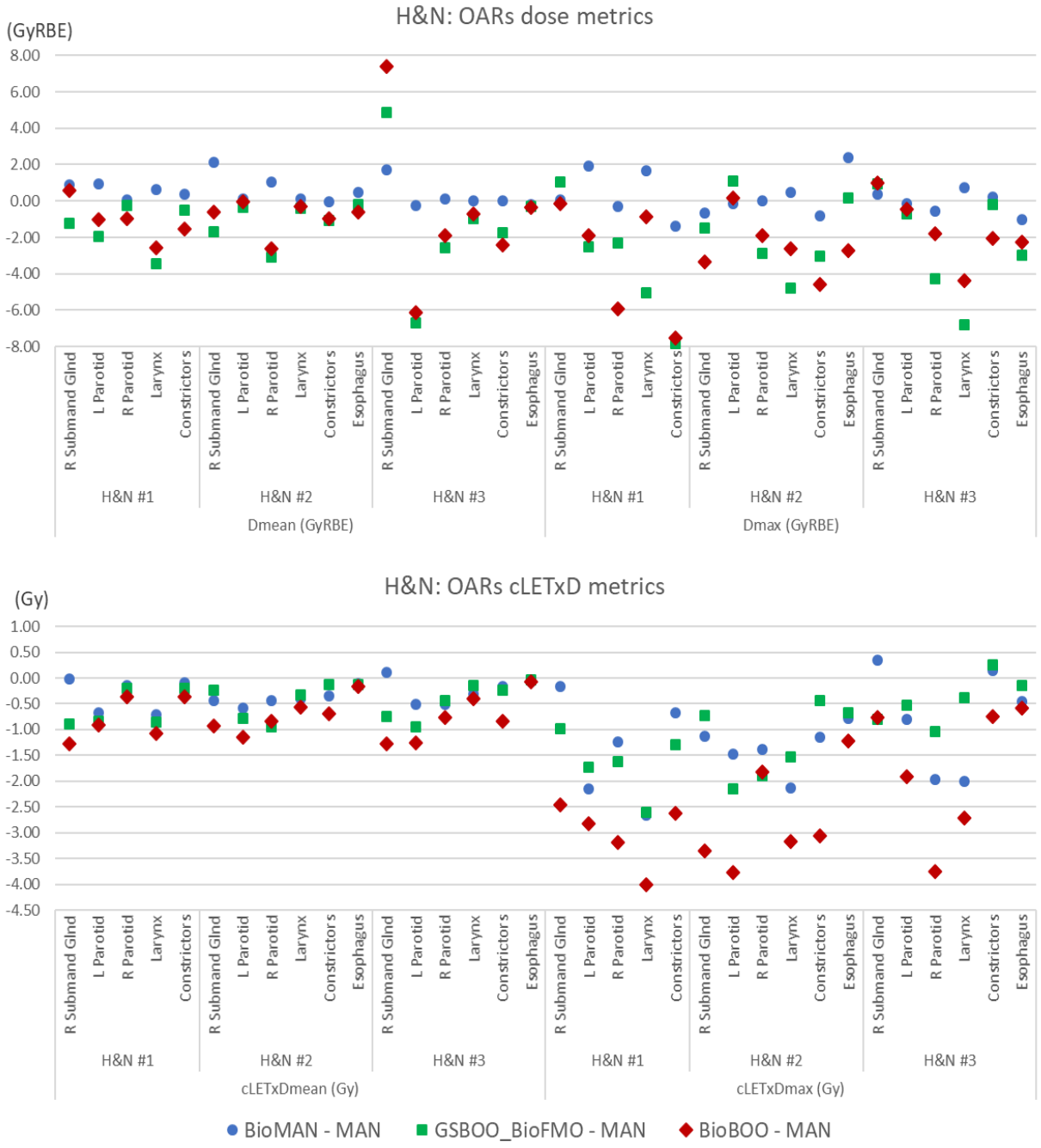

Figure 3-10: The difference of OAR dose and cLET×D metrics of BioMAN, GSBOO_BioFMO and BioBOO from MAN for the three H&N patients. A negative sign represents a reduction from the MAN plan and a positive sign represents an increase.

The OARs doses in the BioBOO plans were consistently reduced compared with the MAN plans except the right submandibular gland in H&N #1 and H&N #3. The structure of the largest reduction of maximal dose from the MAN plan in each case was the right parotid

(5.9 GyRBE) for H&N #1, constrictors (4.6 GyRBE) for H&N #2, and larynx (4.4 GyRBE) for H&N #3. The averaged reduction of [Dmean, Dmax] of the BioBOO plans from the MAN plans were [0.9, 2.5] GyRBE, while on average the [Dmean, Dmax] of the BioMAN plans were increased by [0.5, 0.2] GyRBE from the MAN plans.

Even though the doses to the right submandibular gland in the three BioBOO plans were comparable to the MAN plans, an effective reduction of LET×D by the BioBOO method was observed. Compared with the BioMAN plans, the maximal cLET×D of the right submandibular gland in the three cases was 2.3 Gy, 2.2 Gy and 1.2 Gy lower in the BioBOO plans. The averaged reduction of [cLET×Dmean, cLET×Dmax] of the BioBOO plans from the MAN plans were [0.8, 2.6] Gy, while the averaged reduction of [cLET×Dmean, cLET×Dmax] of the BioMAN plans from the MAN plans were [0.3, 1.2] Gy.

3.2.4 Discussion

Unexpected high LET-weighted pencil beam dose deposition inside a sensitive OAR has raised considerable concerns in IMPT plans. Although the beam direction strongly correlates with the distal proton biological effect, IMPT optimization may scatter high-LET beams throughout the target volume or sometimes within OARs. Manual selection of proton fields for multiple planning goals including optimal dosimetry, biological effectiveness and robustness is a computational task beyond the capability of human planners. To the best of our knowledge, this work describes the first mathematical framework and actual implementation that include biological effectiveness in IMPT beam orientation optimization. In contrast to the limited manual trial and error approach, the BioBOO framework performs a global search among all feasible candidate beams by solving

a group sparsity problem integrating physical dose objectives and biological dose constraints.

Compared with the previous approaches appending LET optimization to the physical dose optimization using manually selected beams, which were shown to reduce LET in the OARs at the cost of the physical dose[30], BioBOO further reduced *both* the physical dose and LET×D in the OARs. The OAR dose and LET reduction was achieved while maintaining the physical target dose and LET×D coverage, eliminating concerns due to uncertain tumor radiobiology and RBE modeling. It is worth noting that the current framework is flexible to increase LET×D in the tumor for potentially greater tumor cell killing by setting the $(LD)^{\mathrm{ref}}$ value in Equation 3-8 higher.

A limitation of the current BioBOO framework is that the reference LET×D values for targets and normal tissues are unknown. Subsequently, the$(LD)^{\mathrm{ref}}$and weighting hyperparameters for the LET×D constraints cannot be mechanistically determined. In this study, the $(LD)^{\mathrm{ref}}$ value for the BioMAN and BioBOO plans of each patient was extracted from the corresponding conventional MAN plan so that we could compare the biological effectiveness with the MAN plan. However, in the clinical setting, the lack of reference values makes it difficult to directly use BioBOO for a new case. A potential solution is to set a typical SOBP single field plan before treatment planning and calculate the mean LET×D for the targets as the reference value. Another alternative solution is that templates can be built for different sites, e.g., H&N and skull base in this study. Meanwhile, further preclinical and clinical research is needed for more quantitative integration of RBE modeling in treatment planning as recommended by AAPM TG-256[19].

Due to the prohibitively long time required to calculate dose and LET for over 500 candidate beams using Monte Carlo, the current study used an analytical method. While the analytical method was shown to be acceptably accurate for dose calculation and BOO planning in our previous paper[35,38], it has compromised accuracy for LET calculation due to reasons including failing to account for secondary protons[88]. Despite the limitation, the proof-of-principle study used the same analytical calculation engine across different methods for a fair comparison. Without changing the optimization framework, the dose calculation engine can be replaced by fast Monte Carlo or analytical calculation models, including secondary protons[90] in future work.

3.2.5 Conclusion

We developed a novel biological effectiveness-coupled BOO method for IMPT based on group sparsity regularization and LET×D constraint. Beams and plans with superior physical dose and biological OAR sparing are generated.

3.3 A Novel Energy Layer Optimization Framework for Spot-Scanning Proton Arc Therapy[36]

3.3.1 Introduction

To further reduce healthy tissue irradiation, the concept of proton arc therapy (PAT) has been proposed since 1997[104–106], to combine the unique dose deposition curve of protons and the benefit of rotating beams. Passive-scattering based proton arc therapy is not practical for clinical application mainly due to the difficulty of changing beam-specific compensator and range modulation wheel during gantry rotation. But with the recent

development and increasing adoption of the spot-scanning technique[4,107,108], the modulation is integrated into the gantry, making PAT technically viable.

Using the modern scanning nozzle, the modern proton systems deliver treatments spot-by-spot and layer-by-layer[109–111]. The time of spot scanning within the same energy layer is on the order of milliseconds, but it requires seconds to change energy to another layer[112,113], particularly from low to high energies. The slow energy layer switch is mainly due to magnetic hysteresis accompanying changing magnetic field strengths in the energy selection system. The energy layer switching time (ELST) cannot be easily reduced. Therefore, for practical proton arc delivery, reduction of the energy switching steps is an essential consideration besides the dosimetric quality.

Different delivery methods have been proposed for spot-scanning proton arc therapy (SPAT), like multiple static fields[114,115], distal edge tracking[116,117], and single energy modulation[118–120]. However, these delivery methods either cannot perform continuous rotation-delivery, or fail to fully utilize the freedoms in spot-scanning techniques, and the delivery efficiency is not optimized.

Ding et al[63] proposed a delivery-efficient and practical algorithm called SPArc. Similar to volumetric modulated arc therapy (VMAT)[121], this greedy algorithm starts with a coarse sampling of beams, also known as control points, and iteratively increases the sampling frequency while redistributing the energy layers, until reaching the desired sampling frequency. By this method, 1-3 energy layers remain active at each control point, ensuring acceptable delivery time. Later this algorithm was updated to optimize the energy delivery sequence from high to low instead of arbitrary switching to further shorten the

delivery time[122]. Retrospective studies show the potential of SPArc plans that improve plan dosimetry compared with intensity-modulated proton therapy (IMPT) to lung cancer[61,63], prostate cancer[62], and whole-brain radiotherapy[59]. Recently, the first prototype of SPArc delivery was performed on a clinical IBA Proteus One proton machine, with a Proton Dynamic Arc Delivery (PDAD) module. It demonstrated the feasibility of SPArc treatment within the clinical requirements[58].

However, in the current SPArc algorithm, the energy layer selection and optimization are greedy and heuristic. Due to the separate sequencing and plan optimization steps[122], the optimality of the plan delivery efficiency and the dosimetric quality cannot be promised. The alternating back-and-forth operation between the fluence map optimization and energy layer processing required for final plan creation is computationally inefficient.

To further improve SPAT, in this work, we present a novel optimization method to integrate energy layer selection and sequencing with scanning-spot optimization in a single framework, which affords a global search of all feasible energy layers and then simultaneously optimizes the energy sequence. The energy layer selection is achieved by the group sparsity regularization developed in Chapter 2. An energy-sequencing regularization is developed to improve the SPAT delivery efficiency.

3.3.2 Methods

The proposed Energy Layer Optimization incorporated Spot-scanning Proton Arc Therapy (ELO-SPAT) optimization framework aims to select as few as possible energy layers from the available candidate layers and then encourage energy switch from high to

low, for a predefined control point. Considering the continuity in gantry rotating, the optimization goal to be exactly one active energy layer at each control point.

Motivated by this consideration, the ELO-SPAT framework is formulated with a dose fidelity, a group sparsity regularization, a log barrier regularization, and an energy-sequencing (ES) penalty term. The details are described in the following sections.

3.3.2.1 Notations

Before presenting the optimization framework, we establish the following notations.

- B is the number of static beams used as sampled control points. The fixed spacing of 2.5° or 2° is used in this study.
- E is the number of candidate energy layers in each beam. To simplify the notation, it is assumed that each beam has the same candidate energy layers. The infeasible layers can be eliminated during optimization. So the number of all candidate energy layers is $B \times E$.
- The vector $\boldsymbol{x}_{be}$ is the spot intensities of eth energy layer in bth beam. The length of $\boldsymbol{x}_{be}$, denoted as N_{be}, which is the number of scanning-spots in the specific layer, varies with beams and layers.
- The vector $\boldsymbol{x}_b$ is the concatenation of $\boldsymbol{x}_{be}$ (for $e = 1, \dots, E$) with increasing energy, representing the spot intensities of bth beam. And the vector $\boldsymbol{x}$ is the concatenation of $\boldsymbol{x}_b$ (for $b = 1, \dots, B$), following the sequence of gantry rotation.

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_b \\ \vdots \\ \boldsymbol{x}_B \end{bmatrix} \text{ and } \boldsymbol{x}_b = \begin{bmatrix} \boldsymbol{x}_{b1} \\ \boldsymbol{x}_{b2} \\ \vdots \\ \boldsymbol{x}_{be} \\ \vdots \\ \boldsymbol{x}_{bE} \end{bmatrix}.$$

Equation 3-9

- The vectors $\boldsymbol{y}$ and $\boldsymbol{y}_b$ are compact representations of $\boldsymbol{x}$ and $\boldsymbol{x}_{be}$ to eliminate the dimension of scanning-spot, with the element y_{be} being the sum of all elements in $\boldsymbol{x}_{be}$. Therefore, all the $\boldsymbol{y}_b$ (for $b = 1, \dots, B$) are in the same length of E.

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_b \\ \vdots \\ \boldsymbol{y}_B \end{bmatrix}, \boldsymbol{y}_b = \begin{bmatrix} y_{b1} \\ y_{b2} \\ \vdots \\ y_{be} \\ \vdots \\ y_{bE} \end{bmatrix} \text{ and } y_{be} = \sum_{i=1}^{N_{be}} x_{bei}.$$

Equation 3-10

Equation 3-15 can also be written as matrix-vector multiplication:

$$\boldsymbol{y} = W\boldsymbol{x},$$

Equation 3-11

where W is a summation matrix, to sum up $\boldsymbol{x}$ along the spot dimension.

- A new variable $\tilde{\boldsymbol{y}}_b$ is defined by replacing each element in $\boldsymbol{y}_b$ with 0 except the maximal element.

$$\tilde{\boldsymbol{y}} = \begin{bmatrix} \tilde{\boldsymbol{y}}_1 \\ \vdots \\ \tilde{\boldsymbol{y}}_b \\ \vdots \\ \tilde{\boldsymbol{y}}_B \end{bmatrix}, \tilde{\boldsymbol{y}}_b = \begin{bmatrix} \tilde{y}_{b1} \\ \vdots \\ \tilde{y}_{be} \\ \vdots \\ \tilde{y}_{bB} \end{bmatrix} \text{ and } \tilde{y}_{be} = \begin{cases} y_{be}, \text{if } y_{be} = \max(\boldsymbol{y}_b) \\ 0, \text{otherwise} \end{cases}.$$

Equation 3-12

- The matrix D_b^B is a discrete gradient operator for bth beam along the beam direction. For example, $D_b^B \boldsymbol{y}_b$ is a vector of the intensity difference between $\boldsymbol{y}_{b+1}$ and $\boldsymbol{y}_{b,}$

$$D_b^B \boldsymbol{y}_b = \boldsymbol{y}_{b+1} - \boldsymbol{y}_{b.}$$

Equation 3-13

- $\widetilde{D}_b^B$ is a specially designed gradient operator to make

$$\widetilde{D}_b^B \boldsymbol{y}_b = D_b^B \tilde{\boldsymbol{y}}_b.$$

Equation 3-14

- The matrix D_b^E is a discrete gradient operator for bth beam along energy direction, while ignoring all zero elements. For example,

$$D_b^E \begin{bmatrix} 0 \\ \vdots \\ y_{be_1} \\ 0 \\ y_{be_2} \\ \vdots \\ 0 \\ y_{be_3} \end{bmatrix} = \begin{bmatrix} y_{be_2} - y_{be_1} \\ y_{be_3} - y_{be_2} \end{bmatrix}, \text{and } y_{be_1}, y_{be_2}, y_{be_3} \neq 0 .$$

Equation 3-15

- A is the dose-calculation matrix. A contains all the spots from the entire $B \times E$ candidate layers.
- A sigmoid operator $\mathcal{S}$ on a vector $\boldsymbol{u}$ of length K, is defined as

$$\mathcal{S}(\boldsymbol{u}) = \begin{bmatrix} s(u_1) \\ \vdots \\ s(u_k) \\ \vdots \\ s(u_K) \end{bmatrix} \text{ and } s(t) = \frac{2}{1 + e^{-\eta t}} - 1,$$

Equation 3-16

where $s(t)$ is a modified sigmoid function, as a smooth approximation of sign function, and η controls the level of smoothness. The function of $\mathcal{S}$ is to normalize each element in $\boldsymbol{u}$ to -1, 0 or +1.

3.3.2.2 Formulation of ELO-SPAT

The ELO-SPAT is formulated as follows:

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \quad \Gamma(A\boldsymbol{x}) + \sum_{b=1}^{B}\sum_{e=1}^{E} \alpha_{be}\|\boldsymbol{x}_{be}\|_2^{1/2} - \beta \sum_{b=1}^{B} \log\left(\sum_{e=1}^{E} y_{be}\right) + \gamma \sum_{b=1}^{B-1} h(D_b^E \mathcal{S}(\widetilde{D}_b^B \boldsymbol{y}_b))$$

$$\text{subject to} \quad \boldsymbol{x} \geq 0,$$

$$\boldsymbol{y} = W\boldsymbol{x}.$$

Equation 3-17

The first term is the dose fidelity term, penalizing the actual dose, calculated by $A\boldsymbol{x}$, from the prescription dose. The same quadratic function as Equation 2-3 is used in this study, but the choice of dose fidelity cost is flexible. The second term is the L2,1/2-norm group sparsity term. With proper tuning of the weighting hyperparameter α_{be}, the non-convex ½ norm effectively turns off most candidate layers. But the term alone can result in aggregated layers in some beam blocks and leaving some control points with no layers active, which does not fully utilize the rotating beams. In the third term, a log barrier regularization function is used to distribute the selected layers to the whole gantry rotating range. The term sums up the intensity of each beam and penalizes the zero intensities, therefore forcing each beam to keep at least one layer selected. β is the regularization parameter for the log barrier function. By picking a proper value of β and setting α_{be} large enough, one energy layer per beam can be ensured.

The fourth term regularizes energy-sequencing (ES) with a weighting parameter γ. ES regularization asymmetrically penalizes energy switching low-to-high harder than high-to-low. The details of ES regularization can be found in Section 3.3.2.3.

3.3.2.3 Energy-sequencing regularization

The method using the group sparsity to select a few layers out of the candidates is by gradually reducing the $\boldsymbol{x}_{be}$ of the layer with a lower weight to zero during the iterations. As a result, in each beam, the layer with the maximal intensity is most likely kept in each iteration. Therefore, in the energy-sequencing term, instead of $\boldsymbol{y}_b$, we control $\widetilde{\boldsymbol{y}}_b$, which only keeps the maximal element in $\boldsymbol{y}_b$ and sets all others to zero, as defined in Section 3.3.2.1.

To better understand how energy sequencing works, we consider two adjacent beams during gantry rotation, beam b, and beam $b+1$. Assume the only nonzero elements of $\widetilde{\boldsymbol{y}}_b$ and $\widetilde{\boldsymbol{y}}_{b+1}$ are y_{be_1} and y_{be_2}, respectively, at the position of e_1 and e_2. As shown in Figure 3-11, if $e_2 > e_1$, meaning energy going up from beam b to beam $b+1$, the vector $\widetilde{\boldsymbol{y}}_{b+1} - \widetilde{\boldsymbol{y}}_b$, or $D_b^B \widetilde{\boldsymbol{y}}_b$, shows a pattern of transitioning from a negative value to a positive value, with possible zeros before, between, and after. A sigmoid operator $\mathcal{S}$ is suited to normalize each nonzero element in $D_b^B \widetilde{\boldsymbol{y}}_b$ to -1 or +1.

With normalization, $\mathcal{S}(D_b^B \widetilde{\boldsymbol{y}}_b)$ is a vector with only two non-zero elements -1, and +1, respectively. When the energy goes up, taking the difference of the non-zero elements along e direction (the D_b^E operator defined in Section 3.3.2.1) results in +2.

Similarly, if the energy goes from high to low, the result of D_b^E operation on $\mathcal{S}(D_b^B \widetilde{\boldsymbol{y}}_b)$ is -2. Maintaining the same energy would result in zero values.

The above process can be written as $D_b^E \mathcal{S}(D_b^B \tilde{\boldsymbol{y}}_b)$ or equivalently $D_b^E \mathcal{S}(\widetilde{D}_b^B \boldsymbol{y}_b)$, as defined in Section 3.3.2.1. In summary, the value of $D_b^E \mathcal{S}(D_b^B \tilde{\boldsymbol{y}}_b)$ indicates the energy changing pattern between adjacent beams in the following relationship:

$$D_b^E \mathcal{S}\left(\widetilde{D}_b^B \boldsymbol{y}_b\right) = \begin{cases} +2, & \text{energy switch} - \text{up}, \\ -2, & \text{energy switch} - \text{down}, \\ 0, & \text{energy unchanged}. \end{cases}$$

Equation 3-18

For delivery efficiency, fewer energy switch-ups during gantry rotating are encouraged. Therefore, positive $D_b^E \mathcal{S}\left(\widetilde{D}_b^B \boldsymbol{y}_b\right)$ is more heavily penalized. In this work, the energy switch-down is less penalized than staying unchanged for two reasons. First, energy switching down has a small impact on the total delivery time. For example, the switching-down time is 0.6 s, according to the IBA Proton Dynamic Arc Delivery module, which is considered a negligible increase compared with staying unchanged in this study. Second, doing so encourages more layers to be used for better dosimetry. Mathematically, a one-sided quadratic cost function is used to penalize energy switching. The cost function, $h(t)$, is defined as:

$$h(t) = \begin{cases} \frac{1}{4}(t+2)^2 - 1, & \text{if } t \geq -2, \\ -1, & \text{otherwise}. \end{cases}$$

Equation 3-19

The quadratic term makes the function $h(t)$ convex, smooth and differentiable, and the function definition is designed to give a value of 0 at 0 for simplicity.

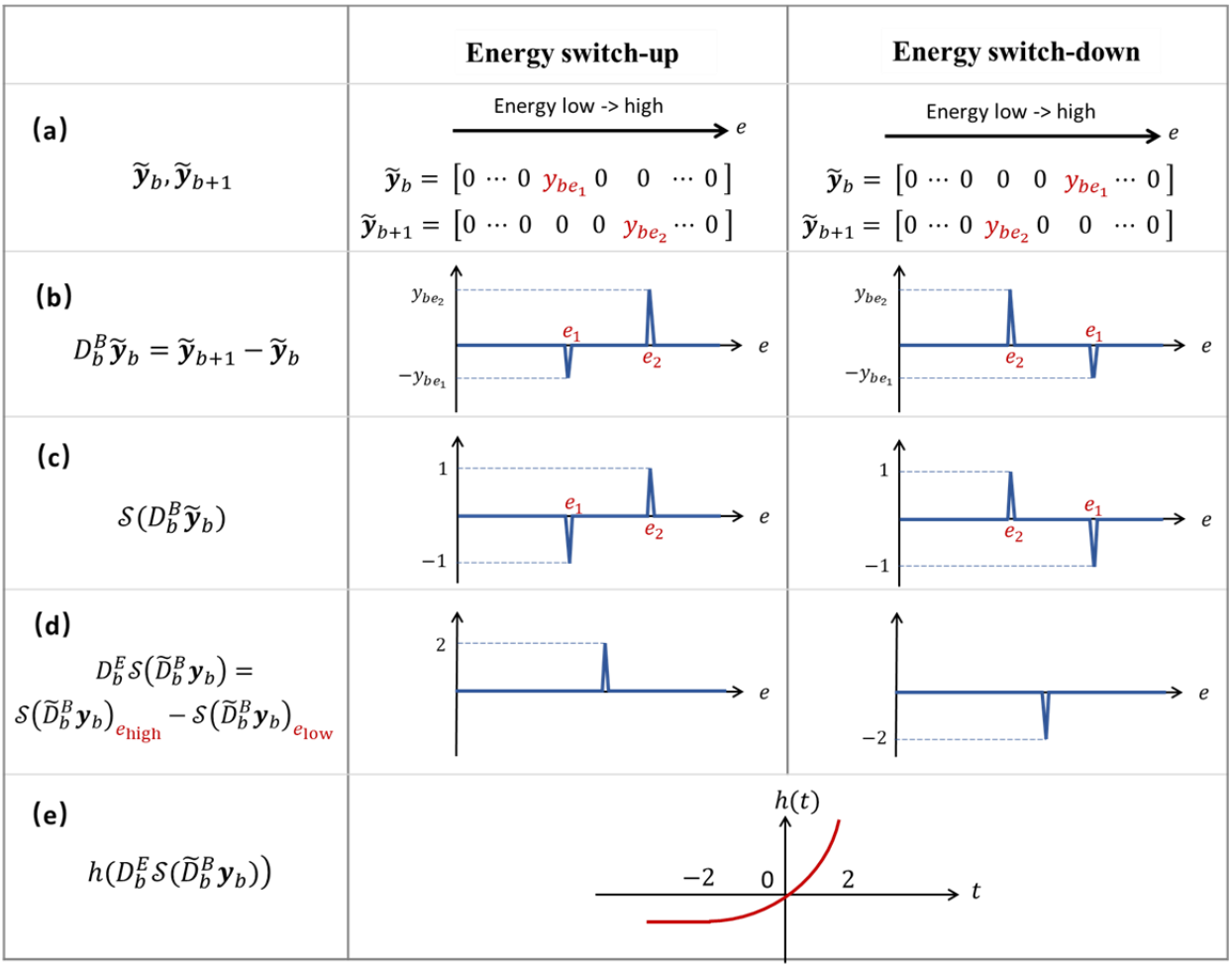

Figure 3-11: Schematic workflow of energy-sequencing regularization. Two adjacent beams, beam b and beam b + 1, are shown, with the situation of energy switch-up in the middle column and energy switch-down in the right column. (a) $\tilde{y}_b$ and $\tilde{y}_{b+1}$ are shown as row vectors, while energy increases from left to right. y_{be_1} and y_{be_2} are the sole nonzero element of $\tilde{y}_b$ and $\tilde{y}_{b+1}$, respectively. (b) $D_b^B \tilde{y}_b = \tilde{y}_{b+1} - \tilde{y}_b$. (c) $D_b^B \tilde{y}_b$ is normalized to -1 or $+1$. (d) Take the gradient of $\mathcal{S}(D_b^B \tilde{y}_b)$ along e direction, which is the difference of the element at high e index and that at low e index. It yields a positive gradient when energy switches up and a negative gradient when energy switches down. (e) The positive gradient is penalized harder to encourage less energy switch-up.

3.3.2.4 Evaluations

ELO-SPAT was tested on one frontal base-of-skull (BOS) patient, one chordoma (CHDM) patient with the simultaneous integrated boost, one bilateral head-and-neck (H&N) patient, and one lung (LNG) patient. A full arc was used for the H&N case, and a partial arc was used for the rest of the cases. Gantry rotation was assumed clockwise. The

control points for individual beams were spaced 2° in the LNG case and 2.5° otherwise. Dose calculation for the scanning spots covering the PTV and a 5 mm margin was performed using matRad[51,52]. We assumed a constant RBE of 1.1. IMPT plans with 2~4 manually selected beams were created for these tested patients for comparison. The prescription dose, target volume, arc range and IMPT beam angles for each patient are shown in Table 3-7 .

To investigate the effectiveness of ES regularization, we created two proton arc plans for each patient with or without ES regularization in Equation 3-17. In the latter plan, the energy layers were selected by group sparsity and log barrier regularization, but not sequenced.

We compared the ELO-SPAT plans against the arc plans created using the SPArc method proposed by Ding et al[63]. Because robustness is not considered yet in this work, the robust optimization used in SPArc is replaced with a conventional PTV-based fluence map optimization for a fair comparison. We created the SPArc plans using arc setting and dose calculation identical to ELO-SPAT. To match the two plans, in SPArc optimization, we pushed the number of layers per beam to be 1 for as many beams as possible, leaving only a few beams to have two energy layers. Similarly, we created SPArc plans with[122] or without energy sequencing to compare with ELO-SPAT. The SPArc method with or without ES is denoted as SPArc-ES or SPArc-noES, and the ELO-SPAT method with or without ES is denoted as ELO-ES or ELO-noES.

All plans were normalized to deliver prescription dose to 95% of volume. The time spent on treatment planning and delivery of ELO-SPAT and SPArc were compared. For

delivery, because energy layer switching time (ELST) is the major factor affecting the total delivery time, we use the total time spent on energy layer switching as the surrogate. The times required for switching energy up, down and keeping it unchanged were 5.5 s, 0.6 s, and 0 s according to the IBA Proton Dynamic Arc Delivery (PDAD) module. A constant ELST of 2.1 s was also used for calculation according to the M.D. Anderson proton therapy system[109] with a synchrotron accelerator.

Case		Prescription Dose (GyRBE)	PTV Volume (cc)	Arc angle (degree)			IMPT (gantry, couch) angle (degree)
				Start angle	Stop angle	Spacing	
BOS		56	66.8	225	135	2.5	(60, 273), (270, 0), (90, 0), (180, 0).
CHDM	PTV63	63	128.9	225	135	2.5	(60, 273), (270, 0), (90, 0), (180, 0).
	PTV74	74	26.6				
H&N	PTV54	54	179.1	180	180	2.5	(0, 0), (160, 0), (200, 0).
	PTV60	60	204.4				
LNG		42	297.8	160	0	2	(180, 0), (315, 0).

Table 3-7: Prescription doses, PTV volumes, arc range and IMPT beam angles for each patient.

3.3.3 Results

3.3.3.1 Optimization and delivery efficiency

The dose calculation and optimization were performed on a Xeon 28-core CPU server operating at 2.40 GHz clock, with Matlab and its Parallel Computing Toolbox. The energy layer delivery sequence for each patient using SPArc or ELO-SPAT with and without

energy sequencing is shown in Figure 3-12. The number of energy switches for different arc plans is plotted in Figure 3-13. The optimization time and expected delivery time are also shown in Figure 3-13.

Without ES regularization, although single energy layer at each control point is achieved, the energies are not ordered, resulting in 40 to 60 switch-up for the tested cases, adding a non-trivial amount of time to delivery. With ES regularization, the energy layer was sequenced to reduce the number of energy switch-up to fewer than 20, which are comparable to that of the SPArc plans with energy sequencing. Despite the similar number of energy up-switching, SPArc uses a regular sequencing pattern with the same number of down-switching between up-switchings. In comparison, ELO-SPAT sequencing patterns vary to meet dosimetric optimization needs. For the synchrotron plans, with a constant ELST of 2.1 s, the total ELST of ELO-ES plans was similar with the ELO-noES plans, with an averaged time reduction of 9%. For cyclotron plans, with an ELST-up of 5.5 s and ELST-down of 0.6s, the total ELST was reduced to around 2 min in ELO-ES from the 4-7 min in ELO-noES, with an averaged reduction of 61%. Meanwhile, considering both the ELO-SPAT and SPArc plans with ES, the ELO plans had 10-30 more unchanged energies between adjacent control points compared with SPArc, therefore leading to an averaged reduction of total ELST time by 24% for the synchrotron plans and by 14% for the cyclotron plans.

In addition to efficient delivery, the ELO-SPAT reduced the runtime of optimization by 84% on average, from the 0.5-2 hours in the SPArc plans to 5-30 min.

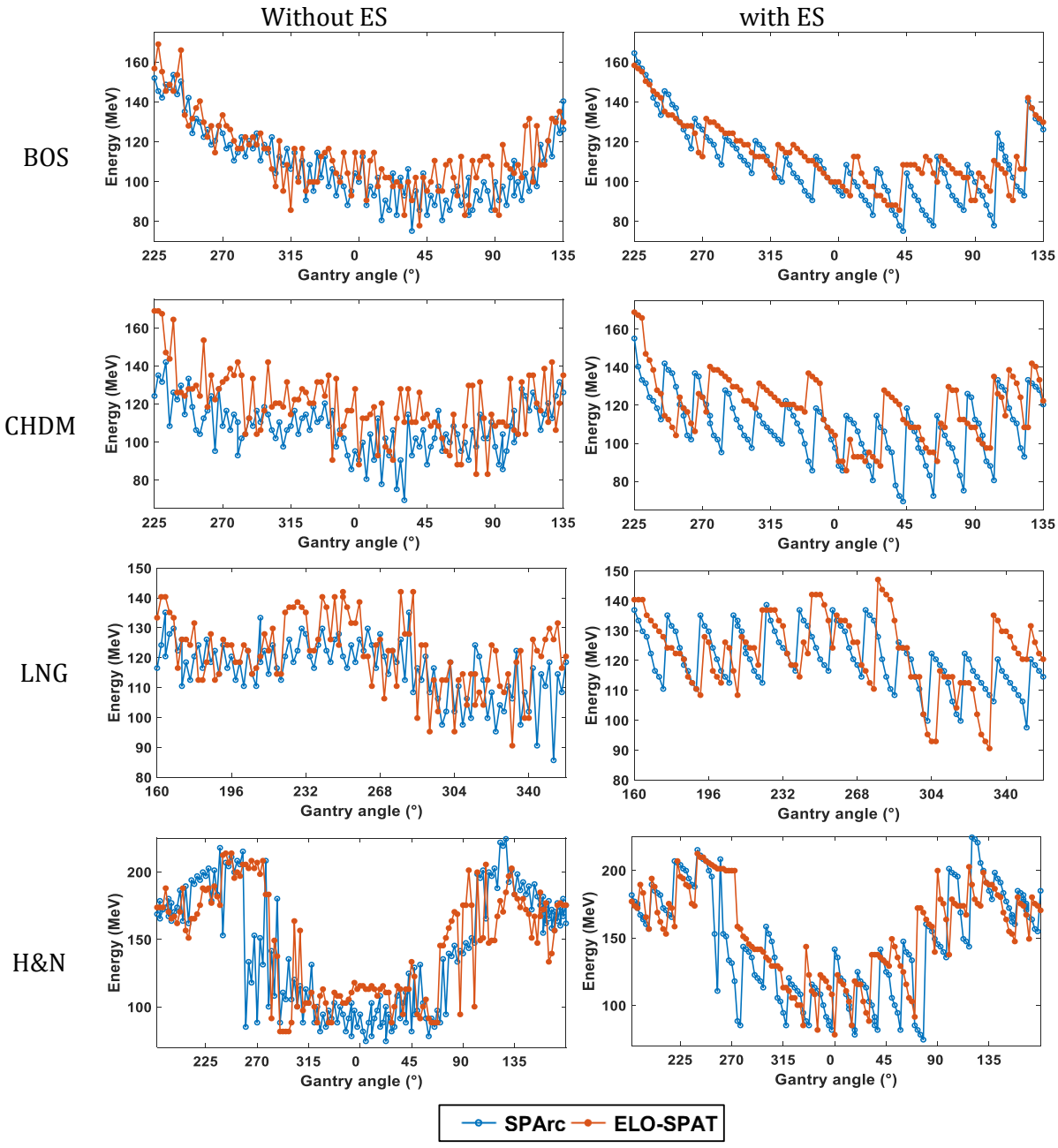

Figure 3-12: Energy layer delivery trajectory comparison between SPArc (blue) and ELO-SPAT (red). The gantry rotates in clockwise following the angle of x-axis from left to right. The delivery sequences without ES are shown in the left column and with ES in the right column.

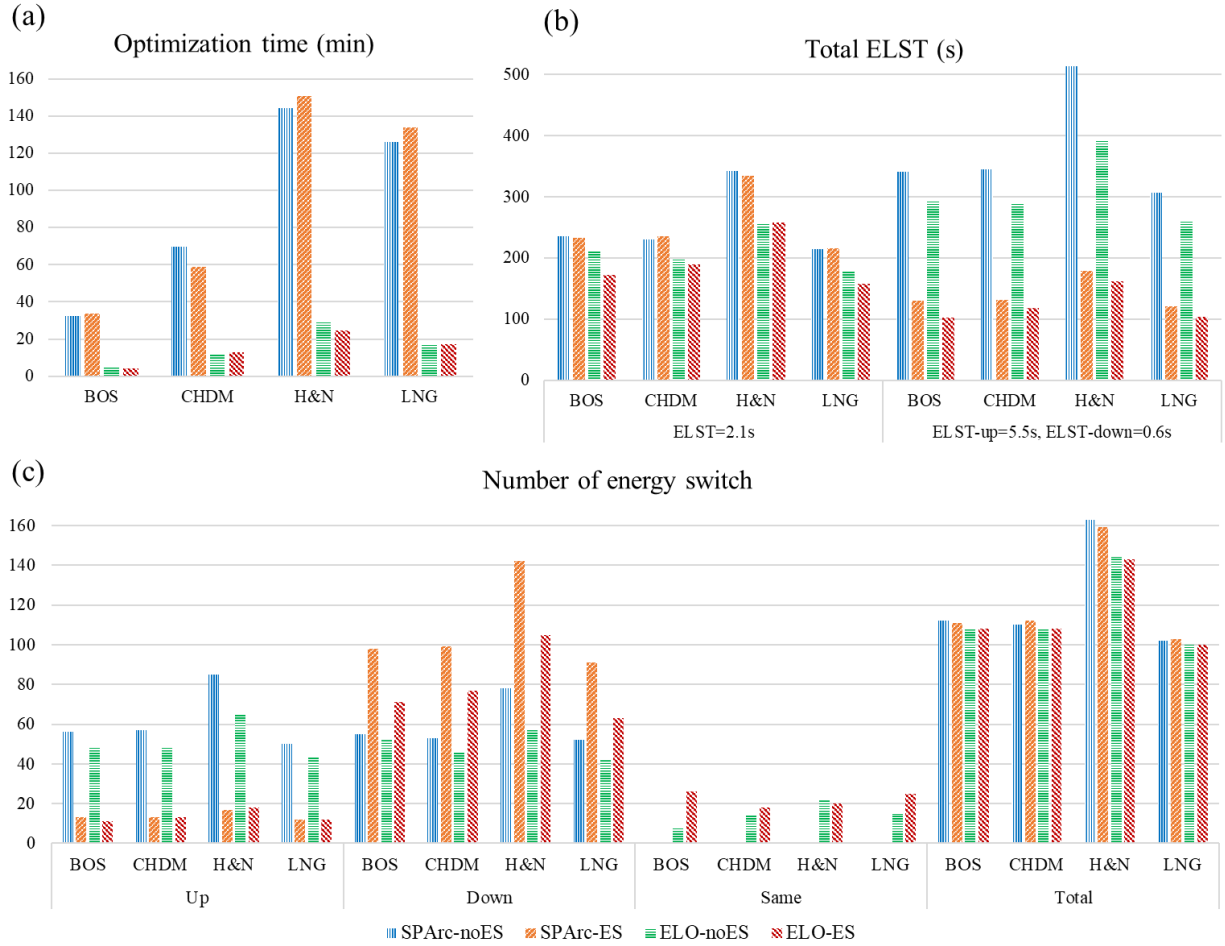

Figure 3-13: (a) Optimization runtime of the four arc plans. (b) The total ELST time of the four arc plans when the ELST time is 2.1s (left) and the ELST-up is 5.5 s and ELST-down is 0.6s (right). (c) The number of energy switch up, down, and staying the same. The total number of energy switches is also plotted.

3.3.3.2 Optimization and delivery efficiency

The DVH comparison of ELO-SPAT and SPArc without ES is shown in Figure 3-14, the DVH comparison of the two arc plans with ES is shown in Figure 3-15. The IMPT plan is plotted in both figures for comparison. The mean dose and max dose for several selected OARs were evaluated in the four arc plans and their differences from the IMPT plans are plotted in Figure 3-16.

All compared plans achieved similar PTV dose coverage. Qualitatively, both the ELO and SPArc plans achieved better sparing compared with the IMPT plans for most OARs, either with or without ES. But in the lung case, the low dose region of the right lung is larger in the arc plans compared with the IMPT plans.

Without ES, the ELO-SPAT plans further improved the OAR sparing compared with the SPArc plans. Lower DVH lines are observed in the ELO-SPAT plans. For example, in the CHDM case, the maximum dose to the left and right cochleas were reduced by 8.1 GyRBE and 6.4 GyRBE, respectively. In the lung case, the maximum dose to the spinal cord was reduced by 5.6 GyRBE. On average, the ELO-SPAT plans without ES reduced the [Dmean, Dmax] of the OARs by [1.6, 3.3] GyRBE from the SPArc plans without ES.

While adding ES regularization, the dosimetry of the quality of ELO-SPAT plans degraded but was still slightly better than the SPArc plans with ES. For example, in the CHDM case, the maximum dose to the left and right cochleas were reduced by 2.8 GyRBE and 4.4 GyRBE, respectively. In the lung case, the maximum dose to the spinal cord was reduced by 2.6 GyRBE. On average, the ELO-SPAT plans with ES reduced the [Dmean, Dmax] of the OARs by [1.4, 2.3] GyRBE from the SPArc with ES.

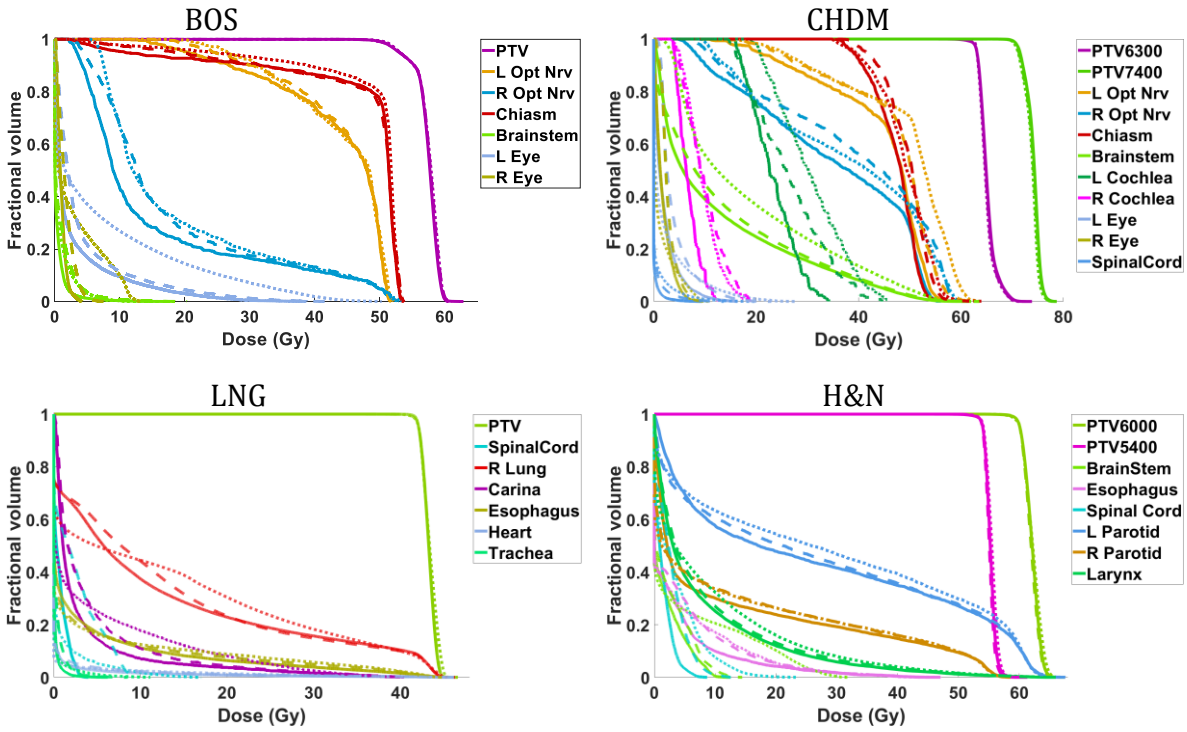

Figure 3-14: DVH comparison of plans without ES. The ELO-SPAT plan is in solid line, the SPArc plan is in dashed line, and the IMPT plan is in the dotted line.

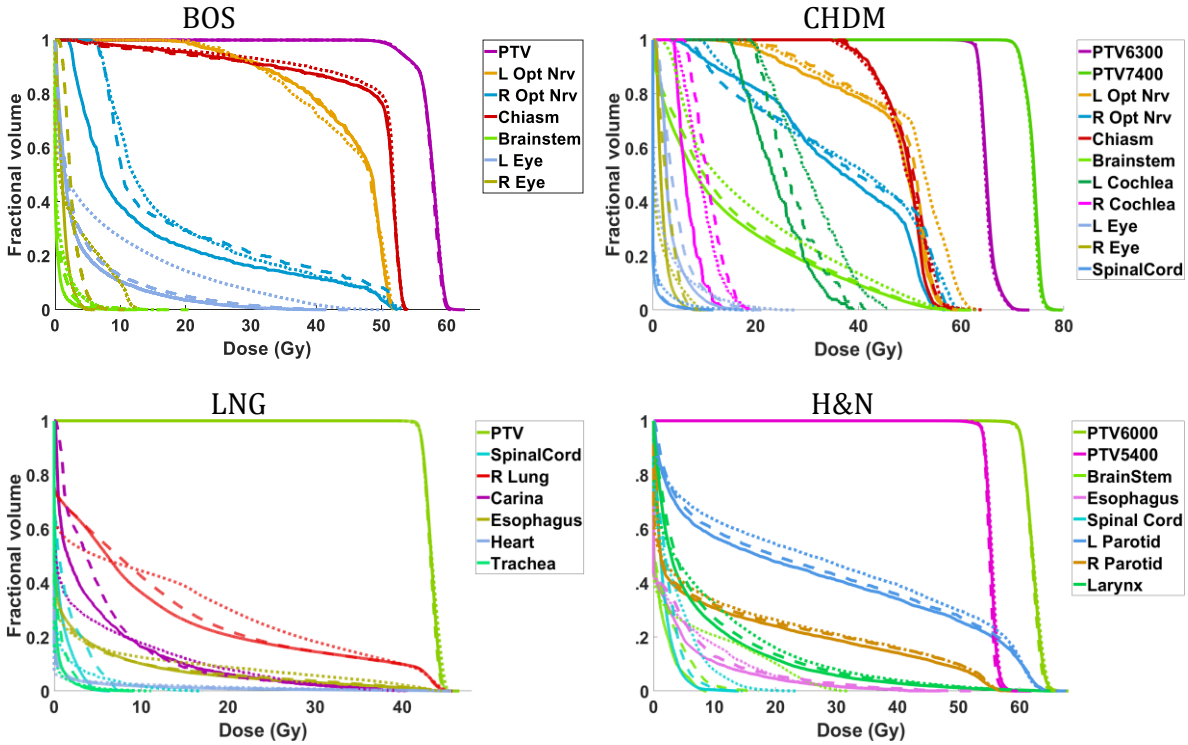

Figure 3-15: DVH comparison of plans with ES. The ELO-SPAT plan is in solid line, the SPArc plan is in the dashed line, and the IMPT plan is in the dotted line.

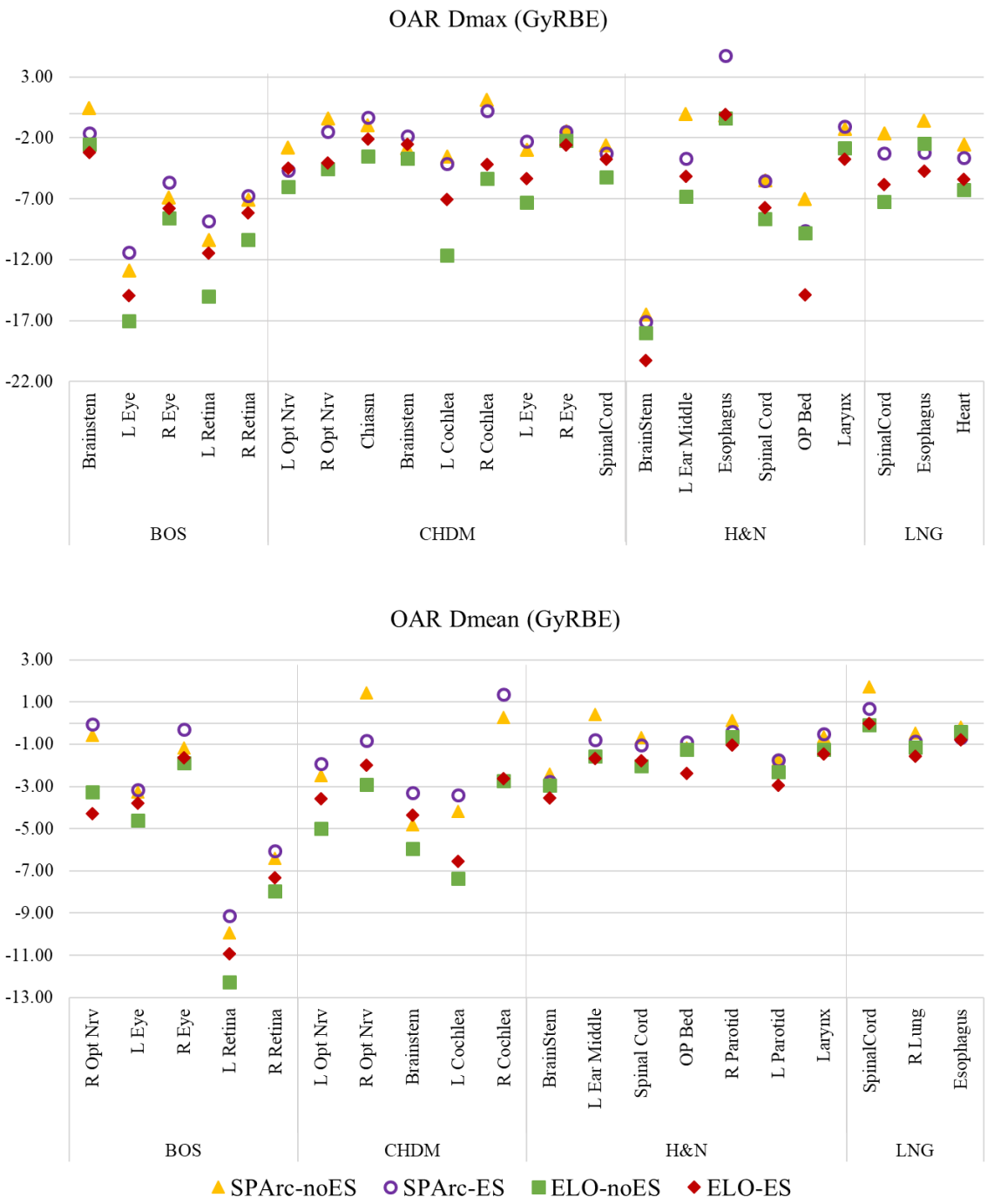

Figure 3-16: The difference of OAR Dmax (top) and Dmean (bottom) in the four arc plans from IMPT. A negative value represents a reduction from the IMPT plan, and a positive value represents an increase.

3.3.3.3 Convergence and effect of γ

A convergence plot of the ELO-SPAT method, for the BOS patient, is shown in Figure 3-17. The cost of each component in Equation 3-17 during the iterations is also plotted. Dose fidelity, group sparsity, and log barrier all converged. The group sparsity is the component with the highest value because the tuning parameter α_{be} need to be large enough to make only one layer selected per control point. Since the differential matrices D_b^E and D_b^B were updated after every iteration, the cost on ES fluctuated during the iterations. But in general, the ES cost started from a high value, meaning lower delivery efficiency, converged to a low value, presenting higher delivery efficiency.

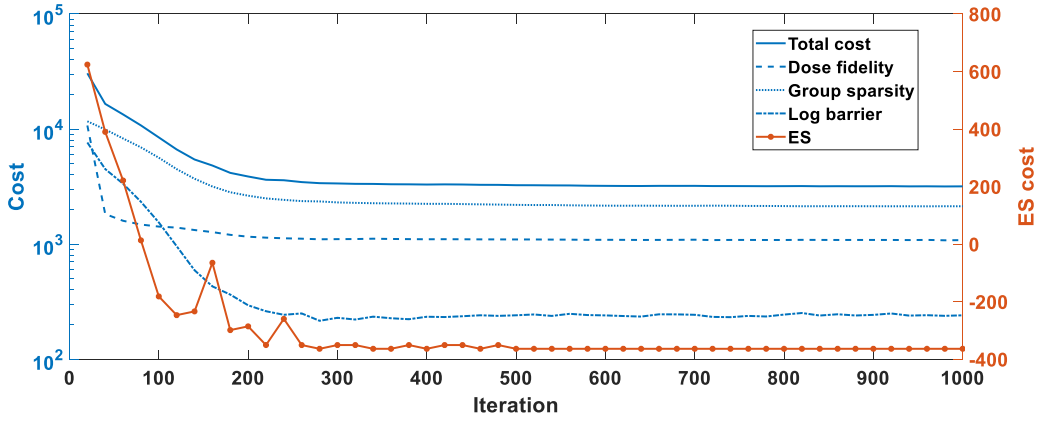

Figure 3-17: Convergence plot of the ELO-SPAT method, for the BOS patient. The total cost and the cost components of dose fidelity, group sparsity, and log barrier are shown in the logarithm scale following the y-axis on the left. The ES cost is shown on a linear scale following the axis on the right.

Figure 3-18 shows how the ES weighting parameter γ affects the number of switch-up and the value of dose fidelity cost. Generally, when γ increases from zero, the number of energy switch-up decreases and the dose fidelity increases. When gamma reaches a certain value, such as 16 in this case, the number of switch-up plateaus, while the

fidelity cost is still in the trend of increasing. In this study, $\gamma = 16$ is picked for a minimum number of energy switch-up and the highest delivery efficiency.

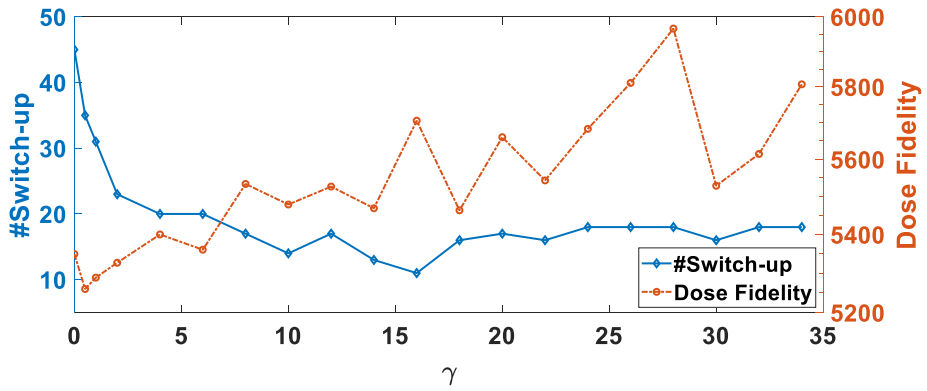

Figure 3-18: The number of energy switch-up and the value of final dose fidelity versus γ, for the BOS patient.

3.3.4 Discussion

We present an integrated energy layer optimization method for scanning-spot proton arc therapy. The novel framework allows an integrated optimization of fluence map optimization, global search of candidate energy layers, and the delivery sequence. The energy-sequencing penalty is added as a soft regularization to dose fidelity term, thus providing a flexible trade-off between dosimetry and delivery speed. In this work, the final ELO-SPAT plans were selected as the ones with the lowest achievable number of energy switch-up for best delivery efficiency, with slight scarification of dosimetry. In clinical practice, the balance between dosimetry and delivery time can be tuned case-by-case. In the case of synchrotron where there is no difference between the time between energy layer switch up and down, the ES regularization can be removed for superior dosimetry.

The number of energy switch using the SPArc energy sequencing method is comparable to our optimization result, showing good performance with the heuristic

method. On the other hand, energy switching patterns are distinctly different. Compared with the regular SPArc pattern, the energy sequencing pattern using ELO-SPAT is flexible to take advantage of the patient and arc geometry. For example, for the patient with a frontal BOS tumor, when the gantry rotates in clockwise from posterior to anterior in the first half arc, the overall energies of the candidate layers decrease because the tumor becomes shallower from the beam's eye view. The ELO-SPAT algorithm exploits the geometry and makes more switch-down before energy going up. The flexibility can facilitate future arc trajectory optimization to further enhance efficiency and dosimetry.

With a similar number of energy switch-up between the ELO-SPAT and SPArc plans, we observed 15-20% less total energy layer switching time in the ELO-SPAT method. This is because SPArc does not allow the energy to stay at the same level between control points due to the progressive sampling scheme and the way energy layers are distributed. Using ELO-SPAT, energy switching-down, and unchanged are both encouraged, thereby shortening the total energy layer switching time.

Another major benefit of ELO-SPAT is the significantly shortened optimization time by 5-10 fold from SPArc. In the progressive SPArc sampling scheme from coarse to fine control point resolution, repetitive fluence map optimization is required after either energy layer filtration or redistribution, resulting in long optimization runtime. In the ELO-SPAT optimization, although the algorithm starts with all candidate layers, the number of active layers is gradually reduced to the desired number, shrinking the size of the dose matrix needed for calculation during computation, shortening the time for each iteration. Furthermore, ELO needs only one run to obtain the final delivery sequence and fluence map. The optimization was further accelerated by FISTA, which converges at an optimal

rate of $O(1/k^2)$ [56]. We expect additional acceleration using the graphics processing unit (GPU) platform and multi-resolution sampling of the dose matrix.

Due to the L2,1/2-norm for group sparsity and the sigmoid function for ES, Equation 3-17 is highly non-convex. Originally, FISTA has been used to solve convex problems. However, recent FISTA work[35,38] and the convergence results (Figure 3-17) in this study suggest that FISTA can be used to solve certain non-convex problems with stable convergence. The differential matrices D_b^E and D_b^B need to be updated after every iteration, but their changes are gradual. Regardless of the ES cost fluctuation, the optimization converges in a few hundred iterations. On the other hand, because of the high non-convexity and the need to update D_b^E and D_b^B, a high weighting parameter on the ES term does not necessarily promote high delivery efficiency. As shown in Figure 3-18, the number of switch-up plateaus after reaching 16 for the BOS case. The main reason is that the problem is trapped in local minima due to the high non-convexity of the problem formulation. Still, the overall trend of decreasing the number of switch-up and increasing fidelity cost is observed can still be used to guide parameter tuning. While a random initialization is used in current work to assign the initial spots intensities during optimization, which is common but not necessarily optimal, other initialization schemes can be explored to improve the convergence of the optimization problem and overcome the local minima problem.

A limitation of the current ELO-SPAT algorithm is that only one energy layer allowed per control point for simplicity, while a few more layers could lead to better dosimetry with small scarification of delivery time. Without the energy-sequencing term, multiple layers are readily achievable by tuning the group sparsity term. However, allowing

multiple energy layers complicates energy sequencing in the current framework that only regulates the layer with maximal weight at each control point. This is a point for future improvement.

In the current problem formulation, the energy switching pattern is used as the surrogate of delivery time and the order of energy switching is penalized by a simple one-sided quadratic Equation 3-19. In future work, Equation 3-19 can be designed to directly correlate the cost with machine-specific energy switching time, thereby allowing intuitive control of actual delivery time. Furthermore, the delivery time penalty can be incorporated as a hard constraint instead of the soft regularization in the current framework for the planner to specify the maximal permissible delivery time directly.

3.3.5 Conclusion

We developed a computationally efficient spot-scanning proton arc method based on the group sparsity penalty and the novel energy-sequencing regularization. It solved energy layer selection and sequencing in an integrated optimization framework, generating proton plans with good dosimetry and high delivery efficiency.

4 INCORPORATING PHYSICAL DOSE ROBUSTNESS INTO BEAM ORIENTATION OPTIMIZATION

4.1 Robust Fluence Map Optimization for IMPT with Soft Spot Sensitivity Regularization[37]

4.1.1 Introduction

Due to the sharp drop-off at the proton Bragg peak[2] and the beam-by-beam dose heterogeneity in the MFO plans[49], IMPT is more susceptible to patient positioning errors or proton beam range uncertainties[7–11]. If the setup and range uncertainties are unaccounted for, dose to the tumor or OARs can substantially differ from what is indicated in the treatment plan. Different from X-ray treatment planning, the proton dose deviation can happen not only at the target boundaries but also inside the target, making traditional PTV

-based optimization, which expands the clinic target volume (CTV) by a safety margin, ineffective for IMPT[48].

Several approaches have been developed to address this problem. Rather than a constant margin, a beam-specific PTV[123] is introduced, to vary the margin based on the field and tissue property for passive scattering and single-field uniform dose IMPT (SFUD-IMPT). Nevertheless, this approach is inapplicable to MFO-IMPT. A theoretically appealing way to account for uncertainties was reported to calculate the dose distribution under random perturbations and optimizes the expectation value of the objective function[10,11]. However, due to the large statistical sampling required, the probabilistic approach is too slow for practical use. An alternative probabilistic approach is analytical probabilistic modeling (APM)[124,125], which uses a Gaussian pencil beam dose calculation algorithm to generate closed-form propagation of probability distributions to quantify uncertainty input for probabilistic optimization. APM is faster because scenario-sampling is not required, but estimation of the covariance requires non-trivial amount of computational resource that increases the optimization time. Furthermore, APM is incompatible with non-model-based pencil beam dose calculation, e.g., Monte Carlo, that is particularly important in handling the lateral dose profile and tissue heterogeneity in proton treatment planning. Coverage optimized planning[126] is also a probabilistic treatment planning based method, which uses dose coverage histogram criteria to replace PTV margin and improves target dose coverage against geometric uncertainties, e.g. setup error. Nonetheless, range uncertainty is not considered in this method. Alternatively, Pflugfelder et al[127] proposed to use a heterogeneity number to quantify lateral tissue heterogeneity of single scanning spot, and incorporated it in the inverse optimization to suppress the spots with a high heterogeneity

number. This empirical method only considers the effect of tissue lateral heterogeneity to setup uncertainty without accounting for the range uncertainties. This method later pivoted towards beam angle selection[128,129] , a separate problem from our current focus of robust scanning spot intensity optimization.

Presently, a class of methods referred to as "worst-case robust optimization" is more commonly used to handle setup and range uncertainties[10,49,130–139]. Instead of considering all possible variations, the worst-case method penalizes the maximal dose deviation for the estimated worst positioning and range estimation errors, to ensure acceptable dose distribution in these cases. In practice, the worst-case approach has reduced plan sensitivity to uncertainties, but on the other hand increased computational cost. Furthermore, the worst cases use generic estimation that may not be applicable to all cases. The actual patient anatomical and range uncertainties may still exceed the estimation, causing unexpected dosimetric deviations.

In this work, we aim to overcome these limitations and develop a novel mathematical framework to exploit the intricate balance between the proton scanning spot distribution, robustness and dose conformality. The plan robustness is incorporated as a sensitivity term in IMPT fluence map optimization, which minimizes the dose deviation from ideal dose distribution and penalizes the combination of scanning spots with high sensitivity.

4.1.2 Method

The sensitivity-based robust optimization problem is formulated with a dose fidelity term and a robustness regularization term. The details are described as follows.

4.1.2.1 Sensitivity analysis

The dose calculation matrix, or the dose influence matrix, denoted as A, contains the vectorized dose information delivered to the patient volume from scanning spots of unit intensities. In this study, the position of individual scanning spots is denoted by the location of the Bragg peak in the patient volume. The sensitivity of a spot is determined by the magnitude of dose distribution for the perturbation due to patient position and range variations. To make the plan more resilient to changes, we penalize the spot position combinations with high sensitivity. The spatial dose gradient, which is used as a surrogate of spot sensitivity, is mathematically described as follows.

As shown in Figure 4-1, a coordinate system $(\boldsymbol{u}_b, \boldsymbol{v}_b, \boldsymbol{w}_b)$ is first designated for the beam b, with the origin centered at the isocenter. $\boldsymbol{u}_b$ represents the beam direction pointing from the source to the isocenter, and $\boldsymbol{v}_b$ and $\boldsymbol{w}_b$ are orthogonal vectors in the plane perpendicular to the beam direction. We define $\boldsymbol{p}_{b,i}$ as the spatial position of scanning-spot i from beam b, which points from the isocenter to the position of its Bragg peak in the patient. $\boldsymbol{a}_{b,i}$ is the full dosimetric contribution of spot i in beam b to all voxels of the patient, embedded as a column vector in the dose calculation matrix A, and $\boldsymbol{a}_b$ is the submatrix of A that contains only the $\boldsymbol{a}_{b,i}$ for all the spots in the same beam b. Then we evaluate the gradient field of $\boldsymbol{a}_{b,i}$ with respect to the spot position $\boldsymbol{p}$, denoted as $\nabla_{\boldsymbol{p}}\boldsymbol{a}_{b,i}$. If there are m elements (meaning m voxels in the patient volume) in the vector $\boldsymbol{a}_{b,i}$, then $\nabla_{\boldsymbol{p}}\boldsymbol{a}_{b,i}$ is a $3 \times m$ matrix, with each row representing a directional derivative.

Then the directional derivatives of $\nabla_{\boldsymbol{p}}\boldsymbol{a}_{b,i}$ along $\boldsymbol{u}_b$, $\boldsymbol{v}_b$ and $\boldsymbol{w}_b$ are in the respective functional forms:

$$D_{\boldsymbol{u}_b}\boldsymbol{a}_{b,i} = (\nabla_{\boldsymbol{p}}\boldsymbol{a}_{b,i}) \cdot \boldsymbol{u}_b,$$

$$D_{\boldsymbol{v}_b}\boldsymbol{a}_{b,i} = \left(\nabla_{\boldsymbol{p}}\boldsymbol{a}_{b,i}\right) \cdot \boldsymbol{v}_b,$$

$$D_{\boldsymbol{w}_b}\boldsymbol{a}_{b,i} = \left(\nabla_{\boldsymbol{p}}\boldsymbol{a}_{b,i}\right) \cdot \boldsymbol{w}_b.$$

Equation 4-1

This equation set evaluates the dose sensitivity level at each voxel from a specific scanning spot along the longitudinal direction (beam direction) and the lateral directions (orthogonal to beam direction). Since both $D_{\boldsymbol{v}_b}\boldsymbol{a}_{b,i}$ and $D_{\boldsymbol{w}_b}\boldsymbol{a}_{b,i}$ represent the lateral sensitivity, only $D_{\boldsymbol{u}_b}\boldsymbol{a}_{b,i}$ and $D_{\boldsymbol{v}_b}\boldsymbol{a}_{b,i}$ are used for optimization in the following sections.

We can obtain the vector specific to spot i of beam b in each direction, $\boldsymbol{u}_b$ or $\boldsymbol{v}_b$, by simply extracting column i from $D_{\boldsymbol{u}_b}\boldsymbol{a}_b$ or $D_{\boldsymbol{v}_b}\boldsymbol{a}_b$, respectively. After performing this operation on every beam-specific submatrix of the A, we can obtain two sensitivity matrices, written as $D_u A$ and $D_v A$.

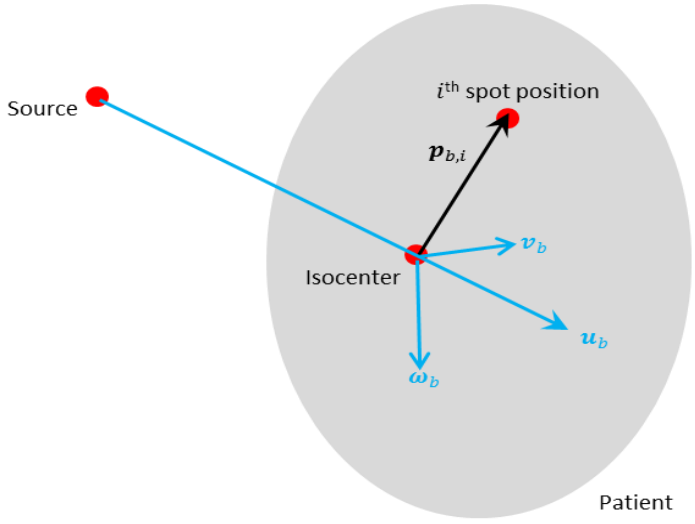

Figure 4-1: Diagram showing the coordinates and the vectors used in spot sensitivity calculation. The beam divergence due to spot lateral distance to the isocenter is exaggerated for illustration purposes. The actual proton system source-to-axis distance is substantially greater than the target size and the individual pencil beams in the same beam direction are nearly parallel.

4.1.2.2 Problem formation

As mentioned before, the spots are penalized based on their sensitivities. With the formation of sensitivity matrices along the beam direction and perpendicular to the beam direction, an intuitive approach is to penalize the L2,2-norm of $D_k A\boldsymbol{x}$, ($k \in \{\boldsymbol{u}, \boldsymbol{v}\}$), which is formulated as:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \Gamma(A\boldsymbol{x}) + \lambda_{\boldsymbol{u}} \|D_u A\boldsymbol{x}\|_2^2 + \lambda_{\boldsymbol{v}} \|D_v A\boldsymbol{x}\|_2^2,$$

$$\text{subject to} \quad \boldsymbol{x} \geq 0,$$

Equation 4-2

where $\boldsymbol{x}$ is the optimization variable representing the scanning spot intensities, $\Gamma(A\boldsymbol{x})$ is the dose fidelity term penalizing the dose deviation from ideal dose distribution as defined in Section 2.2.1, and $\lambda_{\boldsymbol{u}}$ and $\lambda_{\boldsymbol{v}}$ are the sensitivity regularization parameters.

However, the matrix $D_k A\boldsymbol{x}$, ($k \in \{\boldsymbol{u}, \boldsymbol{v}\}$) has the same size as the matrix A, which makes it time- and memory-expensive to solve the problem Equation 4-2. To improve the computational efficiency, as suggested by Ungun et al.[140], an L1-norm is used as a surrogate of the L2,2-norm and column clustering on the sensitivity matrix is performed to reduce the problem size. The problem is then formulated as:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \Gamma(A\boldsymbol{x}) + \sum_{k \in \{\boldsymbol{u}, \boldsymbol{v}\}} \lambda_k \|D_k A\boldsymbol{x}\|_1,$$

$$\text{subject to} \quad \boldsymbol{x} \geq 0,$$

Equation 4-3

Then the absolute values of the rows of $D_u A$ and $D_v A$ are summed up, and the two resulting row vectors are transposed to produce the longitudinal and lateral sensitivity vectors,

denoted as $\boldsymbol{s_u}$ and $\boldsymbol{s_v}$, respectively. Then the sensitivity-regularized robust optimization problem is written as:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \Gamma(A\boldsymbol{x}) + \sum_{k\in\{\boldsymbol{u},\boldsymbol{v}\}} \lambda_k \boldsymbol{s}_k^T \boldsymbol{x}\,,$$

$$\text{subject to} \quad \boldsymbol{x} \geq 0,$$

Equation 4-4

The initial large-scale matrix and vector multiplication in Equation 4-2 is reduced to a vector inner product, which is computationally inexpensive. Moreover, Equation 4-4 is a convex problem and can be solved using FISTA[56].

4.1.2.3 Evaluation

This proposed *Sen*sitivity *R*egularized (SenR) method was tested on three patients with skull base tumor (SBT) and three bilateral head-and-neck (H&N) patients, and was compared against conventional PTV-based optimization method (Conv) and voxel-wise worst-case optimization method (WC)[10,49,132,134]. The voxel-wise worst-case optimization considered nine scenarios, including one nominal scenario and 8 worst-case scenarios. The 8 worst-case scenarios consist of (1) 6 setup uncertainties scenarios, by shifting the beam isocenter by ±3 mm along anteroposterior, superior-inferior, and mediolateral directions; (2) 2 range uncertainties scenarios, by scaling the CT number by ±3%. The worst-case method is solved by a first order primal-dual algorithm known as Chambolle-Pock algorithm[141].

For every patient, the same beam arrangement, scanning spot population scheme and dose calculation engine were used for the three methods. The dose calculation for all

scanning spots covering the CTV and a 5 mm margin was performed by matRad[51,52]. The target volume for worst-case approach was chosen to be CTV, and the conventional method was planned based on the PTV, which was a 3-mm isotropic expansion of the CTV. Our sensitivity-regularized method was applied to both CTV and PTV to investigate the impact of margin in the new optimization framework. The prescription dose, target volume and the beam arrangement are shown in Table 4-1.

Case		Prescription Dose (GyRBE)	CTV Volume (cc)	Beam Angle
SBT #1	CTV63	63	86.07	(270, 0) (90, 0) (180, 0) (60, 275)
	CTV74	74	26.42	
SBT #2		70	36.8	
SBT #3		56	33.7	
H&N #1	CTV54	54	141.29	(0,0) (160,0) (200,0)
	CTV60	60	160.89	
	CTV63	63	68.00	
H&N #2	CTV54	54	108.00	
	CTV60	60	127.26	
H&N #3	CTV54	54	110.38	
	CTV60	60	98.94	
	CTV63	63	10.23	

Table 4-1: Prescription doses, CTV volumes, and the beam angles (gantry, couch).

The nominal dose distribution and robustness against range uncertainties and setup uncertainties were both evaluated. The robustness was evaluated by the same 9 scenarios used for worst-case optimization. The DVH band plot, as well as the worst dose metrics occurred among uncertainties scenarios, was used for analysis. In addition to the 3% range uncertainty, a stress test was performed on the normalized CTV volume covered

by the 100% prescription dose for the range estimation error varying from 0.5% to 4.0%, with 0.5% interval.

4.1.3 Results

Using an i7 6-core CPU desktop, the time spent on dose calculation, sensitivity vector evaluation and optimization of each method are list in Table 4-2. Parallel computing toolbox in Matlab was used to accelerate the worst-case dose calculation and sensitivity evaluation. The preparation time before optimization for the WC method and the SenR method was comparable. During optimization, the SenR method using PTV as target volume (SenR-PTV) was as efficient as the Conv method, and it was on average 22 times faster than the WC method. And the SenR plans using CTV as target volume (SenR-CTV) were faster than the SenR-PTV plans due to fewer voxels to consider during optimization. One thing to note is that the computational time comparison is based on the solvers developed in our group, and the actual time of voxel-wise worst-case method will be different in commercial treatment planning system.

Case	Pre-optimization time (s)			Optimization runtime (s)			
	Nominal dose calculation	Worst-case dose calculation	Sensitivity calculation	Conv	WC	SenR-PTV	SenR-CTV
SBT #1	41.8	106.3	129.8	75.5	2296.5	74.1	66.7
SBT #2	30.0	66.8	55.3	73.1	1070.6	73.4	73.0
SBT #3	28.2	75.5	44.7	76.0	909.5	75.1	59.9
H&N #1	210.0	638.1	408.6	129.7	2211.6	129.4	93.8
H&N #2	208.7	649.2	346.6	114.7	2477.0	105.4	80.9
H&N #3	174.3	546.6	408.6	121.5	3269.0	133.0	103.4

Table 4-2: Computational time comparison of the four plans of each patient.

4.1.3.1 Nominal dose comparison

Figure 4-2 shows the nominal DVHs comparison among the WC plans, SenR-CTV plans and SenR-PTV plans for the SBT #1 patient and H&N #2 patient. Several OARs are selected for the SBT and H&N sites, respectively, and the differences of their mean and maximum doses between the SenR plans and the WC plans are presented in Table 4-3 and Table 4-4. Without uncertainties, the Conv, WC, SenR-PTV and SenR-CTV methods achieved similar CTV dose coverage.

The SenR-CTV plans had better OAR sparing compared with the WC plans. For example, in the SBT #1 patient, SenR-CTV reduced the mean dose and max dose of the left cochlea by 20.2 GyREB and 18 GyRBE from WC, and the dose sparing of other OARs were also improved except the max dose to the right optical nerve. In the H&N #1 patient, the doses to the parotids were also lower in SenR-CTV plan compared with the WC plan. The average reduction of [Dmean, Dmax] of the SenR-CTV plans from the WC plans were [4.7, 3.4] GyRBE for the SBT cases and [2.5, 3.3] GyRBE for the H&N cases.

The overall OAR sparing of SenR-PTV was comparable with the WC. For example, in the three SBT cases, the mean and max brainstem doses were both reduced in SenR-PTV relative to WC. SenR-PTV plans also achieved lower Dmax to the left optical nerve and chiasm, but the Dmax to the left cochlea in the SBT #2 patient was greater due to an overlap with PTV. In the H&N case, a reduction of dose in the brainstem, larynx and spinal cord was observed in the SenR-PTV plans. The average reduction of [Dmean, Dmax] of the SenR-PTV plans from the WC plans were of [2.0, 1.5] GyRBE for the SBT cases, and [0.8, 0.8] GyRBE for the H&N cases.

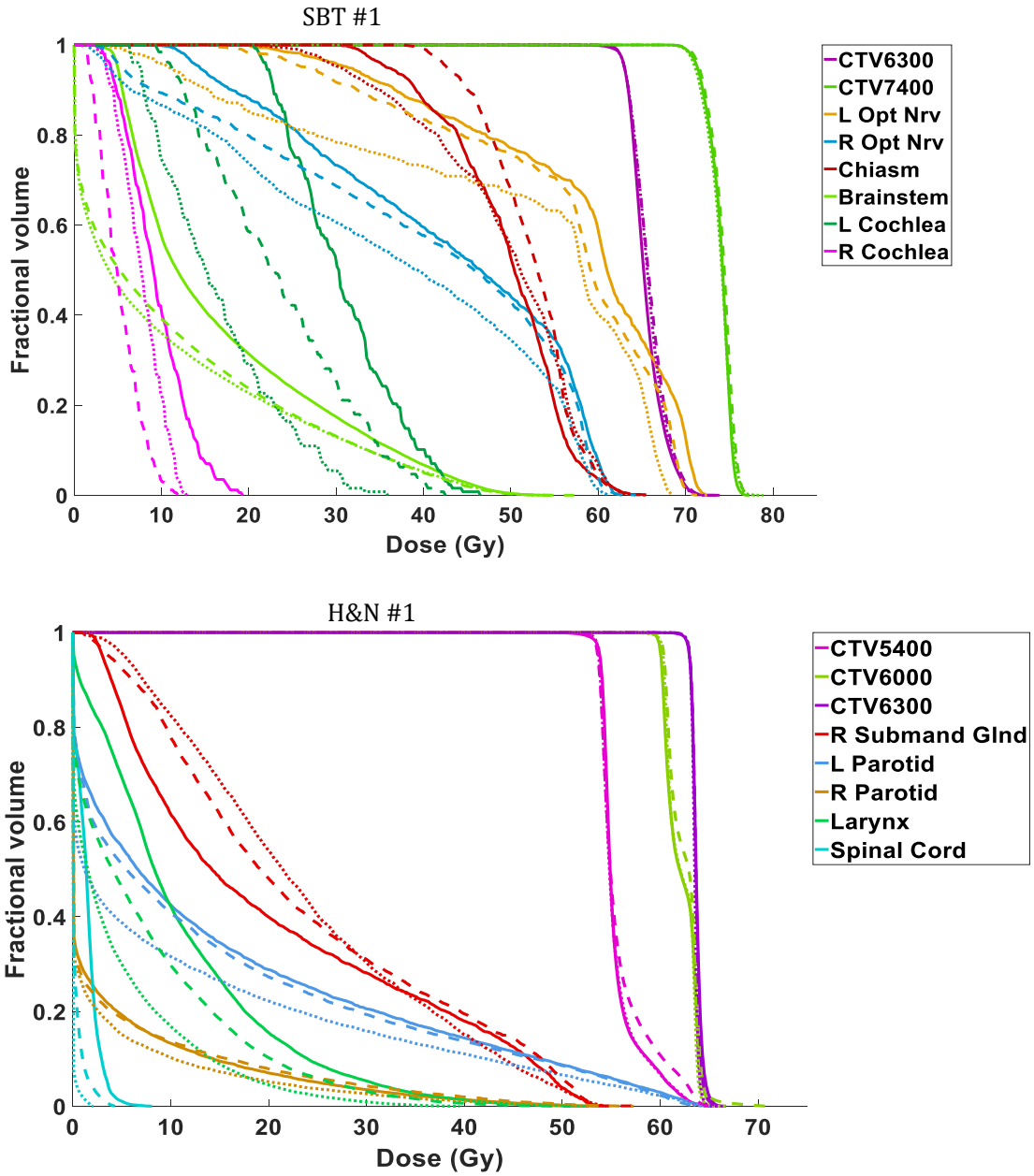

Figure 4-2: Comparison of nominal DVHs for patients SBT #1 and H&N #1 of the WC method (solid), SenR-CTV method (dotted), and SenR-PTV method (dashed).

	SenR-CTV – WC (GyRBE)						SenR-PTV – WC (GyRBE)					
	Dmean			Dmax			Dmean			Dmax		
SBT Case	#1	#2	#3	#1	#2	#3	#1	#2	#3	#1	#2	#3
L Opt Nrv	-8.4	-4.3	-4.8	-3.5	-2.9	-1.1	-2.3	-5.0	+1.1	-1.5	-2.1	-0.7
R Opt Nrv	-4.5	-1.0	-9.9	+0.5	-4.2	-2.7	-0.7	-0.7	-2.0	+1.3	-1.0	+2.0

Chiasm	-1.2	-8.9	-1.6	0.0	-4.2	-0.3	+2.1	-3.3	+0.5	-0.1	-8.0	-0.9
Brainstem	-6.8	-3.6	-1.3	-1.9	-7.6	-3.7	-6.2	-2.5	-1.2	-2.4	-0.9	-2.7
L Cochlea	-20.2	-3.5	0.0	-18.0	-1.0	0.0	-13.1	+6.2	0.0	-9.0	+11.4	0.0
R Cochlea	-4.4	-0.6	0.0	-9.6	-0.8	0.0	-7.0	-2.0	0.0	-11.2	-1.4	0.0

Table 4-3: OAR mean dose and max dose reduction of the SenR plans from the WC plans, for the SBT cases under nominal situation. A negative sign represents a dose reduction from the WC plans.

	SenR-CTV – WC (GyRBE)						SenR-PTV – WC (GyRBE)					
	Dmean			Dmax			Dmean			Dmax		
H&N Case	#1	#2	#3	#1	#2	#3	#1	#2	#3	#1	#2	#3
Brainstem	-0.5	-1.1	-0.2	-2.0	-4.8	-2.6	-0.1	-1.2	-0.1	-1.0	-8.3	-1.6
Constrictors	-3.5	-3.2	-2.9	-10.6	-3.1	-1.2	-1.0	+0.5	+3.3	+3.7	+2.4	+3.2
R Submandibular Gland	+3.4	-16.0	+1.2	0.0	-3.7	-1.2	+3.0	-0.6	-10.0	+0.4	-1.2	-1.5
Larynx	-6.3	-1.9	-2.8	-11.7	-4.9	-7.5	-3.4	-0.8	-2.7	-4.0	-1.0	-6.3
L Parotid	-3.8	-9.5	-1.2	-0.5	-2.1	-0.3	-0.8	-2.1	+1.6	-0.7	-1.1	+1.7
R Parotid	-0.8	-0.7	-0.1	-3.1	-0.2	-0.7	+0.2	-0.8	+1.1	+3.4	-1.8	+3.9
Spinal Cord	-1.3	-1.1	-1.3	-3.3	-2.3	-4.3	-1.1	-0.7	-1.3	-1.5	-0.1	-5.4

Table 4-4: OAR mean dose and max dose reduction of the SenR plans from the WC plans, for the H&N cases under nominal situation. A negative sign represents a dose reduction from the WC plans.

4.1.3.2 Robust analysis

The DVH bands of CTVs and selected OARs indicating the plan robustness with range and setup uncertainties for the SBT #1 and the H&N #2 patients are shown in Figure

4-3 and

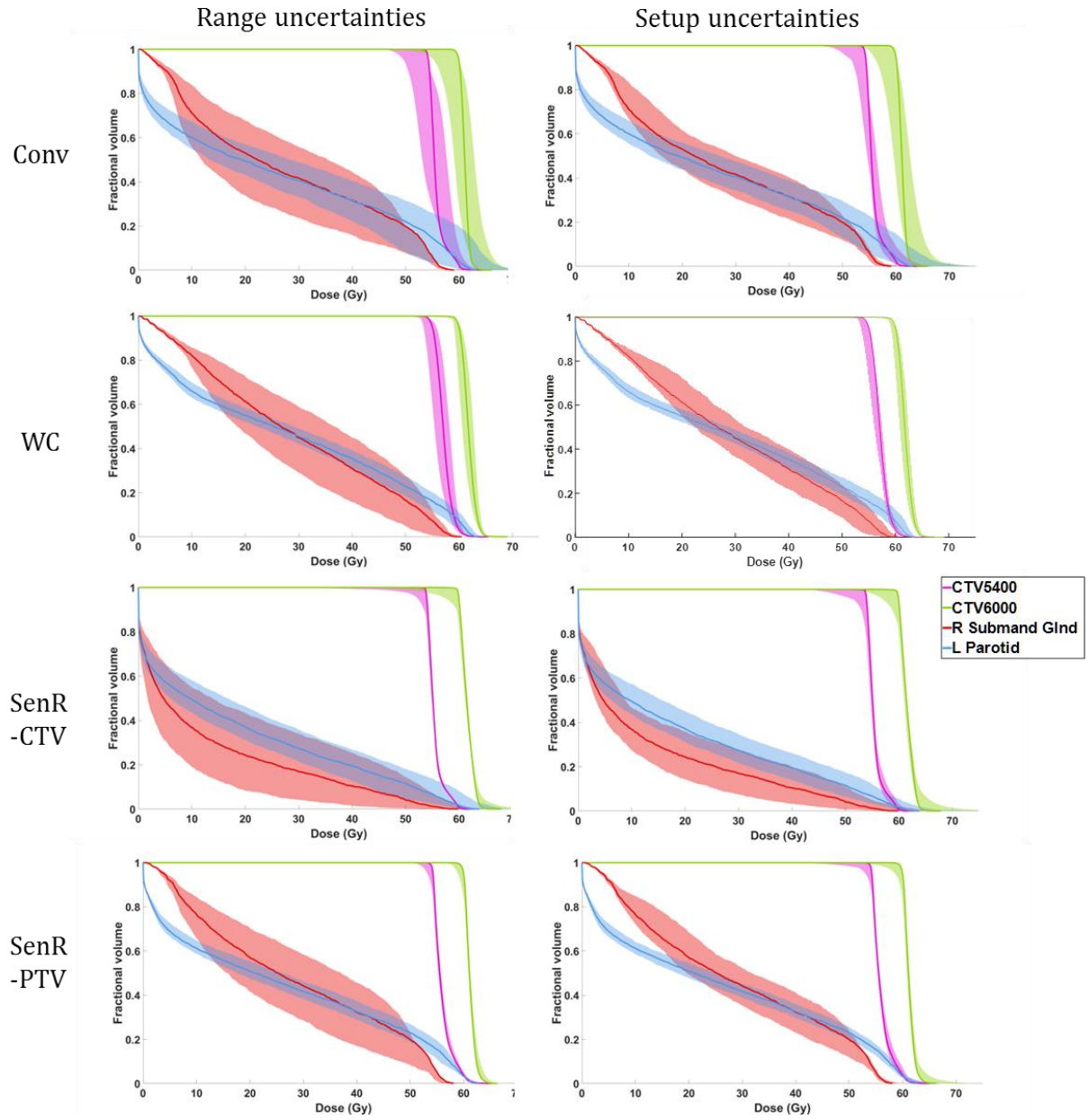

Figure 4-4, where the solid lines in each plot are the DVHs of the nominal case, and the bands bound the worst-case distributions. A narrower band means greater resilience to uncertainties. Qualitatively, both the SenR approach and WC method improved the robustness of CTVs and OARs from conventional PTV-based method for the two disease sites.

With range uncertainties, similar or more compacted CTV bands were observed in the SenR-PTV plans compared with the WC plans. The SenR-CTV plans also resulted in narrow CTV bands, but there was a slightly larger underdosed region of CTV in these plans. The robustness against setup uncertainties was similarly improved by SenR-PTV and SenR-CTV.

In addition to better target volume robustness, a decrease in OAR sensitivity was observed in both the SenR-PTV and SenR-CTV plans. For example, the DVH bands of the optical nerves and optical chiasm in the SBT #1 patient, and the left parotid in the H&N#2 patient are narrower than that in Conv.

Compared with Conv plans, the lowest D95%, V95% and V100% were improved by SenR and WC. Overall SenR-PTV and WC achieved better CTV dose metrics. On average, under range uncertainties, the lowest [D95%, V95%, V100%] of CTV were increased from [93.8%, 88.5%, 47.4%] in Conv, to [99.3%, 99.5%, 86.6%] in WC, [97.7%, 97.9%, 81. %] in SenR-CTV and [98.8%, 99.3%, 85.1%] in SenR-PTV, respectively. Under setup uncertainties, the average lowest [D95%, V95%, V100%] of CTV were increased from [95.4%, 94.9%, 65.1%] in Conv, to [99.4%, 99.6%, 87.1%] in WC, [97.0%, 97.1%, 77.9%] in SenR-CTV and [98.2%, 98.3%, 83.9%] in SenR-PTV, respectively.

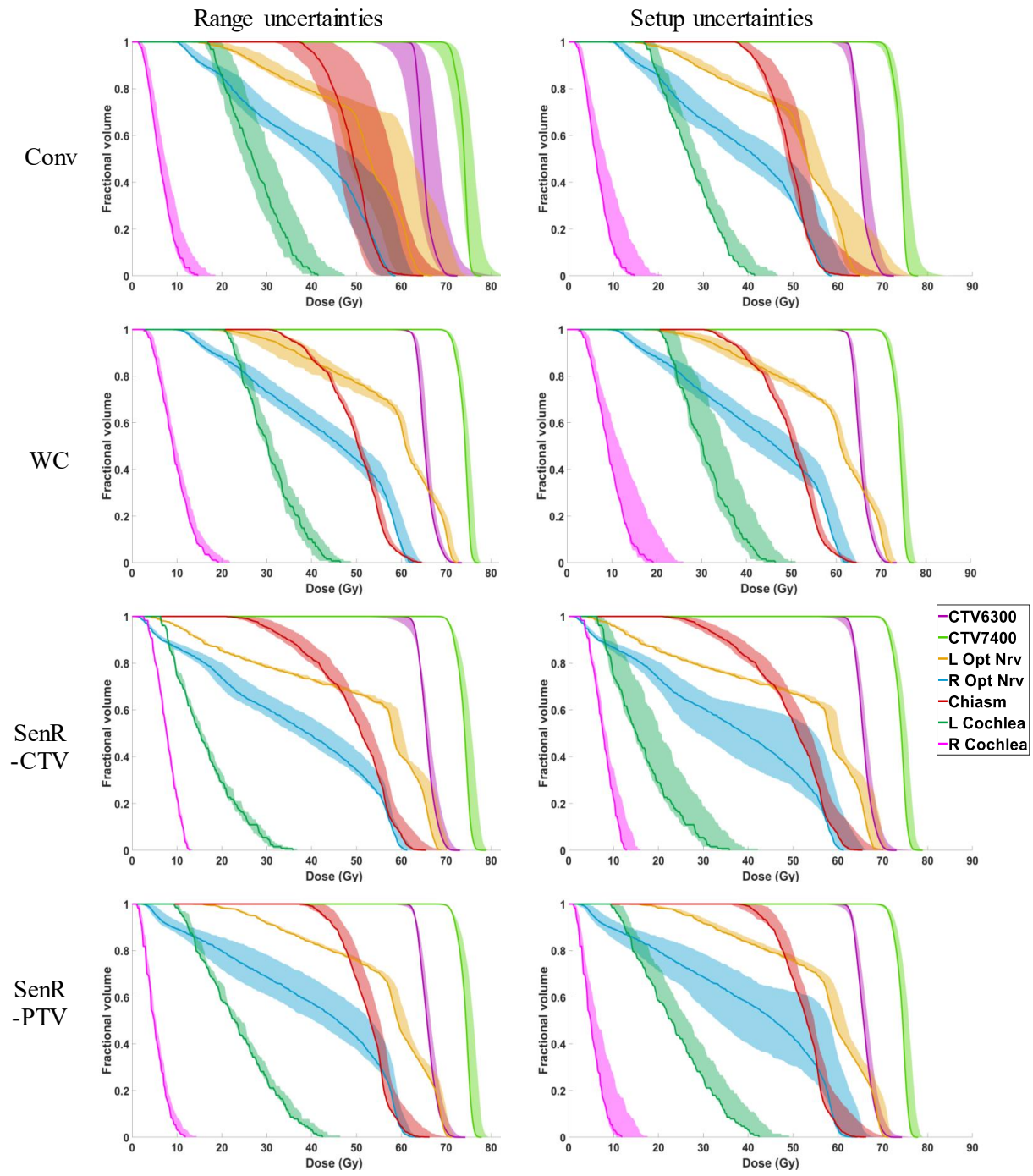

Figure 4-3: DVH bands of the SBT #1 patient including 2 range uncertainties (left column) and 6 setup uncertainties (right column). The first row is Conv plans, the second row is the WC plans, the third row is SenR-CTV plans, and the last row is the SenR-PTV plans.

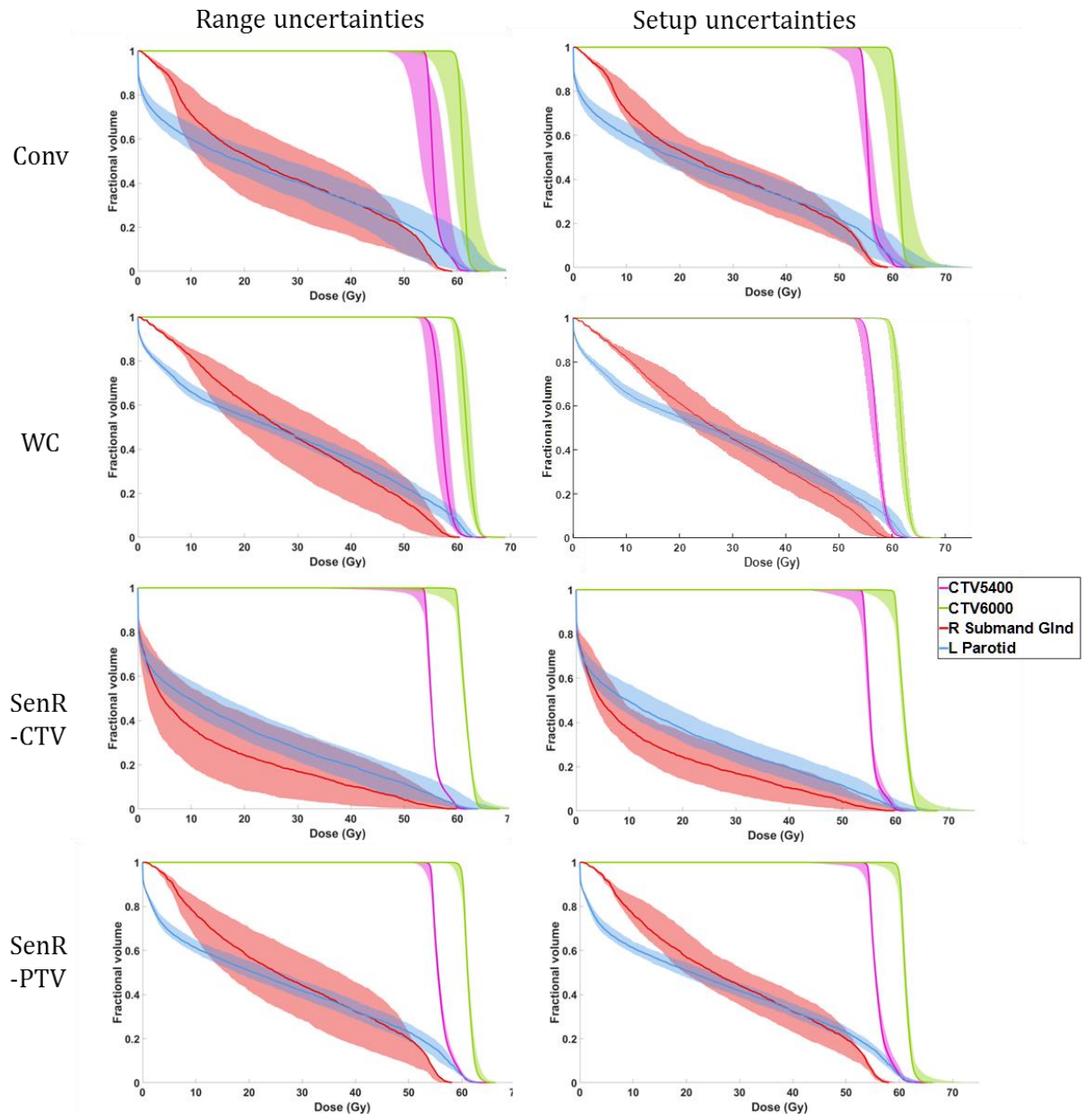

Figure 4-4: DVH bands of the H&N#2 patient including 2 range uncertainties (left column) and 6 setup uncertainties (right column). The first row is Conv plans, the second row is the WC plans, the third row is SenR-CTV plans, and the last row is the SenR-PTV plans.

Figure 4-5 shows the V100% stress test results, where the range estimation error increased from 0% to 4%, which is 1.0% outside of the expected worst case. V100% degrades with increasing range estimation error but the SenR-PTV method shows slower degradation and greater robustness than WC.

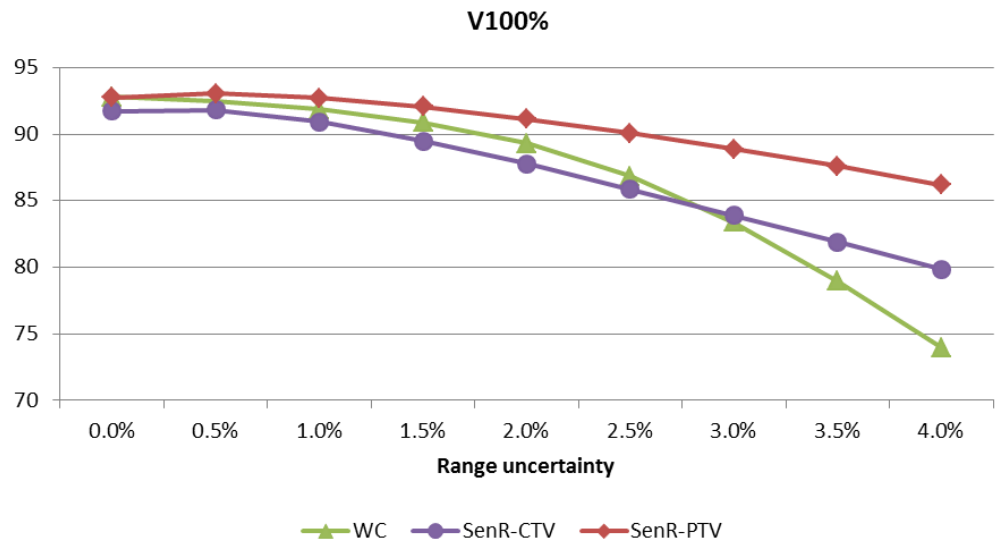

Figure 4-5: The patient-averaged worst V100% of the three methods, when range uncertainty varies from 0.0% to 4.0%.

4.1.3.3 Spot level analysis

In order to better understand the mechanism of SenR method, an analysis on the scanning spot level is demonstrated using the SBT #1 patient as an example. The spot-level dose difference between Conv method and SenR-CTV method when undershooting (+3% range uncertainties) happens is shown in Figure 4-6. In this analysis, a point of interest in the target, which is inside an underdosing area when undershooting, is found and the scanning spots located within 2 cm radius sphere of this cold spot are extracted. These scanning spots from four different beam directions are the main contributors to the dose of the point of interest. The total dose from these local scanning spots is shown in Figure 4-6.A, with the first row being the transverse plane and the second row being the sagittal plane. From left to right, each column represents the Conv nominal, the Conv undershooting, SenR-CTV nominal and SenR-CTV undershooting conditions, respectively. The peak position of the dose distribution in the Conv (SenR) nominal plan is marked $\mathbf{P_1}$ ($\mathbf{P_2}$), denoted as the crosshairs in the first (last) two columns of images in Figure 4-6. $\mathbf{P_1}$

and $\mathbf{P_2}$ are used as the reference points when comparing the dose change when undershooting. For comparison, the isodose display is normalized to the P-point dose of the corresponding nominal case without range error. When the range is overestimated, a 20% reduction in the P-point dose is observed in the Conv case. However, the high sensitivity combination is quantified in the new optimization framework and correctly penalized. As a result, to deliver dose to the same point of interest, a different combination of spots is selected. When the same undershooting happens, the P-point dose only drops 5% in the SenR plan. A closer examination of the scanning spots distribution reveals why the SenR optimized combination is more resilient to the range estimation error. Different from the Conv approach that chooses spots that match their distal edges, in the SenR approach, spots are slightly mismatched. Spots from beam 1 and 4 contribute their proximal edges to **P** and the spot from beam 3 contributes its lateral edge. When the range is over-estimated, the slightly undershoot spots from beam 1 and 4 would retract while the contribution from beam 3 remains unchanged due to the smooth lateral dose profile. This combination is equally resilient to range underestimation due to the same reason that the dose to a given point is contributed by a mixture of distal, lateral, and proximal edges. The last two are not as sharp as the distal edge thus more resilient to range estimation error.

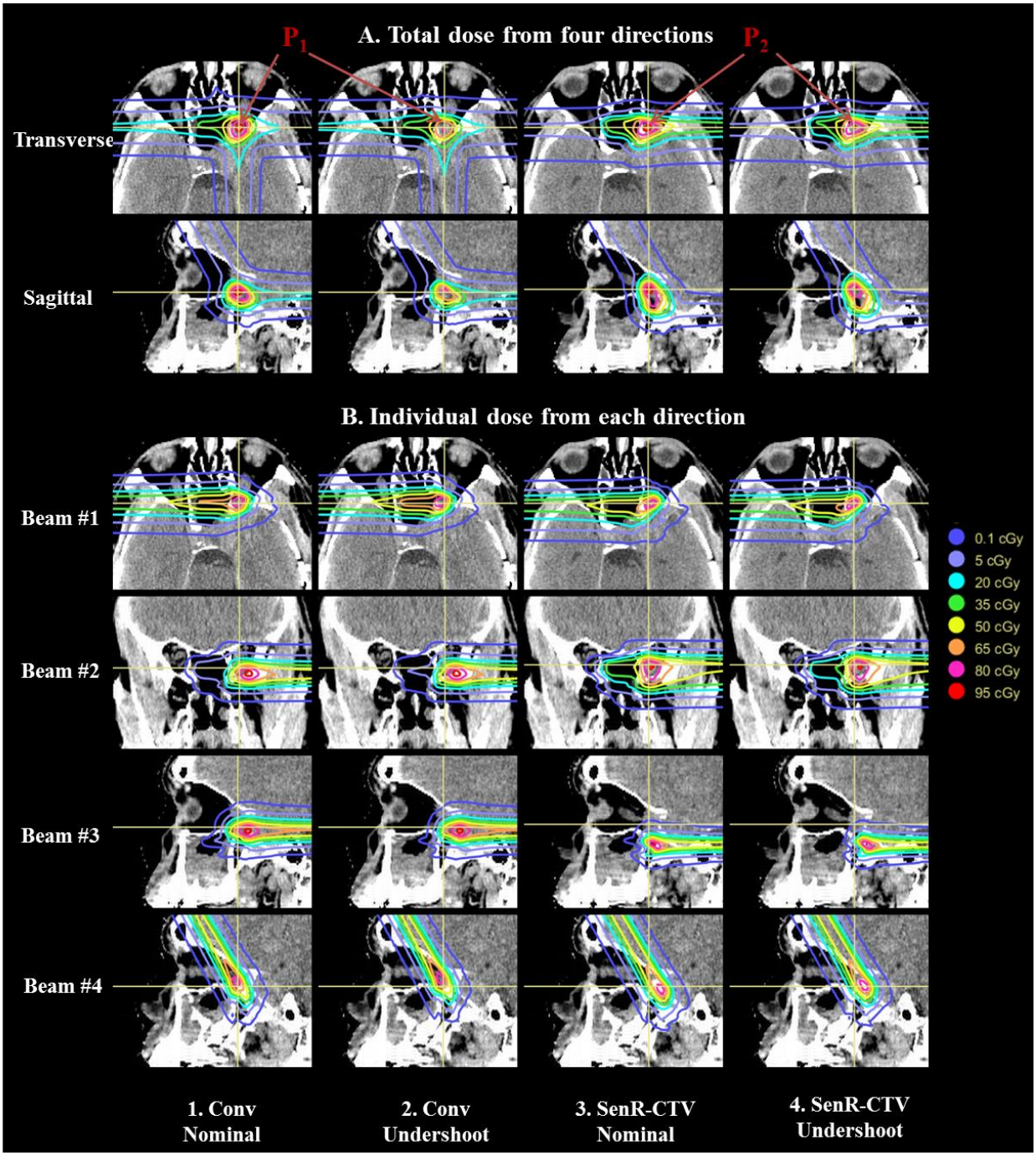

Figure 4-6: Spot-level analysis around a cold spot for the SBT #1 patient when range undershooting. (A) The total dose from the local scanning spots within 2 cm radius sphere of the cold spot. The first row is the transverse plane and the second row is the sagittal plane. (B) The dose contribution of the local spots from each beam direction. From left to right, each column represents the Conv nominal condition, Conv undershooting condition, SenR-CTV nominal condition and SenR-CTV undershooting condition.

4.1.4 Discussion

Current proton treatment planning methods manage robustness by performing optimization on a finite number of hypothetical worst cases. A drawback for the existing worst-case method is that it may be too conservative in certain cases, resulting in unacceptable dosimetric compromise[142] yet is inadequate for extreme case where the error exceeds expectation. The uncertainties are sparsely sampled in the worst-case approach, which is unprepared for positioning and range errors different from these sparsely-sampled cases. In comparison, in the SenR framework, robustness is included as a linear regularization term that not only softens the impact of robustness consideration, but also allows flexible adjustment of the robustness to meet varying requirements. In the nominal cases where the uncertainties are low, the dosimetric quality is better preserved. Due to the differences, SenR method may particularly benefit cases where the uncertainties are difficult to accurately estimate, highly heterogeneous in the same cohort, or variable over the treatment course. Since the sensitivity is calculated as a gradient of the spot dose distribution, our method does not depend on a specific set of expected positioning or range uncertainties, which is needed in the worst-case optimization. This difference lends the flexibility of trading off the robustness with dosimetry by adjusting the sensitivity term weighting without needing to estimate the uncertainties explicitly. This new robust optimization method is thus different from previous approaches of adjusting the worst case weights[132], using multi-criteria optimization[135], and using the normalized dose interval volume constraints[143].

Another drawback of worst-case methods is that they are computationally inefficient due to the time needed to optimize a significantly larger optimization problem

for all scenarios. The runtime of the SenR optimization is 22 times shorter than that of the voxel-wise worst-case method excluding pre-optimization calculation of the sensitivity matrix and worst-case doses, and 8 times shorter including pre-optimization calculation, while achieving comparable robustness in the hypothetical worst cases.

In this study, the SenR method is implemented on both CTV and PTV. The SenR-PTV method achieves comparable robustness towards the expected worst cases and OAR sparing as the WC method. Sen-CTV attains superior OAR sparing with a slight compromise in the CTV robustness while avoiding the substantial degradation seen in the conventional PTV plans. The different target volumes offer additional flexibility in clinical practice for the trade-off between OAR sparing and CTV coverage robustness. This is also feasible due to the demonstrated fast SenR planning speed. As an additional advantage, SenR is versatile and independent of the underlying proton dose calculation algorithms, of which, a model-based method and a Monte Carlo method were used showing consistent results.

The proposed method is particularly effective for targets in the heterogeneous environment where the sensitivity is captured in the perturbation term. The effectiveness of the regularized-sensitivity is highly dependent on the beam and spot arrangement. As shown by the spot level analysis, instead of matching the distal edges, SenR tends to combine distal, proximal and lateral edges of spots for improved robustness. Figure 4-6 shows one such possible robust combinations and the new optimization framework allows us to efficiently and globally find these combinations. The importance of combining spots for both plan robustness and conformality was discussed by Liu et al[49,144]. One of the main contributions here is to describe the intricate spot interdependence with a new mathematical model that can be efficiently solved.

The proposed method applies to scenarios where the same location is covered by multiple beams. However, field-matching may happen when different parts of the CTV are treated by different beams. The proposed method may result in a mismatch in the gradients at the field-matching lines that leads to cold and hot spots with position and range uncertainties. Further investigation is needed to understand and mitigate such dose heterogeneities.

4.1.5 Conclusion

We developed a novel computationally efficient robust optimization method for IMPT. The robustness is calculated as the spot sensitivity to both range and shift perturbations. The dose fidelity term is then regularized by the sensitivity term. The new SenR method offers the flexibility to balance between the dosimetry and the robustness. In the stress test, SenR is shown resilient to greater than expected uncertainties. The SenR FMO method is incorporated into the group sparsity based framework for robust BOO in the next section.

4.2 Robust Beam Orientation Optimization for Intensity-Modulated Proton Therapy[38]

4.2.1 Introduction

In general, both plan robustness and plan quality/conformality of IMPT depend on beam angle selection. An ideal IMPT treatment planning process should include beam angle selection and fluence map optimization (FMO) simultaneously. In current clinical practice, the proton beam angles are manually selected first by a planner. Different from X-ray

therapy where equiangular or arc beams are often acceptable, the proton beam orientations are typically asymmetric, and need to be more carefully considered for factors such as the water-equivalent thickness to the target, nearby OAR sparing, heterogeneity of tissues in the beam path, and setup robustness etc[12–14]. To minimize low dose regions and speed up treatment delivery, there are practically fewer beam angles in a typical proton plan, which makes the selection of proton beam angle particularly important. Planners' experience and skill can heavily influence the final treatment plan quality. For complicated patient cases, tedious trial-and-error attempts may be needed to find better beam configurations. Yet, human operators cannot effectively search the enormous coplanar and non-coplanar beam space, resulting in inconsistent planning results. Beam orientation optimization (BOO) using a computational model is, therefore essential for improving IMPT.

There have been limited IMPT BOO studies. Cao et al.[46] applied a local neighborhood search (LNS) algorithm to the IMPT BOO problem and implemented it on prostate cancer to improve beam arrangement[145]. The LNS is confined to be within a small search space near the initial condition, which still has to be manually selected. Later Lim et al.[146] used global search methods, such as branch-and-bound and simulated annealing, to find a good feasible solution as the initial condition for LNS, but these stochastic methods were only demonstrated on much smaller coplanar IMPT problems. In Chapter 2, we developed an integrated BOO and FMO framework for non-coplanar IMPT[35]. Based on group sparsity regularization, this algorithm efficiently performs a global search on non-coplanar candidate beams and finds a dosimetrically optimal solution.

Besides plan dosimetric quality, uncertainty or plan robustness is important for IMPT. The proton dose uncertainties sensitive to both patient positioning and range

estimation uncertainties[8–10,49,127] can lead to severely under-dosed target and over-dosed OARs. For IMPT, the commonly worst-case optimization method[21-40] improves the plan robustness at the cost of substantially increased computational cost[10,133]. To avoid the additional burden of calculating the worst cases and provide the robustness consideration as a soft constraint, we modeled the scanning spot sensitivity concerning range and positioning uncertainties as a regularization term in the optimization in Section 4.1[37]. We showed improved dosimetry, robustness to larger range uncertainties, and an order of magnitude faster optimization time than the worst case approach.

In the previous IMPT frameworks, robustness and BOO were studied separately, despite their obvious inter-dependence. For instance, beams passing through highly heterogeneous tissues are likely more sensitive to range and positioning uncertainties than beams passing through homogeneous tissues. It may cause more dosimetric compromise to achieve robustness for these beams. The robustness consideration complicates beam selection in manual IMPT planning, making integrated robust BOO and FMO even more urgently needed. Pflugfelder et al[127] modeled the interdependence of beam orientation and robustness as a lateral tissue heterogeneity across the proton pencil beams. Their *heterogeneity number*, is then used to guide beam angle selection[128,129]. After evaluating the heterogeneity of each beam, Bueno et al[128] recommended to change the beam direction if the heterogeneity exceeded a threshold, and Toramatsu et al[129] proposed to use the beams with minimum heterogeneity in single field uniform dose (SFUD) plans. These heuristic heterogeneity-guided beam angle selection methods have not quantitatively incoporated the robustness consideration in IMPT optimization and potentially dismiss dosimetrically superior beam orientations. Cao et al[46,145,146] combines the worst-case approach and local

neighborhood search algorithm to achieve robust beam angle selection. However, in addition to the limitations above being confined to the local search, in each search step, a subproblem of worst-case FMO is solved, making the method impractically slow.

In this work, we develop a novel unified robust optimization framework for IMPT, that integrates robust beam orientation selection and robust fluence map optimization in a single problem and then solve this global optimization problem. The BOO is achieved by group sparsity regularization, and robustness is promoted by the lateral tissue heterogeneity penalty and dose sensitivity regularization.

4.2.2 Methods

The integrated robust BOO and FMO framework is formulated with a dose fidelity term, a heterogeneity-weighted group sparsity term, and a dose sensitivity regularization term. The details are described as follows.

4.2.2.1 Heterogeneity-weighted group sparsity

The group sparsity-based BOO (GSBOO) framework presented in Equation 2-1 is designed to select beams for good dosimetry, and the robustness is not considered yet. In order to select beams with less sensitivity to setup uncertainties, we incorporate lateral tissue heterogeneity into the current group sparsity term, to encourage the algorithm to choose beams with less lateral tissue heterogeneity. The lateral tissue heterogeneity observed along beam b is quantified by its heterogeneity index h_b, which is defined as follows.

First, as shown in Figure 4-7, a coordinate system is created for each pencil beam (scanning spot) in beam b, with the z axis along the central axis of the pencil beam and pointing from the source to the patient. The central axis of ith pencil beam is located at (x_i, y_i), and the position of $(x_i, y_i, 0)$ is where the pencil beam enters the patient.

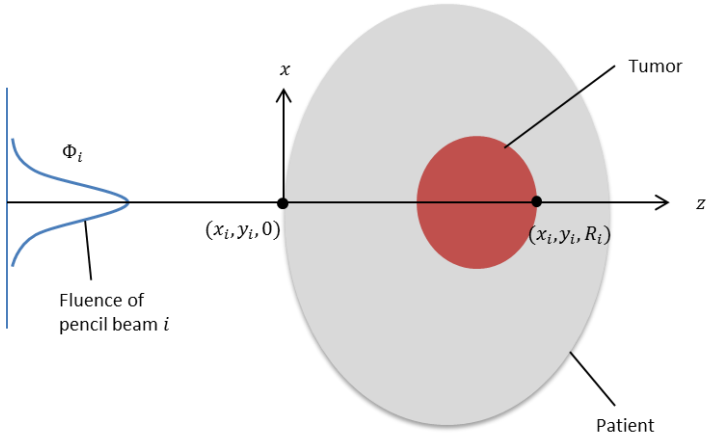

Figure 4-7: Diagram showing the coordinates used in the heterogeneity index calculation for a specific pencil beam.

With discrete sampling, the heterogeneity index of ith pencil beam in beam b at the depth z_k, denoted as $h_{b,i}^k$, is defined as:

$$h_{b,i}^k = \left(\frac{\sum_{j \in \mathcal{S}_i(z_k)} \phi_i(x_j, y_j, z_k) \cdot \left[S_{\text{rel}}(x_j, y_j, z_k) - S_{\text{rel}}(x_i, y_i, z_k) \right]^2}{\sum_{j \in \mathcal{S}_i(z_k)} \phi_i(x_j,\ y_j, z_k)} \right)^{1/2},$$

Equation 4-5

where $S_{\text{rel}}(x_j, y_j, z_k)$ is the relative stopping power ratio at the voxel (x_j, y_j, z_k), and $\phi_i(x_j,\ y_j, z_k)$ is the particle fluence at (x_j, y_j, z_k) for the ith pencil beam. The sampling set of lateral voxels at depth z_k is written as $\mathcal{S}_i(z_k)$. In the analytical model, the lateral dose distribution of pencil beam i is approximated as a single Gaussian distribution, with a standard deviation of $\sigma_i(z_k)$ at depth z_k. The sampling set $\mathcal{S}_i(z_k)$ at each depth is selected to include the voxels within $3\sigma_i(z_k)$ from the central axis.

The depth-specific $h_{b,i}^{k}$ is evaluated and summed up from $z_k = 0 \ldots R_{b,i}$, which is the path spanning from where the pencil beam enters the patient to the end of its range. The sum generates a single metric to indicate the lateral heterogeneity affecting the ith pencil beam in beam b:

$$h_{b,i} = \sum_{k=0}^{R_{b,i}} h_{b,i}^{k}.$$

Equation 4-6

The heterogeneity index values of all scanning spots in the same beam b are then averaged to represent the beam heterogeneity. Therefore, the heterogeneity index of beam b, denoted as h_b, is calculated as:

$$h_b = \frac{1}{n_b} \sum_{i=1}^{n_b} h_{b,i},$$

Equation 4-7

where n_b is the number of scanning spots in beam b.

Then h_b is evaluated for each candidate beam and used to weigh the group sparsity in Equation 2-1. The heterogeneity-weighted group sparsity BOO (HBOO) is thus formulated as:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \Gamma(A\boldsymbol{x}) + \sum_{b \in \mathcal{B}} \alpha_b h_b \|\boldsymbol{x}_b\|_2^{1/2}$$

$$\text{subject to} \quad \boldsymbol{x} \geq 0.$$

Equation 4-8

In this algorithm, the beams with higher lateral heterogeneity are more heavily penalized in the group sparsity term, resulting in selecting beams with higher dose fidelity and less sensitivity to setup errors.

4.2.2.2 Robust BOO-FMO

Even though the beams more resilient to setup errors are preferred in Equation 4-8, the range uncertainty has not been considered in FMO. Sensitivity regularization[37] is thus incorporated into Equation 4-8 to achieve simultaneous robust beam angle selection and robust fluence map optimization. The integrated robust BOO and FMO framework is written as:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \Gamma(A\boldsymbol{x}) + \sum_{b\in\mathcal{B}} \alpha_b h_b \|\boldsymbol{x}_b\|_2^{1/2} + \sum_{k\in\{\boldsymbol{u},\boldsymbol{v}\}} \lambda_k \boldsymbol{s}_k^T \boldsymbol{x}$$

$$\text{subject to} \quad \boldsymbol{x} \geq 0,$$

Equation 4-9

This Sensitivity regularization and Heterogeneity weighting based BOO and FMO framework (SHBOO-FMO), allows robust beams to be selected and robust fluence map to be generated in a single equation. SHBOO will be used in place of SHBOO-FMO for the rest of the section when referring to the BOO algorithm and the selected beams for brevity.

4.2.2.3 Evaluations

This SHBOO-FMO method was tested on two patients with the skull base tumor (SBT) and two bilateral head-and-neck (H&N) patients. Four beams were selected for the SBT patients and three beams for the H&N patients. For each patient, there were about 700 to 800 candidate beams. For each candidate beam, dose calculation for the scanning spots

covering the CTV and a 5 mm margin was performed by matRad[51,52]. In matRad, the lateral beam width is calculated as the root sum square of the initial beam width from Safai et al[153] and the lateral broadening from Gottschalk et al[154]. The CTV was set as the optimization target. The prescription dose, target volume, and average spot count per beam for each patient are shown in Table 4-5.

Case		Prescription Dose (GyRBE)	CTV Volume (cc)	Average Spots per Beam
SBT #1		56	33.7	2537
SBT #2		70	36.8	2650
H&N #1	CTV54	54	108.0	10077
	CTV60	60	127.3	
H&N #2	CTV54	54	110.4	9433
	CTV60	60	99.0	
	CTV63	63	10.2	

Table 4-5: Prescription doses, CTV volumes and average number of spots per beam for each patient.

The dosimetry and plan robustness of the proposed SHBOO-FMO method was compared against 1) the voxel-wise worst-case FMO method with manually selected beams (MAN-WC), and 2) sensitivity-regularized FMO method with the same manual beams (MAN-SenR). The voxel-wise worst-case optimization considered nine scenarios, including one nominal scenario and the 8 worst-case scenarios same as those in Section 4.1.2.3.

In addition to the robustness of the final plan, the sole robustness of the selected beams by SHBOO-FMO, was also evaluated and compared with the following beam sets: 1) manually selected beams, 2) GSBOO beams selected by Equation 2-1, and 3) HBOO beams selected by Equation 4-8. The comparison was performed by creating plans using the same conventional CTV-based FMO method (Conv), using the aforementioned beam sets. Same candidate beam set, spot population, and dose calculation scheme were used for different

BOO algorithms. The acronym used for each method and its definition can be found in Table 4-6.

Acronym	Definition
SHBOO-FMO	Group sparsity based integrated BOO and FMO framework with sensitivity regularization and heterogeneity weighting
SHBOO	Short for SHBOO-FMO when referring to the BOO algorithm and the beams selected by SHBOO-FMO
MAN	Manually selected beams
GSBOO	Group sparsity based BOO algorithm
HBOO	Heterogeneity-weighted group sparsity BOO algorithm
Conv	Conventional CTV-based FMO method
MAN-Conv GSBOO-Conv HBOO-Conv SHBOO-Conv	Conventional CTV-based FMO plan with MAN, GSBOO, HBOO, and SHBOO beams, respectively
MAN-WC	CTV-based voxel-wise worst-case FMO method with manually selected beams
MAN-SenR	Sensitivity-regularized FMO method with manually selected beams

Table 4-6: Acronym of each method and its definition.

4.2.3 Results

4.2.3.1 Runtime and selected beams

The dose, sensitivity and heterogeneity calculation for all the candidate beams were performed on a Xeon 20-core CPU server operating at 3.10 GHz clock, with Matlab and its Parallel Computing Toolbox. The averaged time per beam to calculate the three data is listed in Table 4-7. The most time-consuming step during preparation is the evaluation of the sensitivity vector. The averaged runtime for GSBOO, HBOO, and SHBOO, on an i7 CPU

desktop, is also shown in Table 4-7. Depending on the target size, these BOO process took about 6-75 minutes to complete. With the additional heterogeneity weighting and sensitivity regularization, the SHBOO method reduced the runtime from the original GSBOO method approximately by half.

Case	Calculation time per beam (s)			BOO runtime (s)		
	Dose	Sensitivity	Heterogeneity	GSBOO	HBOO	SHBOO
SBT #1	0.4	1.5	1.5	804	745	362
SBT #2	0.6	2.0	1.6	1102	999	682
H&N #1	1.9	24.0	8.2	3214	2978	1446
H&N #2	1.4	14.9	7.2	4407	3996	2728

Table 4-7: Preparation time and runtime of each BOO method for the tested patients.

4.2.3.2 Beam robustness

The beam robustness was compared among the plans using different BOO methods but the same conventional CTV-based approach (Conv) for fluence map optimization.

Figure 4-8 shows the DVH bands of the CTVs of these Conv plans with range uncertainty and setup uncertainty for the SBT patients and H&N patients. In these DVH band plots, the solid lines are the nominal DVHs without uncertainties, the dotted lines and bands bound the worst-case dose distributions, and the horizontal and vertical lines label the worst D95% of each method for each CTV. For the tested cases, the beam robustness of the GSBOO method is not maintained. For example, the GSBOO beams lead to wide DVH bands under range uncertainties for the SBT #2 and H&N #2 patients, and wide bands under setup uncertainties for the two SBT patients. With heterogeneity-weighted group sparsity, the beam robustness against setup uncertainty is improved from the GSBOO beams for the four tested patients, while the robustness against range uncertainty varies

among patients. With the SHBOO method, the beams with superior range robustness over manual beams and HBOO beams are selected while the setup robustness is maintained or improved.

The D98%, D95%, V95% and V100% of each CTV with range uncertainties and setup uncertainties were evaluated and plotted in Figure 4-9. Compared with the manual selection, the lowest (worst) D98%, D95%, V95%, and V100% were improved by the SHBOO method. On average, the lowest [D98%, D95%, V95%, V100%] of CTV increased from [90.9%, 93.9%, 91.1%, 70.6%] in MAN beams, to [96.0%, 98.6%, 98.6%, 96.1%] in SHBOO beams. Under setup uncertainties, the average lowest [D98%, D95%, V95%, V100%] of CTV increased from [92.1%, 94.8%, 94.3%, 78.9%] in MAN beams, to [93.5%, 96.6%, 97.0%, 92.0%] in SHBOO beams.

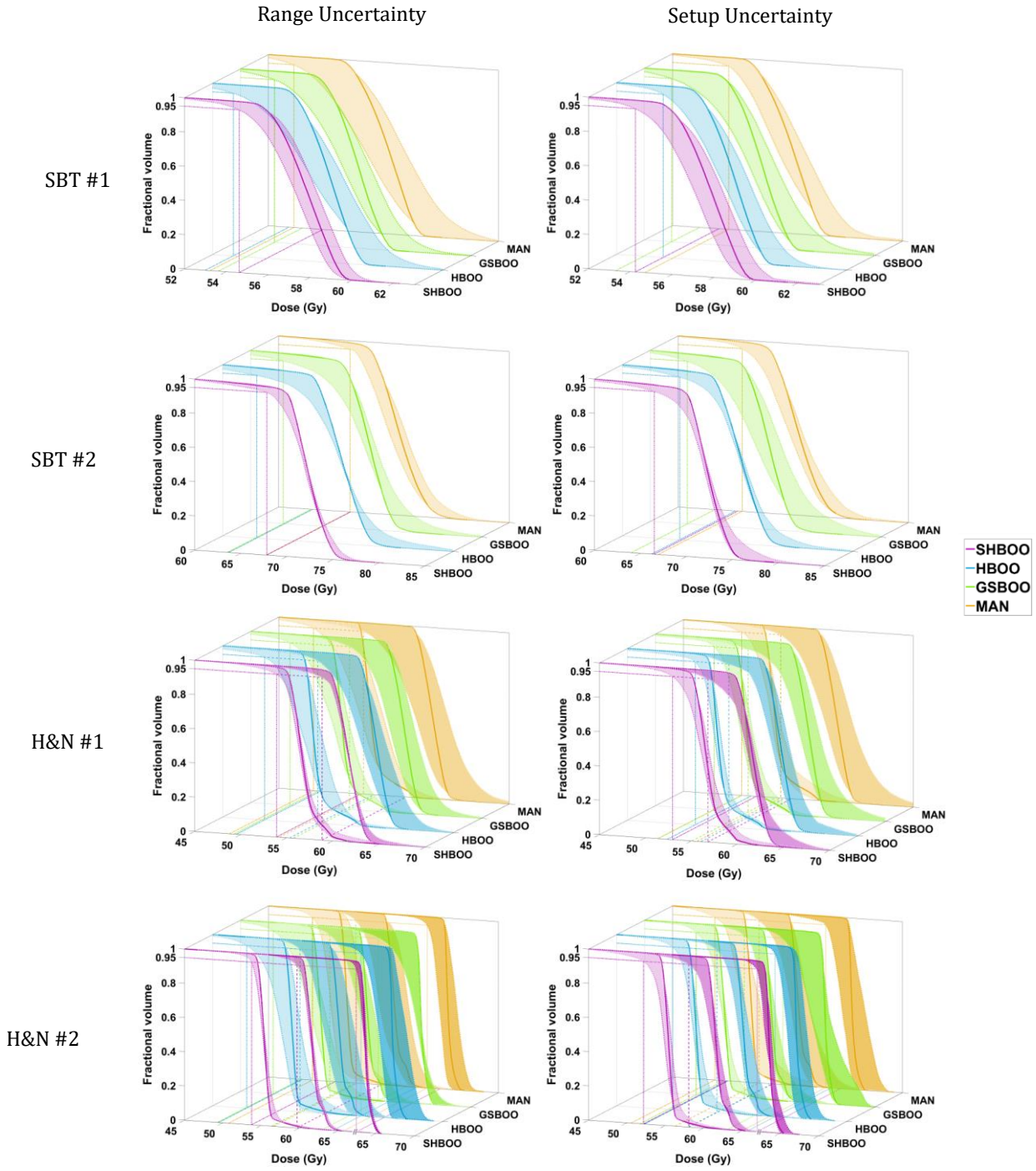

Figure 4-8: CTV DVH bands of the four patients, indicating the robustness of the beams chosen by different methods. The situation with only range uncertainty is shown on the left and situation with only setup uncertainty is shown on the right. The worst D95% of each method is labeled by reference lines in the x-y plane. The two CTVs in the H&N #1 patient are plotted together in the third row, and the three CTVs in the H&N #2 patient are plotted together in the fourth row, with different transparencies.

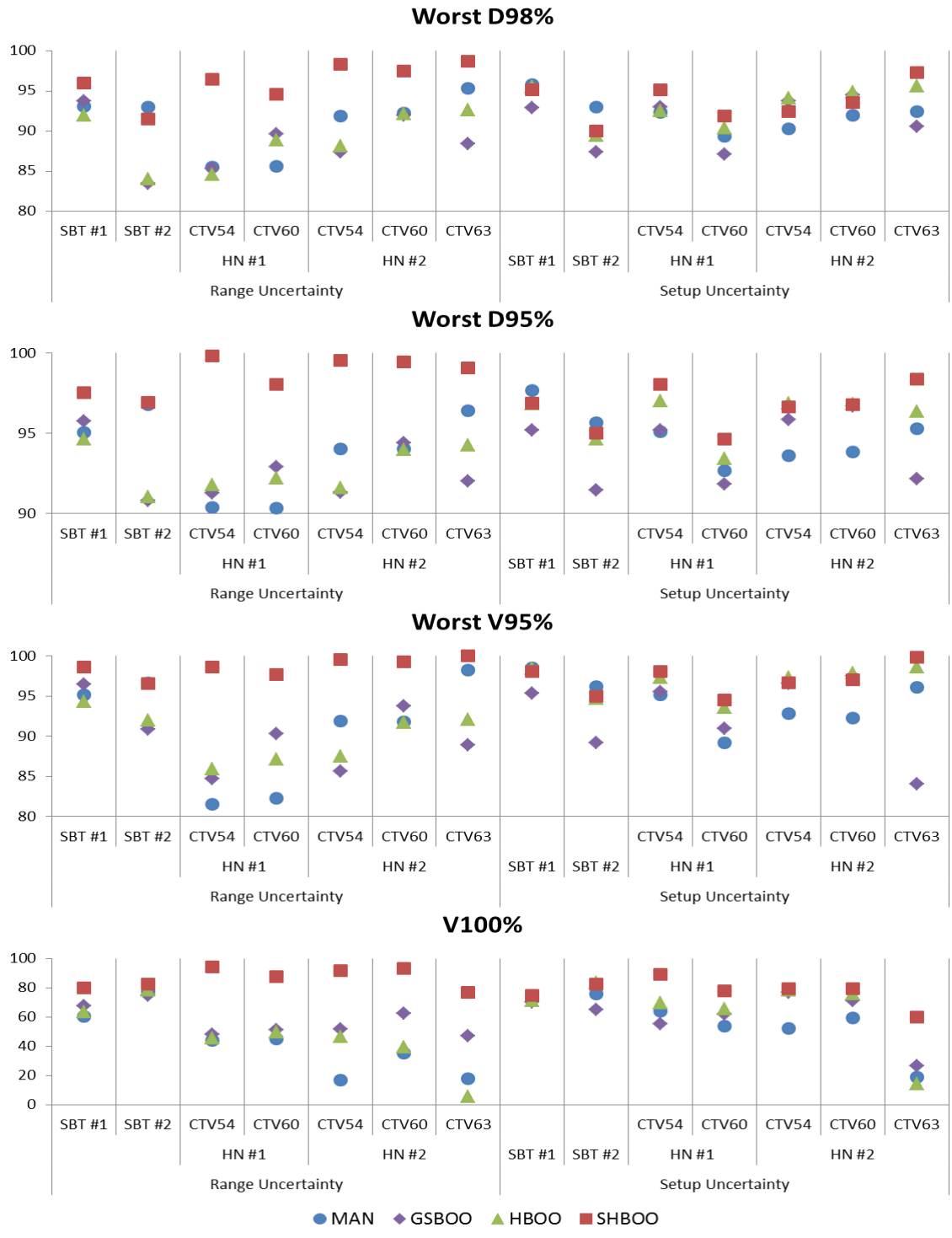

Figure 4-9: The comparison of worst D98% (top row), D95% (second row), V95% (third row), and V100% (bottom row) of the CTVs as a percentage of prescription doses, for every patient, between the plans with Conv FMO and MAN, GSBOO, HBOO and SHBOO beams, respectively. The situation with only range uncertainty is shown on the left and situation with only setup uncertainty is shown on the right in each plot.

4.2.3.3 Plan robustness

The plan robustness of SHBOO-FMO method was compared with the plan with manual beams and voxel-wise worst-case FMO (MAN-WC) as well as that with manual beams and SenR FMO (MAN-SenR). The CTV DVH bands of the three methods are shown in Figure 4-10 for the SBT patients and the H&N patients. Under range uncertainties, narrower DVH bands were observed in the SHBOO-FMO plans compared with the MAN-WC plans, and the CTV underdosage in the MAN-SenR plans was also improved by the SHBOO-FMO method. Under setup uncertainties, the SHBOO-FMO method was less robust than MAN-WC but comparable with or more robust than MAN-SenR.

The lowest (worst) D98%, D95%, V95% and V100% of each CTV with range uncertainties and setup uncertainties were also evaluated and plotted in Figure 4-11. Compared with MAN-SenR, the D98%, D95%, V95% and V100% were improved by the SHBOO-FMO method. On average, the lowest [D98%, D95%, V95%, V100%] of CTV were increased from [94.0%, 97.4%, 97.6%, 94.6%] in MAN-SenR plans, to [96.2%, 98.8%, 98.7%, 96.7%] in SHBOO-FMO plans. Under setup uncertainties, the averaged lowest [D98%, D95%, V95%, V100%] of CTV were increased from [93.1%, 96.5%, 96.9%, 92.0%] in MAN-SenR plans, to [93.8%, 96.9%, 97.3%, 93.0%] in SHBOO-FMO plans. Overall the MAN-WC method achieved the best CTV metrics, with the averaged lowest [D98%, D95%, V95%, V100%] of [97.5%, 98.8%, 99.4%, 97.4%] under range uncertainties and [97.9%, 99.1%, 99.6%, 97.9%] under setup uncertainties.

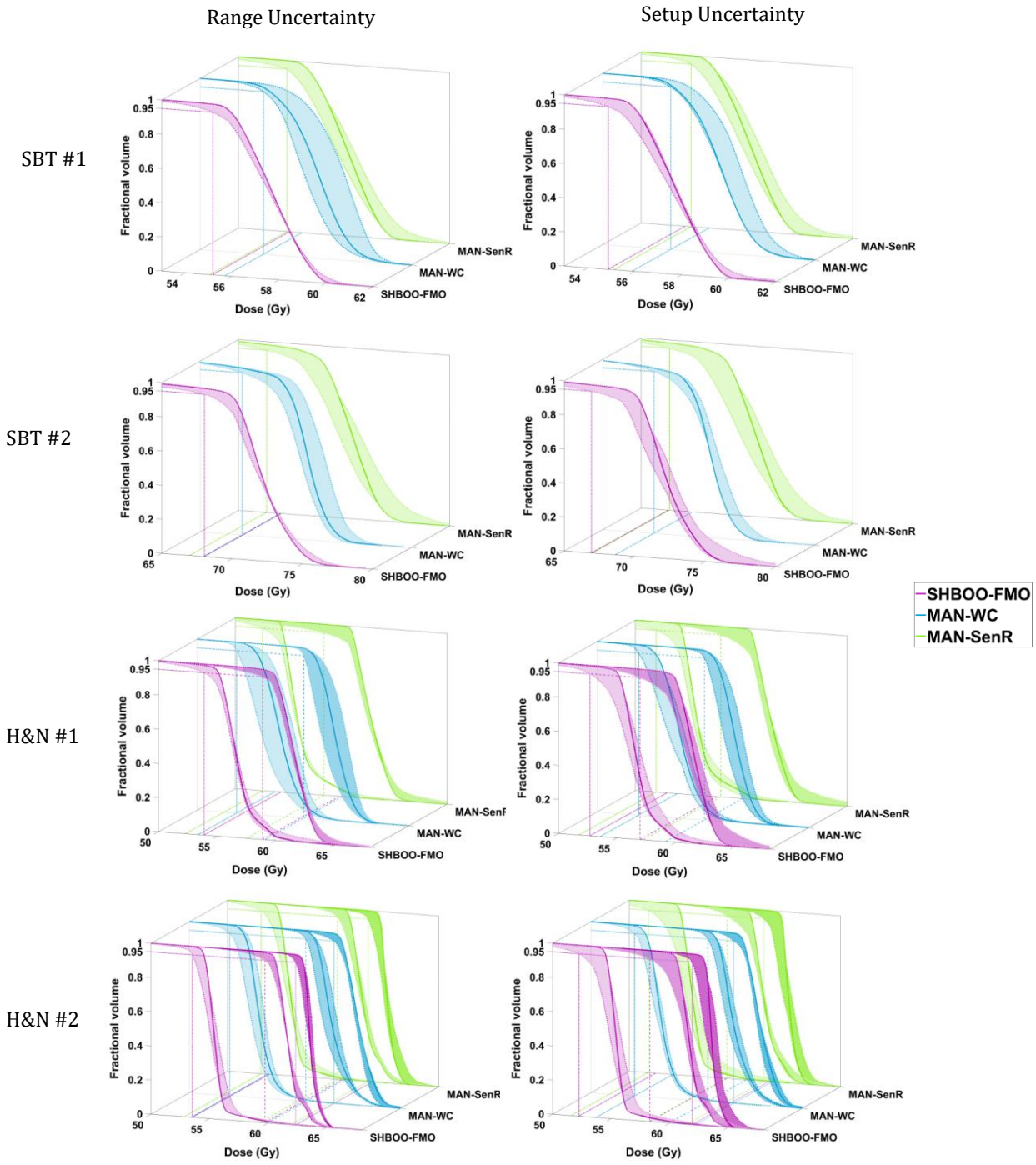

Figure 4-10: CTV DVH bands of the four patients, indicating the robustness of the plans generated by SHBOO-FMO, MAN-WC and MAN-SenR. Situation with only range uncertainty is shown on the left and situation with only setup uncertainty is shown on the right. The two CTVs in the H&N #1 patient are plotted together in the third row, and the three CTVs in the H&N #2 patient are plotted together in two figures in the fourth row. The worst D95% of each method is labeled by reference lines in the x-y plane.

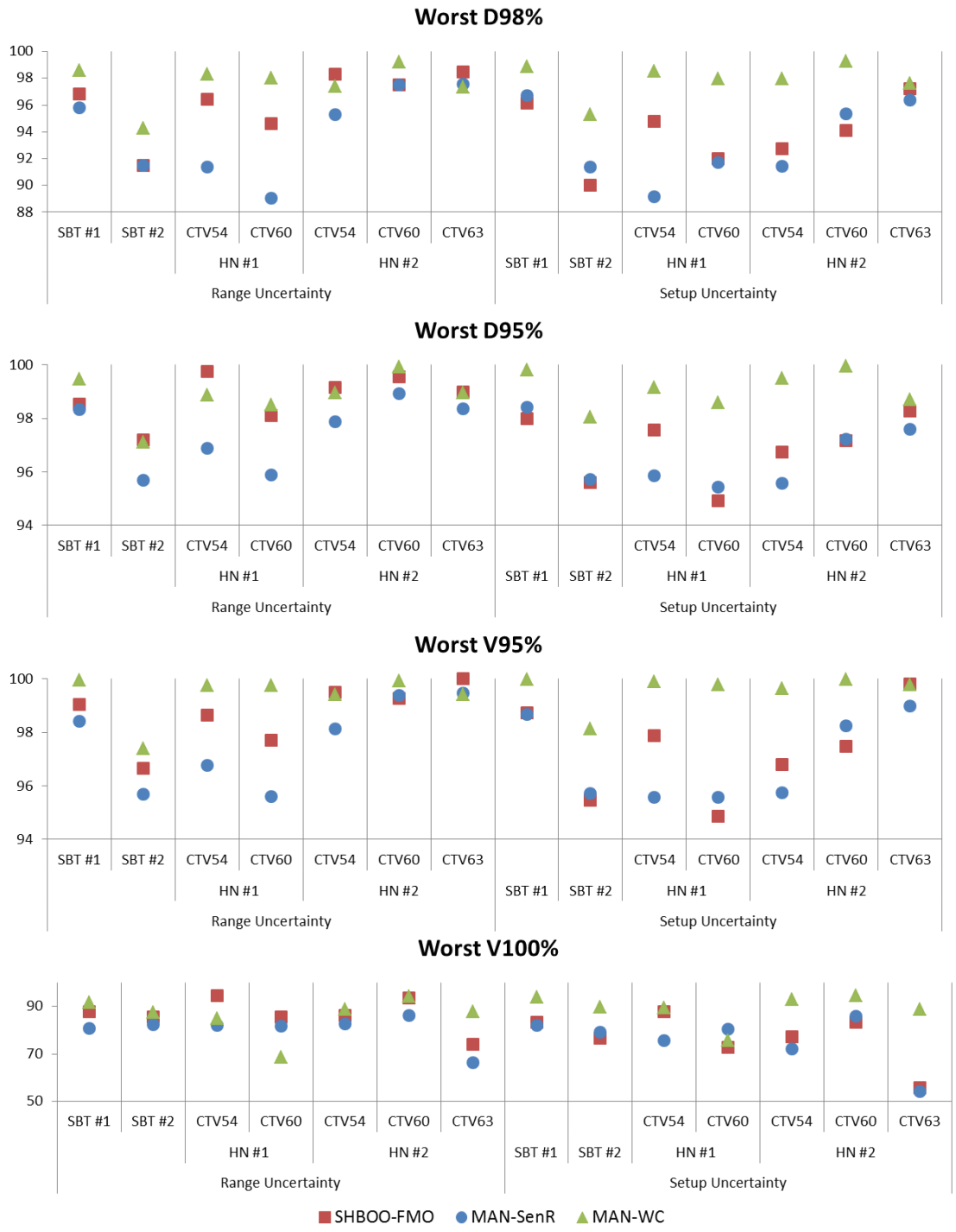

Figure 4-11: The comparison of worst D98% (top row), D95% (second row), V95% (third row), and V100% (bottom row) of the CTVs as a percentage of prescription doses, for every patient, between the MAN-WC plan, MAN-SenR plan and SHBOO-FMO plan Situation with only range uncertainty is shown on the left and situation with only setup uncertainty is shown on the right.

4.2.3.4 Nominal dose comparison

Several OARs are selected for the SBT and H&N sites, respectively, and the differences of their mean and maximum doses between the SHBOO plans and the MAN plans are presented in Table 4-8, and Table 4-9, respectively. Figure 4-12 shows the nominal DVHs comparison between the SHBOO-FMO method and MAN-WC method for the four tested patients.

The SHBOO-FMO plans achieved substantially better OAR sparing compared with the MAN-WC plans. For example, in the SBT cases, the dose sparing of all the OARs was improved. In the SBT #2 patient, the SHBOO-FMO plan reduced the max dose to the right optical nerve and left eye by 13.9 GyRBE and 25.6 GyRBE from the MAN-WC plan. In the H&N cases, the overall OAR sparing was also improved by SHBOO-FMO method from MAN-WC method, except for the increase of mean dose to the right submandibular gland. The average reduction of [Dmean, Dmax] of the SHBOO-FMO plans from the MAN-WC plans were [6.3, 6.6] GyRBE for the SBT cases and [1.9, 5.1] GyRBE for the H&N cases.

From Table 4-8 and Table 4-9, the overall OAR sparing of SHBOO-FMO was better than MAN-SenR in the SBT cases and comparable with the MAN-SenR in the H&N cases. The average reduction of [Dmean, Dmax] of the SHBOO-FMO plans from the MAN-SenR plans were of [2.1, 2.4] GyRBE for the SBT cases, and [-0.4, 2.5] GyRBE for the H&N cases.

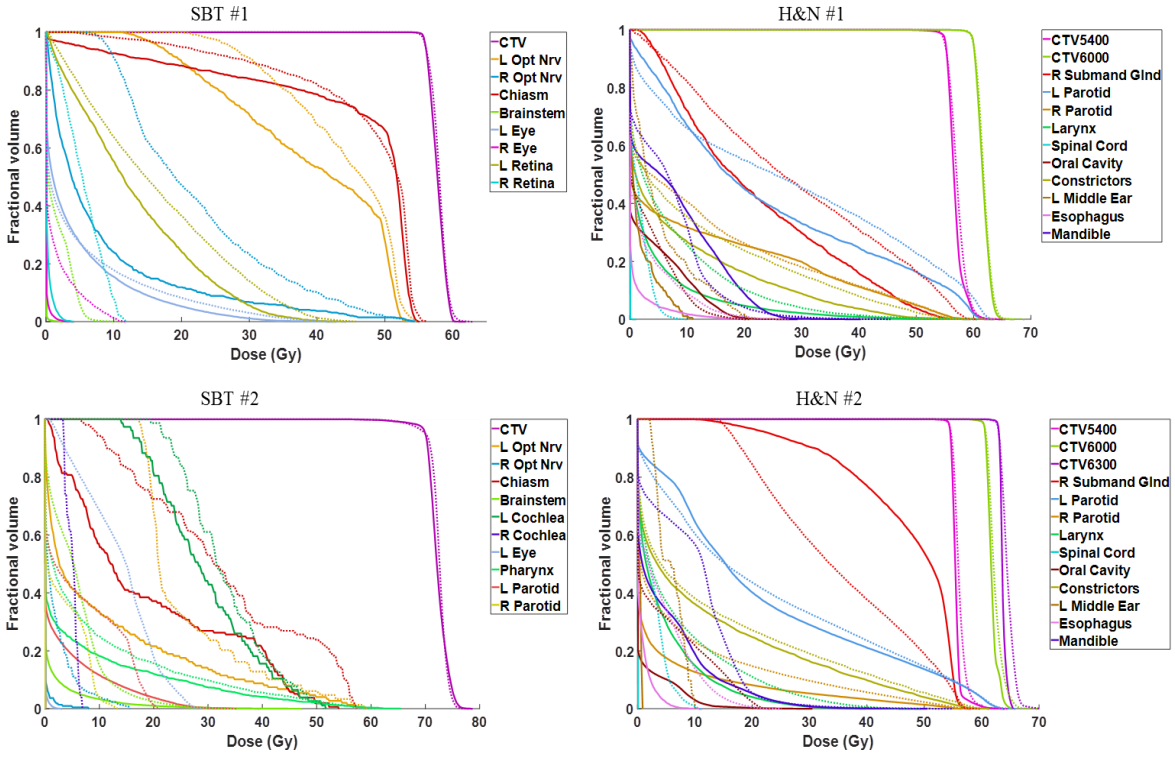

Figure 4-12: Comparison of nominal DVHs for four patients between the SHBOO-FMO method (solid) and MAN-WC method (dotted).

SBT Case	SHBOO-FMO – MAN-WC (GyRBE)				SHBOO-FMO – MAN-SenR (GyRBE)			
	Dmean		Dmax		Dmean		Dmax	
	#1	#2	#1	#2	#1	#2	#1	#2
L Opt Nrv	-5.1	-15.1	-1.4	-0.9	+0.1	-10.1	+0.8	+1.4
R Opt Nrv	-13.9	-1.8	-4.6	-11.2	-2.1	-0.2	-2.0	-4.6
Chiasm	-2.4	-13.7	-0.8	-7.2	-2.1	-5.1	+0.2	-4.2
Brainstem	-1.5	-4.6	-5.8	-7.3	-0.4	-0.4	-3.3	2.3
L Eye	-0.6	-13.7	-5.4	-25.6	+0.5	-6.0	-2.2	-16.7
R Eye	-1.2	0.0	-8.4	0.0	-0.1	0.0	-0.7	0.0
L Cochlea	0.0	-3.6	0.0	0.0	0.0	0.8	0.0	+1.2
R Cochlea	0.0	-4.9	0.0	-6.7	0.0	-2.1	0.0	-3.5

Table 4-8: OAR mean dose and max dose reduction of the SHBOO-FMO plans from the MAN-WC plans and MAN-SenR, for the SBT cases under nominal situation.

H&N Case	SHBOO-FMO – MAN-WC(GyRBE)				SHBOO-FMO – MAN-SenR(GyRBE)			
	Dmean		Dmax		Dmean		Dmax	
	#1	#2	#1	#2	#1	#2	#1	#2
R Submandibular Gland	-6.8	+11.5	-5.3	-0.3	+9.2	+10.3	-1.6	+0.8
L Parotid	-3.8	-0.4	-1.9	-0.3	+5.7	+0.9	+0.3	0.1
R Parotid	-1.5	-3.3	0.0	-4.3	-0.8	-3.2	+0.2	-4.0
Larynx	-3.2	-2.3	-5.4	-9.6	-1.3	0.0	-0.5	-3.9
Spinal Cord	-1.6	-2.3	-5.9	-8.7	-0.5	-0.3	-3.6	-2.8
BrainStem	-1.8	-0.4	-13.6	-4.1	-0.7	-0.1	-8.8	-0.8
Oral Cavity	+0.1	-3.3	+2.9	-9.2	+1.9	-0.4	+5.5	+1.5
Constrictors	-3.4	-1.2	-6.3	-2.2	-0.2	+1.0	-3.2	0.0
L Middle Ear	-3.4	-5.3	-10.1	-9.1	-1.8	-10.5	-4.8	-16.5
Esophagus	-1.8	-3.4	-6.8	-11.9	-0.8	-1.3	-2.0	-7.7
Mandible	+0.7	-4.9	0.2	-0.1	+4.0	-3.5	+3.0	-6.0

Table 4-9: OAR mean dose and max dose reduction of the SHBOO-FMO plans from the MAN-WC plans and MAN-SenR, for the H&N cases under the nominal situation.

4.2.4 Discussion

To the best of our knowledge, this work describes the first integrated IMPT optimization method that optimizes beam orientation and scanning-spot intensities for both nominal dose conformality and robustness. In clinical IMPT planning, it is known that the beam orientation directly impacts the IMPT dose conformality and robustness, requiring substantial manual effort from the dosimetrists to find better beam angles. However, a manual search is ineffective in identifying beams from the enormous non-coplanar space for *both* dosimetry and robustness goals. The combination of group sparsity,

lateral heterogeneity, and sensitivity into a formulation that allows global search on all feasible candidate beams is a major contribution of this study.

Proton beam has a unique feature that protons stop at the end of its Bragg Peak. This is different from the photon beam. As a result, the experience in beam angle selection is different from the photon experience, in particular for non-coplanar beams. The results on tested patients show that the proposed robust BOO algorithm selects beams that are more resilient to range and setup uncertainties. The final SHBOO-FMO plans better spared the OAR sparing compared with the voxel-wise worst-case method on the manual beams while maintaining similar robustness. Compared with the plans using manually selected beams and SenR FMO, the proposed method achieved better target coverage under simulated uncertainties.

Furthermore, the sensitivity regularization term helps to directly generate the fluence map which is more robust to range and setup uncertainties. In the limited existing IMPT BOO studies, the FMO is a nested subproblem that is solved post-hoc, which not only is inefficient but also compromises plan optimality as the FMO results could influence the selected beams. Our algorithm integrates FMO and BOO in a single function to ensure that both the beam orientations and the spot intensities are matched for the desired dosimetry and robustness. The second important aspect of our study is that rather than the commonly used worst-case scenario optimization method for FMO, we apply sensitivity regularization to improve the plan robustness against errors. This non-scenario-based method can be easily and efficiently incorporated into the optimization framework and provides the flexibility between the dosimetry and the robustness. Our previous study[37] showed that the sensitivity regularization is more effective to mitigate range uncertainties than setup

uncertainties. The latter weakness is largely remedied in the current framework by incorporating lateral tissue heterogeneity in the BOO.

Compared with GSBOO[35] or SenR[37] alone, the planning time of SHBOO-FMO is longer. The computational cost of the proposed method attributes to two main components: the pre-optimization calculation and optimization of the objective function. Preoptimization includes calculating a dose calculation matrix, heterogeneity index, and sensitivity vector for each candidate beam. Under the analytical calculation model, the dose calculation and heterogeneity evaluation, in theory, could have shared the same ray tracing step to reduce the calculation time shown in Table III. Calculation of the sensitivity is more time-consuming. However, this parallel calculation process can be accelerated using the modern graphics processing unit (GPU) platform. Further acceleration is expected using a non-uniform sampling of the dose matrix to have a higher resolution in the CTV and its vicinity and lower resolution elsewhere.

For the optimization step, the Equation 4-9 itself is a large-scale problem due to the extra freedom of proton energy in IMPT and a large number of non-coplanar candidate beams used in this study. With the linear formulation of the sensitivity regularization term and the proximal operators derived in Section 2[35], we are able to efficiently solve the problem with FISTA, which converges at a rate of $O(1/k^2)$ amongst the first-order methods[56]. Moreover, by adding the sensitivity regularization term, the time spent on beam pruning within the SHBOO method is reduced to approximately half of the initial group-sparsity based BOO method.

It is necessary to clarify that the study only handles range uncertainties and setup errors from interfractional setup variations. Other sources of uncertainties such as intrafractional respiratory motion and anatomy changes, which heavily affect the beam selection process, require separate approaches to tackle. Biological effect is another important factor to consider in BOO. In our future work, linear energy transfer (LET) will be included in this framework to encourage selecting beams with a higher biological effect on the target and lower biological risk on the OARs.

4.2.5 Conclusion

We developed a novel IMPT robust optimization method, which efficiently solved robust BOO and FMO in a unified framework, generating plans with superior dosimetry and good robustness.

5 Appendix

THE FISTA ALGORITHM

The Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)[56], an accelerated proximal gradient algorithm, has been successfully utilized to solve all of the problems presented in this dissertation. In this appendix, the format of the optimization problem required for this algorithm, and line-search procedure, will be described.

5.1 Optimization Problem Formulation

The FISTA algorithm solves problems in the following canonical form

$$minimize\ f(\boldsymbol{x}) + g(\boldsymbol{x}),$$

Equation 5-1

where f is a smooth convex function, which is continuously differentiable with Lipschitz continuous gradient (∇f); g is a function which is possibly nonsmooth, but has a proximal operator that can be evaluated efficiently. The proximal operator with step size $t > 0$ for function g is defined by

$$\text{prox}_{tg}(\boldsymbol{x}) = \underset{\boldsymbol{y}}{\text{argmin}}\ \ g(\boldsymbol{y}) + \frac{1}{2t}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2.$$

Equation 5-2

To solve an optimization problem with FISTA, the optimization must be rewritten to fit this canonical form. All problems in this dissertation can easily be expressed in this canonical form.

5.2 The Line-Search Algorithm

The implementation of line-search FISTA algorithm is shown below.

FISTA with line-search

Initialize $\boldsymbol{x_0 := 0, v_0 := x_0}$ **,** $\boldsymbol{t_0 > 0}$ **,** $\boldsymbol{0 < r < 1}$

for $\boldsymbol{k = 1, 2, \dots, n}$, do

$\boldsymbol{t := t_{k-1}/r}$

Repeat

$$\boldsymbol{\theta := \begin{cases} 1 & \text{if } k = 1 \\ \text{positive root of } t_{k-1}\theta^2 = t\theta_{k-1}^2(1-\theta) & \text{if } k > 1 \end{cases}}$$

$\boldsymbol{y := (1-\theta)x_{k-1} + \theta v_{k-1}}$

$\boldsymbol{x := \mathbf{prox}_{tg}\big(y - t\nabla f(y)\big)}$

break if $\boldsymbol{f(x) \le f(y) + < \nabla f(y), x - y > + \frac{1}{2t}\|x - y\|_2^2}$

$\boldsymbol{t := rt}$

$\boldsymbol{t_k := t}$

$\boldsymbol{\theta_k := \theta}$

$\boldsymbol{x_k := x}$

$\boldsymbol{v_k := x_k + \frac{1}{\theta_k}(x_k - x_{k-1})}$

end

return $\boldsymbol{x}$

Table 5-1: Pseudo code for FISTA with line search.

5.3 Solving Group Sparsity based BOO Problem

The group sparsity based problems are re-written to fit the canonical form in Equation 5-1. The process of re-writing Equation 2-1 is shown in this section as an example. Solving the rest of group-sparsity based problems follows the similar rule. The objective function in Equation 2-1 can be re-written in the following format

$$f(\boldsymbol{x}) = \Gamma(A\boldsymbol{x}),$$

$$g(\boldsymbol{x}) = \sum_{b \in B} \alpha_b \|\boldsymbol{x}_b\|_2^p + I_{\ge 0}(\boldsymbol{x}),$$

Equation 5-3

where $I_{\geq 0}(\boldsymbol{x})$ is an indicator function, defined as

$$I_{\geq 0}(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} \geq 0 \\ \infty & \text{otherwise} \end{cases}$$

Equation 5-4

The gradient of $f(\boldsymbol{x})$ is calculated as

$$\nabla f(\boldsymbol{x}) = A^{\mathrm{T}} \nabla \Gamma(A\boldsymbol{x})$$

Equation 5-5

After obtaining the gradient of the function f, the next step is to derive a formula for the proximal operator of the function g. In the BOO problem, $g(\boldsymbol{x})$ is a separable sum: $g(\boldsymbol{x}) = \sum_{b=1}^{B} g_b(\boldsymbol{x}_b)$, where

$$g_b(\boldsymbol{x}_b) = \sum_{b \in B} \alpha_b \|\boldsymbol{x}_b\|_2^p + I_{\geq 0}(\boldsymbol{x}_b)$$

Equation 5-6

It follows from the separable sum rule for proximal operators that the problem evaluating the proximal operator of $g(\boldsymbol{x})$ reduces to independently evaluating the proximal operators of the functions $g_b(\boldsymbol{x_b})$. To simplify notation, we derive an expression for the proximal operator of the function $h(\boldsymbol{x}) = \alpha\|\boldsymbol{x}\|_2^p + I_{\geq 0}(\boldsymbol{x})$. Evaluating the proximal operator of h requires solving the optimization problem:

$$\underset{\boldsymbol{x}}{\text{minimize}} \ \alpha\|\boldsymbol{x}\|_2^p + \frac{1}{2t}\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2$$

$$\text{subject to } \boldsymbol{x} \geq 0$$

Equation 5-7

The proximal operator of function h is:

$$\text{prox}_{th}(\boldsymbol{x}) = \text{prox}_{\alpha t\|\cdot\|_2^p}\big(\max(\boldsymbol{x}, 0)\big)$$

Equation 5-8

There is a known form of proximal operator for both L2,1-norm and L2,1/2-norm[155]:

$$\text{prox}_{t\|\cdot\|_2}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{x} \cdot \min\left(\frac{t}{\|\boldsymbol{x}\|_2}, 1\right)$$

$$\text{prox}_{t\|\cdot\|_2^{\frac{1}{2}}}(\boldsymbol{x}) = \begin{cases} 0, & \text{if } t\|\boldsymbol{x}\|_2^{-1.5} > \dfrac{2\sqrt{6}}{9} \\ \boldsymbol{x}\sqrt{\dfrac{2}{\sqrt{3}}\sin\left(\dfrac{1}{3}\left(\arccos\left(\dfrac{3\sqrt{3}}{4}t\|\boldsymbol{x}\|_2^{-1.5}\right) + \dfrac{\pi}{2}\right)\right)}, & \text{otherwise} \end{cases}$$

Equation 5-9

By deriving the gradient of function f and proximal operator of function g, the Equation 2-1 is then readily solved using FISTA.

6 REFERENCES

1. Khan FM, Gibbons JP. *Khan's the Physics of Radiation Therapy*. Lippincott Williams & Wilkins; 2014.

2. Wilson RR. Radiological use of fast protons. *Radiology*. 1946;47(July):487-491. doi:10.1148/47.5.487

3. Hall EJ. Intensity-modulated radiation therapy, protons, and the risk of second cancers. *Int J Radiat Oncol Biol Phys*. 2006;65(1):1-7. doi:10.1016/j.ijrobp.2006.01.027

4. Kanai T, Kawachi K, Kumamoto Y, et al. Spot scanning system for proton radiotherapy. *Med Phys*. 1980;7(1980):365-369. doi:10.1118/1.594693

5. Lomax a J. Intensity modulation methods for proton radiotherapy. *Phys Med Biol*. 1999;44:185–205. doi:10.1088/0031-9155/44/1/014

6. Lomax AJ, Böhringer T, Bolsi A, et al. Treatment planning and verification of proton

therapy using spot scanning: initial experiences. *Med Phys*. 2004;31(11):3150-3157. doi:10.1118/1.1779371

7. Albertini F, Bolsi A, Lomax AJ, Rutz HP, Timmerman B, Goitein G. Sensitivity of intensity modulated proton therapy plans to changes in patient weight. *Radiother Oncol*. 2008;86(2):187-194. doi:10.1016/j.radonc.2007.11.032

8. Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: The potential effects of calculational uncertainties. *Phys Med Biol*. 2008;53(4):1027-1042. doi:10.1088/0031-9155/53/4/014

9. Lomax AJ, Boehringer T, Coray A, et al. Intensity modulated proton therapy: A clinical example. *Med Phys*. 2001. doi:10.1118/1.1350587

10. Unkelbach J, Chan TCY, Bortfeld T. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. *Phys Med Biol*. 2007;52(10). doi:10.1088/0031-9155/52/10/009

11. Unkelbach J, Bortfeld T, Martin BC, Soukup M. Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning. *Med Phys*. 2009. doi:10.1118/1.3021139

12. Jäkel O, Debus J. Selection of beam angles for radiotherapy of skull base tumours using charged particles. *Phys Med Biol*. 2000;45(5):1229-1241. doi:10.1088/0031-9155/45/5/311

13. Trofimov A, Nguyen PL, Coen JJ, et al. Radiotherapy Treatment of Early-Stage Prostate Cancer with IMRT and Protons: A Treatment Planning Comparison. *Int J Radiat Oncol Biol Phys*. 2007;69(2):444-453. doi:10.1016/j.ijrobp.2007.03.018

14. Kase Y, Yamashita H, Fuji H, et al. A treatment planning comparison of passive-scattering and intensity-modulated proton therapy for typical tumor sites. *J Radiat Res*. 2012;53(2):272-280. doi:10.1269/jrr.11136

15. Steneker M, Lomax A, Schneider U. Intensity modulated photon and proton therapy for the treatment of head and neck tumors. *Radiother Oncol*. 2006;80(2):263-267. doi:10.1016/j.radonc.2006.07.025

16. DeLuca PM, Wambersie A, Whitmore G. Prescribing, recording, and reporting proton-beam therapy. *J ICRU*. 2007. doi:10.1093/jicru/ndm021

17. Gerweck LE, Kozin S V. Relative biological effectiveness of proton beams in clinical therapy. *Radiother Oncol*. 1999;50(2):135-142. doi:10.1016/S0167-8140(98)00092-9

18. Paganetti H, Niemierko A, Ancukiewicz M, et al. Relative biological effectiveness (RBE) values for proton beam therapy. *Int J Radiat Oncol Biol Phys*. 2002;53(2):407-421. doi:10.1016/S0360-3016(02)02754-2

19. Paganetti H, Blakely E, Carabe-Fernandez A, et al. Report of the AAPM TG-256 on the relative biological effectiveness of proton beams in radiation therapy. *Med Phys*. 2019. doi:10.1002/mp.13390

20. Dalrymple G V., Lindsay IR, Ghidoni JJ, et al. Some Effects of 138-Mev Protons on Primates. *Radiat Res*. 1966. doi:10.2307/3572210

21. Dalrymple G V., Lindsay IR, Hall JD, et al. The Relative Biological Effectiveness of 138-Mev Protons as Compared to Cobalt-60 Gamma Radiation. *Radiat Res*. 1966. doi:10.2307/3572211

22. Tepper J, Verhey L, Goitein M, Suit HD, Phil D, Koehler AM. In vivo determinations of RBE in a high energy modulated proton beam using normal tissue reactions and fractionated dose schedules. *Int J Radiat Oncol Biol Phys*. 1977;2(11-12):1115-1122. doi:10.1016/0360-3016(77)90118-3

23. Urano M, Goitein M, Verhey L, Mendiondo O, Suit HD, Koehler A. Relative biological effectiveness of a high energy modulated proton beam using a spontaneous murine tumor In vivo. *Int J Radiat Oncol Biol Phys*. 1980;6(9):1187-1193. doi:10.1016/0360-3016(80)90172-8

24. Wouters BG, Lam GK, Oelfke U, Gardey K, Durand RE, Skarsgard LD. Measurements of relative biological effectiveness of the 70 MeV proton beam at TRIUMF using Chinese hamster V79 cells and the high-precision cell sorter assay. *Radiat Res*. 1996;146(2):159-170. doi:10.2307/3579588

25. Paganetti H, Goitein M. Biophysical modelling of proton radiation effects based on amorphous track models. *Int J Radiat Biol*. 2001;77(9):911-928. doi:10.1080/09553000110066059

26. Wilkens JJ, Oelfke U. A phenomenological model for the relative biological effectiveness in therapeutic proton beams. *Phys Med Biol*. 2004;49(13):2811-2825. doi:10.1088/0031-9155/49/13/004

27. Paganetti H. Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer. *Phys Med Biol*. 2014;59(22):R419-R472. doi:10.1088/0031-9155/59/22/R419

28. Gridley DS, Pecaut MJ, Mao XW, Wroe AJ, Luo-Owen X. Biological Effects of Passive

Versus Active Scanning Proton Beams on Human Lung Epithelial Cells. *Technol Cancer Res Treat*. 2013;14(1):81-98. doi:10.7785/tcrt.2012.500392

29. Mishra M V, Khairnar R, Bentzen SM, et al. Proton beam therapy delivered using pencil beam scanning vs. passive scattering/uniform scanning for localized prostate cancer: Comparative toxicity analysis of PCG 001-09. *Clin Transl Radiat Oncol*. 2019;19:80-86. doi:10.1016/j.ctro.2019.08.006

30. Unkelbach J, Botas P, Giantsoudi D, Gorissen B, Paganetti H. Reoptimization of intensity-modulated proton therapy plans based on linear energy transfer. *Int J Radiat Oncol Biol Phys*. 2016;96(5):1097-1106. doi:10.1016/J.IJROBP.2016.08.038

31. An Y, Shan J, Patel SH, et al. Robust intensity-modulated proton therapy to reduce high linear energy transfer in organs at risk: *Med Phys*. 2017;44(12):6138-6147. doi:10.1002/mp.12610

32. Fjæra LF, Li Z, Ytre-Hauge KS, et al. Linear energy transfer distributions in the brainstem depending on tumour location in intensity-modulated proton therapy of paediatric cancer. *Acta Oncol (Madr)*. 2017. doi:10.1080/0284186X.2017.1314007

33. Giantsoudi D, Adams J, MacDonald SM, Paganetti H. Proton Treatment Techniques for Posterior Fossa Tumors: Consequences for Linear Energy Transfer and Dose-Volume Parameters for the Brainstem and Organs at Risk. *Int J Radiat Oncol Biol Phys*. 2017. doi:10.1016/j.ijrobp.2016.09.042

34. Mohan R, Peeler CR, Guan F, Bronk L, Cao W, Grosshans DR. Radiobiological issues in proton therapy. *Acta Oncol (Madr)*. 2017;56(11):1367-1373. doi:10.1080/0284186X.2017.1348621

35. Gu W, O'Connor D, Nguyen D, et al. Integrated beam orientation and scanning-spot optimization in intensity-modulated proton therapy for brain and unilateral head and neck tumors. *Med Phys*. 2018. doi:10.1002/mp.12788

36. Gu W, Ruan D, Lyu Q, Zou W, Dong L, Sheng K. A Novel Energy Layer Optimization Framework for Spot-Scanning Proton Arc Therapy. *Med Phys*. 2020. doi:10.1002/mp.14083

37. Gu W, Ruan D, O'Connor D, et al. Robust optimization for intensity-modulated proton therapy with soft spot sensitivity regularization. *Med Phys*. 2019. doi:10.1002/mp.13344

38. Gu W, Neph R, Ruan D, Zou W, Dong L, Sheng K. Robust Beam Orientation Optimization for Intensity-Modulated Proton Therapy. *Med Phys*. 2019;0(ja). doi:10.1002/mp.13641

39. Bortfeld T, Schlegel W. Optimization of beam orientations in radiation therapy: some theoretical considerations. *Phys Med Biol*. 1993;38(2):291-304. doi:10.1088/0031-9155/38/2/006

40. Li Y, Yao J, Yao D. Automatic beam angle selection in IMRT planning using genetic algorithm. *Phys Med Biol*. 2004;49(10):1915-1932. doi:10.1088/0031-9155/49/10/007

41. Pugachev A, Li JG, Boyer AL, et al. Role of beam orientation optimization in intensity-modulated radiation therapy. *Int J Radiat Oncol Biol Phys*. 2001;50(2):551-560. doi:10.1016/S0360-3016(01)01502-4

42. Li Y, Yao D, Yao J, Chen W. A particle swarm optimization algorithm for beam angle

selection in intensity-modulated radiotherapy planning. *Phys Med Biol*. 2005;50(15):3491-3514. doi:10.1088/0031-9155/50/15/002

43. Djajaputra D, Wu Q, Wu Y, Mohan R. Algorithm and performance of a clinical IMRT beam-angle optimization system. *Phys Med Biol*. 2003;48(19):3191-3212. doi:10.1088/0031-9155/48/19/007

44. Wang X, Zhang X, Dong L, Liu H, Wu Q, Mohan R. Development of methods for beam angle optimization for IMRT using an accelerated exhaustive search strategy. *Int J Radiat Oncol Biol Phys*. 2004;60(4):1325-1337. doi:10.1016/j.ijrobp.2004.06.007

45. Dias J, Rocha H, Ferreira B, Lopes MC. IMRT beam angle optimization using dynamically dimensioned search. In: *IFMBE Proceedings*. Vol 42. ; 2014:1-4. doi:10.1007/978-3-319-03005-0_1

46. Cao W, Lim GJ, Lee A, et al. Uncertainty incorporated beam angle optimization for IMPT treatment planning. *Med Phys*. 2012;39(8):5248-5256. doi:10.1118/1.4737870

47. Jia X, Men C, Lou Y, Jiang SB. Beam orientation optimization for intensity modulated radiation therapy using adaptive l(2,1)-minimization. *Phys Med Biol*. 2011;56(19):6205-6222. doi:10.1088/0031-9155/56/19/004

48. Albertini F, Hug EB, Lomax AJ. Is it necessary to plan with safety margins for actively scanned proton therapy? *Phys Med Biol*. 2011;56(14):4399-4413. doi:10.1088/0031-9155/56/14/011

49. Liu W, Zhang X, Li Y, Mohan R. Robust optimization of intensity modulated proton therapy. *Med Phys*. 2012;39(2):1079-1091. doi:10.1118/1.3679340

50. Thieke C, Bortfeld T, Küfer K-H. Characterization of dose distributions through the max and mean dose concept. *Acta Oncol*. 2002;41(9):158-161. doi:10.1080/028418602753669535

51. Cisternas E, Mairani A, Ziegenhein P, Jäkel O, Bangert M. matRad – a multi-modality open source 3D treatment planning toolkit. In: *IFMBE Proceedings*. Vol 51. ; 2015:1608-1611. doi:10.1007/978-3-319-19387-8_391

52. Wieser HP, Cisternas E, Wahl N, et al. Development of the open-source dose calculation and optimization toolkit matRad. *Med Phys*. 2017;44(6):2556-2568. doi:10.1002/mp.12251

53. Romeijn HE, Ahuja RK, Dempsey JF, Kumar A. A Column Generation Approach to Radiation Therapy Treatment Planning Using Aperture Modulation. *SIAM J Optim*. 2005;15(3):838-862. doi:10.1137/040606612

54. Nguyen D, Thomas D, Cao M, O'Connor D, Lamb J, Sheng K. Computerized triplet beam orientation optimization for MRI-guided Co-60 radiotherapy. *Med Phys*. 2016;43(10):5667-5675. doi:10.1118/1.4963212

55. Grégoire V, Mackie TR. State of the art on dose prescription, reporting and recording in Intensity-Modulated Radiation Therapy (ICRU report No. 83). *Cancer/Radiotherapie*. 2011;15(6-7):555-559. doi:10.1016/j.canrad.2011.04.003

56. Beck A, Teboulle M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J Imaging Sci*. 2009;2(1):183-202. doi:10.1137/080716542

57. Liu C, Bhangoo RS, Sio TT, et al. Dosimetric comparison of distal esophageal carcinoma plans for patients treated with small-spot intensity-modulated proton

versus volumetric-modulated arc therapies. *J Appl Clin Med Phys*. 2019. doi:10.1002/acm2.12623

58. Li X, Liu G, Janssens G, et al. The first prototype of spot-scanning proton arc treatment delivery. *Radiother Oncol*. 2019. doi:10.1016/j.radonc.2019.04.032

59. Ding X, Zhou J, Li X, et al. Improving dosimetric outcome for hippocampus and cochlea sparing whole brain radiotherapy using spot-scanning proton arc therapy. *Acta Oncol (Madr)*. 2019. doi:10.1080/0284186X.2018.1555374

60. Liu C, Sio TT, Deng W, et al. Small-spot intensity-modulated proton therapy and volumetric-modulated arc therapies for patients with locally advanced non-small-cell lung cancer: A dosimetric comparative study. *J Appl Clin Med Phys*. 2018. doi:10.1002/acm2.12459

61. Li X, Kabolizadeh P, Yan D, et al. Improve dosimetric outcome in stage III non-small-cell lung cancer treatment using spot-scanning proton arc (SPArc) therapy. *Radiat Oncol*. 2018. doi:10.1186/s13014-018-0981-6

62. Ding X, Li X, Qin A, et al. Have we reached proton beam therapy dosimetric limitations?–A novel robust, delivery-efficient and continuous spot-scanning proton arc (SPArc) therapy is to improve the dosimetric outcome in treating prostate cancer. *Acta Oncol (Madr)*. 2018. doi:10.1080/0284186X.2017.1358463

63. Ding X, Li X, Zhang JM, Kabolizadeh P, Stevens C, Yan D. Spot-Scanning Proton Arc (SPArc) Therapy: The First Robust and Delivery-Efficient Spot-Scanning Proton Arc Therapy. *Int J Radiat Oncol Biol Phys*. 2016. doi:10.1016/j.ijrobp.2016.08.049

64. Rah JE, Kim GY, Oh DH, et al. A treatment planning study of proton arc therapy for

para-aortic lymph node tumors: Dosimetric evaluation of conventional proton therapy, proton arc therapy, and intensity modulated radiotherapy. *Radiat Oncol*. 2016. doi:10.1186/s13014-016-0717-4

65. Dink D, Langer MP, Rardin RL, Pekny JF, Reklaitis G V., Saka B. Intensity modulated radiation therapy with field rotation-a time-varying fractionation study. *Health Care Manag Sci*. 2012. doi:10.1007/s10729-012-9190-2

66. O'Connor D, Yu V, Nguyen D, Ruan D, Sheng K. Fraction-variant beam orientation optimization for non-coplanar IMRT. *Phys Med Biol*. 2018. doi:10.1088/1361-6560/aaa94f

67. Girdhani S, Abel E, Katsis A, et al. Abstract LB-280: FLASH: A novel paradigm changing tumor irradiation platform that enhances therapeutic ratio by reducing normal tissue toxicity and activating immune pathways. In: ; 2019. doi:10.1158/1538-7445.sabcs18-lb-280

68. Favaudon V, Caplier L, Monceau V, et al. Ultrahigh dose-rate FLASH irradiation increases the differential response between normal and tumor tissue in mice. *Sci Transl Med*. 2014. doi:10.1126/scitranslmed.3008973

69. Montay-Gruel P, Petersson K, Jaccard M, et al. Irradiation in a flash: Unique sparing of memory in mice after whole brain irradiation with dose rates above 100 Gy/s. *Radiother Oncol*. 2017. doi:10.1016/j.radonc.2017.05.003

70. Unkelbach J, Zeng C, Engelsman M. Simultaneous optimization of dose distributions and fractionation schemes in particle radiotherapy. *Med Phys*. 2013. doi:10.1118/1.4816658

71. Unkelbach J, Papp D. The emergence of nonuniform spatiotemporal fractionation schemes within the standard BED model. *Med Phys*. 2015. doi:10.1118/1.4916684

72. Unkelbach J, Papp D, Gaddy MR, Andratschke N, Hong T, Guckenberger M. Spatiotemporal fractionation schemes for liver stereotactic body radiotherapy. *Radiother Oncol*. 2017. doi:10.1016/j.radonc.2017.09.003

73. MacKin D, Li Y, Taylor MB, et al. Improving spot-scanning proton therapy patient specific quality assurance with HPlusQA, a second-check dose calculation engine. *Med Phys*. 2013. doi:10.1118/1.4828775

74. Zhu XR, Li Y, Mackin D, et al. Towards effective and efficient patient-specific quality assurance for spot scanning proton therapy. *Cancers (Basel)*. 2015. doi:10.3390/cancers7020631

75. Meier G, Besson R, Nanz A, Safai S, Lomax AJ. Independent dose calculations for commissioning, quality assurance and dose reconstruction of PBS proton therapy. *Phys Med Biol*. 2015. doi:10.1088/0031-9155/60/7/2819

76. Li H, Sahoo N, Poenisch F, et al. Use of treatment log files in spot scanning proton therapy as part of patient-specific quality assurance. *Med Phys*. 2013. doi:10.1118/1.4773312

77. Scandurra D, Albertini F, Van Der Meer R, et al. Assessing the quality of proton PBS treatment delivery using machine log files: Comprehensive analysis of clinical treatments delivered at PSI Gantry 2. *Phys Med Biol*. 2016. doi:10.1088/0031-9155/61/3/1171

78. Belosi MF, van der Meer R, Garcia de Acilu Laa P, Bolsi A, Weber DC, Lomax AJ.

Treatment log files as a tool to identify treatment plan sensitivity to inaccuracies in scanned proton beam delivery. *Radiother Oncol*. 2017. doi:10.1016/j.radonc.2017.09.037

79. McNamara AL, Schuemann J, Paganetti H. A phenomenological relative biological effectiveness (RBE) model for proton therapy based on all published in vitro cell survival data. *Phys Med Biol*. 2015;60(21):8399-8416. doi:10.1088/0031-9155/60/21/8399

80. Carabe-Fernandez A, Dale RG, Jones B. The incorporation of the concept of minimum RBE (RbEmin) into the linear-quadratic model and the potential for improved radiobiological analysis of high-LET treatments. *Int J Radiat Biol*. 2007;83(1):27-39. http://www.ncbi.nlm.nih.gov/pubmed/17357437.

81. Wedenberg M, Lind BK, Hårdemark B. A model for the relative biological effectiveness of photons is a predictor for the sensitivity to LET changes. *Acta Oncol (Madr)*. 2012;52(June 2012):1-9. doi:10.3109/0284186X.2012.705892

82. Jones B, McMahon SJ, Prise KM. The Radiobiology of Proton Therapy: Challenges and Opportunities Around Relative Biological Effectiveness. *Clin Oncol*. 2018;30(5):285-292. doi:10.1016/j.clon.2018.01.010

83. Wilkens JJ, Oelfke U. Optimization of radiobiological effects in intensity modulated proton therapy. *Med Phys*. 2005;32(2):455-465. doi:10.1118/1.1851925

84. Frese MC, Wilkens JJ, Huber PE, Jensen AD, Oelfke U, Taheri-Kadkhoda Z. Application of constant vs. variable relative biological effectiveness in treatment planning of intensity-modulated proton therapy. *Int J Radiat Oncol Biol Phys*. 2011;79(1):80-88.

doi:10.1016/j.ijrobp.2009.10.022

85. Guan F, Geng C, Ma D, et al. RBE model-based biological dose optimization for proton radiobiology studies. *Int J Part Ther*. 2019. doi:10.14338/IJPT-18-00007.1

86. Takada K, Sato T, Kumada H, et al. Validation of the physical and RBE-weighted dose estimator based on PHITS coupled with a microdosimetric kinetic model for proton therapy. *J Radiat Res*. 2018. doi:10.1093/jrr/rrx057

87. Grassberger C, Paganetti H. Elevated LET components in clinical proton beams. *Phys Med Biol*. 2011;56(20):6677-6691. doi:10.1088/0031-9155/56/20/011

88. Wilkens JJ, Oelfke U. Analytical linear energy transfer calculations for proton therapy. *Med Phys*. 2003;30(5):806-815. doi:10.1118/1.1567852

89. Wilkens JJ, Oelfke U. Three-dimensional LET calculations for treatment planning of proton therapy. *Z Med Phys*. 2004. doi:10.1078/0939-3889-00191

90. Sanchez-Parcerisa D, Cortés-Giraldo MA, Dolney D, Kondrla M, Fager M, Carabe A. Analytical calculation of proton linear energy transfer in voxelized geometries including secondary protons. *Phys Med Biol*. 2016;61(4):1705-1721. doi:10.1088/0031-9155/61/4/1705

91. Hirayama S, Matsuura T, Ueda H, et al. An analytical dose-averaged LET calculation algorithm considering the off-axis LET enhancement by secondary protons for spot-scanning proton therapy. *Med Phys*. 2018. doi:10.1002/mp.12991

92. Cortés-Giraldo MA, Carabe A. A critical study of different Monte Carlo scoring methods of dose average linear-energy-transfer maps calculated in voxelized

geometries irradiated with clinical proton beams. *Phys Med Biol*. 2015. doi:10.1088/0031-9155/60/7/2645

93. Perl J, Shin J, Faddegon B, Paganetti H. TOPAS : An innovative proton Monte Carlo platform for research. *Med Phys*. 2012;39(11):6818-6837. http://dx.doi.org/10.1118/1.4758060.

94. Grzanka L, Ardenfors O, Bassler N. Monte carlo simulations of spatial let distributions in clinical proton beams. *Radiat Prot Dosimetry*. 2018. doi:10.1093/RPD/NCX272

95. Qin N, Botas P, Giantsoudi D, et al. Recent developments and comprehensive evaluations of a GPU-based Monte Carlo package for proton therapy. *Phys Med Biol*. 2016. doi:10.1088/0031-9155/61/20/7347

96. Grassberger C, Trofimov A, Lomax A, Paganetti H. Variations in linear energy transfer within clinical proton therapy fields and the potential for biological treatment planning. *Int J Radiat Oncol Biol Phys*. 2011;80(5):1559-1566. doi:10.1016/j.ijrobp.2010.10.027

97. Wan Chan Tseung HS, Ma J, Kreofsky CR, Ma DJ, Beltran C. Clinically Applicable Monte Carlo–based Biological Dose Optimization for the Treatment of Head and Neck Cancers With Spot-Scanning Proton Therapy. *Int J Radiat Oncol Biol Phys*. 2016;95(5):1535-1543. doi:10.1016/j.ijrobp.2016.03.041

98. Fager M, Toma-Dasu I, Kirk M, et al. Linear energy transfer painting with proton therapy: A means of reducing radiation doses with equivalent clinical effectiveness. *Int J Radiat Oncol Biol Phys*. 2015;91(5):1057-1064. doi:10.1016/j.ijrobp.2014.12.049

99. Bassler N, Jäkel O, Søndergaard CS, Petersen JB. Dose- and LET-painting with particle therapy. *Acta Oncol (Madr)*. 2010;49(7):1170-1176. doi:10.3109/0284186X.2010.510640

100. Giantsoudi D, Grassberger C, Craft D, Niemierko A, Trofimov A, Paganetti H. Linear energy transfer-guided optimization in intensity modulated proton therapy: Feasibility study and clinical potential. *Int J Radiat Oncol Biol Phys*. 2013;87(1):216-222. doi:10.1016/j.ijrobp.2013.05.013

101. Cao W, Khabazian A, Yepes PP, et al. Linear energy transfer incorporated intensity modulated proton therapy optimization. *Phys Med Biol*. 2018. doi:10.1088/1361-6560/aa9a2e

102. Bai X, Lim G, Grosshans D, Mohan R, Cao W. Robust optimization to reduce the impact of biological effect variation from physical uncertainties in intensity-modulated proton therapy. *Phys Med Biol*. 2019. doi:10.1088/1361-6560/aaf5e9

103. Inaniwa T, Kanematsu N, Noda K, Kamada T. Treatment planning of intensity modulated composite particle therapy with dose and linear energy transfer optimization. *Phys Med Biol*. 2017. doi:10.1088/1361-6560/aa68d7

104. Deasy J, Mackie T, DeLuca P. Method and apparatus for proton therapy. 1997.

105. Sandison GA, Papiez E, Block C, Morphis J. Phantom assessment of lung dose from proton arc therapy. *Int J Radiat Oncol Biol Phys*. 1997. doi:10.1016/S0360-3016(97)00059-X

106. Schreuder AN, Shamblin J. Proton therapy delivery: what is needed in the next ten years? *Br J Radiol*. 2019. doi:10.1259/bjr.20190359

107. Pedroni E, Bacher R, Blattmann H, et al. The 200-Mev proton therapy project at the Paul Scherrer Institute: Conceptual design and practical realization. *Med Phys*. 1995. doi:10.1118/1.597522

108. Pedroni E, Scheib S, Böhringer T, et al. Experimental characterization and physical modelling of the dose distribution of scanned proton pencil beams. *Phys Med Biol*. 2005. doi:10.1088/0031-9155/50/3/011

109. Smith A, Gillin M, Bues M, et al. The M. D. Anderson proton therapy system. *Med Phys*. 2009. doi:10.1118/1.3187229

110. Gillin MT, Sahoo N, Bues M, et al. Commissioning of the discrete spot scanning proton beam delivery system at the University of Texas M.D. Anderson Cancer Center, Proton Therapy Center, Houston. *Med Phys*. 2010. doi:10.1118/1.3259742

111. Zhu XR, Sahoo N, Zhang X, et al. Intensity modulated proton therapy treatment planning using single-field optimization: The impact of monitor unit constraints on plan quality. *Med Phys*. 2010. doi:10.1118/1.3314073

112. Cao W, Lim G, Liao L, et al. Proton energy optimization and reduction for intensity-modulated proton therapy. *Phys Med Biol*. 2014. doi:10.1088/0031-9155/59/21/6341

113. Van De Water S, Kooy HM, Heijmen BJM, Hoogeman MS. Shortening delivery times of intensity modulated proton therapy by reducing proton energy layers during treatment plan optimization. *Int J Radiat Oncol Biol Phys*. 2015. doi:10.1016/j.ijrobp.2015.01.031

114. Seco J, Gu G, Marcelos T, Kooy H, Willers H. Proton arc reduces range uncertainty

effects and improves conformality compared with photon volumetric modulated arc therapy in stereotactic body radiation therapy for non-small cell lung cancer. *Int J Radiat Oncol Biol Phys*. 2013. doi:10.1016/j.ijrobp.2013.04.048

115. Rechner LA, Howell RM, Zhang R, Etzel C, Lee AK, Newhauser WD. Risk of radiogenic second cancers following volumetric modulated arc therapy and proton arc therapy for prostate cancer. *Phys Med Biol*. 2012. doi:10.1088/0031-9155/57/21/7117

116. Flynn RT, Barbee DL, Mackie TR, Jeraj R. Comparison of intensity modulated x-ray therapy and intensity modulated proton therapy for selective subvolume boosting: A phantom study. *Phys Med Biol*. 2007. doi:10.1088/0031-9155/52/20/001

117. Oelfke U, Bortfeld T. Intensity modulated radiotherapy with charged particle beams: Studies of inverse treatment planning for rotation therapy. *Med Phys*. 2000. doi:10.1118/1.599002

118. Sanchez-Parcerisa D, Kirk M, Fager M, et al. Range optimization for mono- and bi-energetic proton modulated arc therapy with pencil beam scanning. *Phys Med Biol*. 2016. doi:10.1088/0031-9155/61/21/N565

119. Blanco Kiely JP, White BM. Dosimetric feasibility of single-energy proton modulated arc therapy for treatment of chordoma at the skull base. *Acta Oncol (Madr)*. 2016. doi:10.3109/0284186X.2016.1170199

120. Langner UW, Eley JG, Guerrero M, et al. A method to deliver energy modulated planar proton Arc therapy (EMPPAT). *J Prot Ther*. 2017;3(312).

121. Otto K. Volumetric modulated arc therapy: IMRT in a single gantry arc. *Med Phys*. 2008. doi:10.1118/1.2818738

122. Ding X, Li X, Liu G, Stevens C, Yan D, Kabolizadeh P. PO-0916 Energy layer switching sequence optimization algorithm for an efficiency proton arc therapy delivery. *Radiother Oncol*. 2019. doi:10.1016/s0167-8140(19)31336-2

123. Park PC, Zhu XR, Lee AK, et al. A beam-specific planning target volume (PTV) design for proton therapy to account for setup and range uncertainties. *Int J Radiat Oncol Biol Phys*. 2012;82(2). doi:10.1016/j.ijrobp.2011.05.011

124. Bangert M, Hennig P, Oelfke U. Analytical probabilistic modeling for radiation therapy treatment planning. *Phys Med Biol*. 2013. doi:10.1088/0031-9155/58/16/5401

125. Wahl N, Hennig P, Wieser HP, Bangert M. Efficiency of analytical and sampling-based uncertainty propagation in intensity-modulated proton therapy. *Phys Med Biol*. 2017. doi:10.1088/1361-6560/aa6ec5

126. Gordon JJ, Sayah N, Weiss E, Siebers J V. Coverage optimized planning: Probabilistic treatment planning based on dose coverage histogram criteria. *Med Phys*. 2010. doi:10.1118/1.3273063

127. Pflugfelder D, Wilkens JJ, Szymanowski H, Oelfke U. Quantifying lateral tissue heterogeneities in hadron therapy. *Med Phys*. 2007;34(4):1506-1513. doi:10.1118/1.2710329

128. Bueno M, Paganetti H, Duch MA, Schuemann J. An algorithm to assess the need for clinical Monte Carlo dose calculation for small proton therapy fields based on quantification of tissue heterogeneity. *Med Phys*. 2013;40(8). doi:10.1118/1.4812682

129. Toramatsu C, Inaniwa T. Beam angle selection incorporation of anatomical

heterogeneities for pencil beam scanning charged-particle therapy. *Phys Med Biol*. 2016;61(24):8664-8675. doi:10.1088/1361-6560/61/24/8664

130. Liu W, Mohan R, Park P, et al. Dosimetric benefits of robust treatment planning for intensity modulated proton therapy for base-of-skull cancers. *Pract Radiat Oncol*. 2014. doi:10.1016/j.prro.2013.12.001

131. Liu W, Frank SJ, Li X, Li Y, Zhu RX, Mohan R. PTV-based IMPT optimization incorporating planning risk volumes vs robust optimization. *Med Phys*. 2013. doi:10.1118/1.4774363

132. Pflugfelder D, Wilkens JJ, Oelfke U. Worst case optimization: A method to account for uncertainties in the optimization of intensity modulated proton therapy. *Phys Med Biol*. 2008;53(6):1689-1700. doi:10.1088/0031-9155/53/6/013

133. Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. *Med Phys*. 2011;38(3):1672-1684. doi:10.1118/1.3556559

134. Fredriksson A, Bokrantz R. A critical evaluation of worst case optimization methods for robust intensity-modulated proton therapy planning. *Med Phys*. 2014;41(8). doi:10.1118/1.4883837

135. Chen W, Unkelbach J, Trofimov A, et al. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. *Phys Med Biol*. 2012;57(3):591-608. doi:10.1088/0031-9155/57/3/591

136. Stuschke M, Kaiser A, Pöttgen C, Lübcke W, Farr J. Potentials of robust intensity modulated scanning proton plans for locally advanced lung cancer in comparison to

intensity modulated photon plans. *Radiother Oncol*. 2012;104(1):45-51. doi:10.1016/j.radonc.2012.03.017

137. Casiraghi M, Albertini F, Lomax AJ. Advantages and limitations of the "worst case scenario" approach in IMPT treatment planning. *Phys Med Biol*. 2013. doi:10.1088/0031-9155/58/5/1323

138. Li Y, Niemela P, Liao L, et al. Selective robust optimization: A new intensity-modulated proton therapy optimization strategy. *Med Phys*. 2015. doi:10.1118/1.4923171

139. Unkelbach J, Alber M, Bangert M, et al. Robust radiotherapy planning. *Phys Med Biol*. 2018. doi:10.1088/1361-6560/aae659

140. Xing L, Boyd S. Real-Time Radiation Treatment Planning with Optimality Guarantees via Cluster and Bound Methods Problem description. 2017:1-24.

141. Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J Math Imaging Vis*. 2011. doi:10.1007/s10851-010-0251-1

142. van de Water S, van Dam I, Schaart DR, Al-Mamgani A, Heijmen BJM, Hoogeman MS. The price of robustness; impact of worst-case optimization on organ-at-risk dose and complication probability in intensity-modulated proton therapy for oropharyngeal cancer patients. *Radiother Oncol*. 2016;120(1):56-62. doi:10.1016/j.radonc.2016.04.038

143. Shan J, Sio TT, Liu C, Schild SE, Bues M, Liu W. A novel and individualized robust optimization method using normalized dose interval volume constraints (NDIVC) for intensity-modulated proton radiotherapy. *Med Phys*. 2018;0(0).

doi:10.1002/mp.13276

144. Liu W, Li Y, Li X, Cao W, Zhang X. Influence of robust optimization in intensity-modulated proton therapy with different dose delivery techniques. *Med Phys*. 2012. doi:10.1118/1.4711909

145. Cao W, Lim GJ, Li Y, Zhu XR, Zhang X. Improved beam angle arrangement in intensity modulated proton therapy treatment planning for localized prostate cancer. *Cancers (Basel)*. 2015. doi:10.3390/cancers7020574

146. Lim GJ, Kardar L, Cao W. A hybrid framework for optimizing beam angles in radiation therapy planning. *Ann Oper Res*. 2014. doi:10.1007/s10479-014-1564-z

147. Liao L, Lim GJ, Li Y, et al. Robust Optimization for Intensity Modulated Proton Therapy Plans with Multi-Isocenter Large Fields. *Int J Part Ther*. 2016. doi:10.14338/ijpt-16-00012.1

148. Liu W, Frank SJ, Li X, et al. Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers. *Med Phys*. 2013. doi:10.1118/1.4801899

149. Liu W, Liao Z, Schild SE, et al. Impact of respiratory motion on worst-case scenario optimized intensity modulated proton therapy for lung cancers. *Pract Radiat Oncol*. 2015. doi:10.1016/j.prro.2014.08.002

150. Li H, Zhang X, Park P, et al. Robust optimization in intensity-modulated proton therapy to account for anatomy changes in lung cancer patients. *Radiother Oncol*. 2015. doi:10.1016/j.radonc.2015.01.017

151. Van Der Voort S, Van De Water S, Perkó Z, Heijmen B, Lathouwers D, Hoogeman M. Robustness Recipes for Minimax Robust Optimization in Intensity Modulated Proton Therapy for Oropharyngeal Cancer Patients. *Int J Radiat Oncol Biol Phys*. 2016. doi:10.1016/j.ijrobp.2016.02.035

152. Van Dijk L V., Steenbakkers RJHM, Ten Haken B, et al. Robust Intensity Modulated Proton Therapy (IMPT) increases estimated clinical benefit in head and neck cancer patients. *PLoS One*. 2016. doi:10.1371/journal.pone.0152477

153. Safai S, Bula C, Meer D, Pedroni E. Improving the precision and performance of proton pencil beam scanning. *Transl Cancer Res*. 2012;1(3). http://tcr.amegroups.com/article/view/599.

154. Gottschalk B, Koehler AM, Schneider RJ, Sisterson JM, Wagner MS. Multiple Coulomb scattering of 160 MeV protons. *Nucl Inst Methods Phys Res B*. 1993. doi:10.1016/0168-583X(93)95944-Z

155. Moellenhoff T, Strekalovskiy E, Möller M, Cremers D. Low Rank Priors for Color Image Regularization. *Submitted*. 2014:1-14. doi:10.1007/978-3-319-14612-6_10