

Optimization of proton arc therapy treatment plans

Sophie Wuyckens

A thesis presented for the degree of
Doctor of Biomedical and Pharmaceutical Sciences



Molecular Imaging, Radiotherapy and Oncology (MIRO)
Institut de Recherche Expérimentale et Clinique (IREC)
Université catholique de Louvain (UCLouvain)
Belgium



Beaumont



Sophie Wuyckens was supported by the Walloon Region as part of the Arc Proton Therapy convention (Pôles Mecatech et Biowin)

Acknowledgements

Over the course of my PhD training, I have had the opportunity to work, present my research, attend specialized courses, and meet amazing people in many cities scattered around the world: Pisa, Padua, Naples, Paris, Washington DC, Geneva, London, Miami, Detroit, Madrid, Lisbon, Stockholm, Vienna, and of course, Louvain-La-Neuve and Brussels. These years of my doctoral program have provided me with a wealth of experiences, both professional and personal. I would like to express my gratitude to all of the individuals who have contributed to what I would call my “journey.”

First and foremost, I would like to thank the supervisors of the first thesis I commenced with, Andrea Giammanco and Eduardo Cortinal Gil, with whom everything began in 2018 in the field of Particle Physics. Even though I decided to move towards Medical Physics after two years, I have no regrets about starting my career with you. You are the individuals who introduced me to the fascinating world of research (and even to the world in general) and taught me so much.

Next, I would like to express my gratitude to my current thesis advisors, John, Kevin, and Edmond. Kevin, you have been more than a mentor to me; you have taken me under your wing. I knew very little about Medical Physics when I first arrived, but you have been patient, supportive, and encouraging throughout. John, thank you for your kindness and all the guidance and insightful discussions on mathematical modeling. Edmond, thank you for being there for me, especially during some trying situations. I would also like to thank the members of my PhD committee who have followed my work over the past four years, offering valuable constructive criticism and feedback to enhance the quality of my research. Additionally, I extend my thanks to the invited committee members for dedicating time to thoroughly review my manuscript.

My journey is about seizing opportunities. Each chapter of this thesis is the

result of a fruitful collaboration. Therefore, there are many individuals I would like to thank;

I would like to thank my friend Michael Saint-Guillain, alias Mick. I came to you at the beginning of my thesis asking for advice in optimization given your huge expertise background in the field. It actually started a great collaboration stretching the proton arc therapy optimization problem in all directions. My first two papers I owe them for a large part to you. I should also thank Guillaume Janssens, who let me re-implement his method presented in Chapter 5 into OpenTPS under his close supervision. Thank you for your availability and kindness, especially lately with my very often meeting requests rushing the last paper.

I was lucky to make two international secondments during this PhD. The first one was at the proton therapy center (PTC) in Beaumont Hospital (MI, USA). I would like to thank my advisor there, Xuanfeng Ding, who has continued to offer me great opportunities since then. I should also thank all other PTC colleagues, in particular Lewei with who I published several papers thanks to his expertise in clinic coupled to mathematical optimization. I recently made a second research stay at RaySearch Laboratories (Stockholm, Sweden) under the supervision of Erik Engwall along with the arc optimization team, Viktor, Otte and Johan. You really took me in as one of yours making my stay so interesting and enjoyable at all levels and I thank you for that. Thank you also to the optimization group and the great fika times. Special mention to Viktor with who whiteboards have never been so colorful with all our great (or bad) ideas but also for your constant help and support.

My heartfelt gratitude also goes out to my colleagues and friends who have supported me throughout my PhD journey. Marie, my first office mate in MIRO, we managed to get close even it was covid times. I really enjoyed our working (chatting ?) days together. Then comes another great person, Macarena, always willing to help even when the task is boring. I would like to thank the whole OpenTPS team: Sylvain, Kevin, Damien, Guillaume, Valentin, Margerie and Ana. What a team work we accomplished! I thank my buddy Loïc with who I started and now ending the PhD with. Congresses are much funnier when you are there. I feel grateful to have had the opportunity to work alongside such talented and dedicated individuals, and I hope to continue collaborating with you in the future. Moreover, thank you to my best friends, les *Vrés* for all the support and belief in me.

Last but certainly not least, I would like to express my deep gratitude to my family. A big thank you for taking care of my beast while I was away (most of the

time). Special mention to my sister, Dr. Gege, who just went through the whole PhD process, you have been a great support and are a real inspiration to me. Finally, I should thank my partner, Timour for your patience and encouragement and for staying by me, despite my hectic professional life.

To all those mentioned above, and to many others whose names I may have overlooked, I owe a debt of gratitude. Your contributions, whether big or small, have been meaningful to the completion of this thesis. Thank you all et merci à tous.

Abstract

Proton arc therapy (PAT) is a novel modality to treat cancer, involving continuous rotation of the irradiation device (gantry) around the patient while delivering a proton beam to the tumor. Compared to conventional fixed-angle proton beam delivery, PAT paves a promising treatment avenue, with numerous possible irradiation directions, to improve plan quality and streamline the treatment workflow. Nevertheless, this increased degree of freedom also becomes a burden for its clinical implementation, in particular in the treatment planning optimization step. The cornucopia of irradiation directions and energy layer candidates increases the complexity of the associated treatment planning optimization problem. It requires substantial computational power as well as the development of new algorithms and objectives to yield optimal treatment plans that are both high in quality and efficient to deliver. To this purpose, one must solve the energy layer selection problem, which constitutes the major challenge of PAT treatment planning optimization. This thesis is dedicated to addressing this challenge and providing potential solutions to it. The chapters unfold as a journey of deeper and deeper exploration of PAT, maturation of the problem statement, and refinement of the proposed solutions. Initially, we tackled the problem with all available degrees of freedom, but practical constraints arose. The computational demand of implemented solvers proved to be too prohibitive. Consequently, we transitioned to a second approach that selectively incorporates a subset of degrees of freedom, aligning with technical constraints. Ultimately, this thesis concludes with an in-depth theoretical exploration of the PAT problem, confirming its inherent complexity. This work contributes to advancing the field of PAT treatment planning optimization, addressing the delicate balance between plan quality and clinical feasibility.

PhD committee

Promoters

Prof. John A. Lee, PhD, Ir

Molecular Imaging, Radiotherapy and Oncology, Université catholique de Louvain
École polytechnique de Louvain, Université catholique de Louvain

Prof. Edmond Sterpin, PhD, Ir

Molecular Imaging, Radiotherapy and Oncology, Université catholique de Louvain
Laboratory of Experimental Radiotherapy, Katholieke Universiteit Leuven
Particle Therapy Interuniversity Center Leuven

Dr. Kevin Souris, PhD

Ion Beam Applications S.A. (IBA)

Molecular Imaging, Radiotherapy and Oncology, Université catholique de Louvain

President of the jury

Prof. Frédéric Lecouvet PhD, MD

Radiology, Cliniques Universitaires Saint Luc

Pôle d'Imagerie Médicale (IMAG), Université catholique de Louvain

Members

Dr. Guillaume Janssens, PhD, Ir

Ion Beam Applications S.A. (IBA)

Prof. Xavier Geets, PhD, MD

Radiation Oncology, Cliniques Universitaires Saint Luc

Molecular Imaging, Radiotherapy and Oncology, Université catholique de Louvain

Prof. Eduardo Cortina Gil, PhD

Centre for Cosmology, Particle Physics and Phenomenology, Université catholique de Louvain

Prof. François Glineur, PhD, Ir

Institute of Information and Communication Technologies, Electronics and Applied Mathematics, Université catholique de Louvain

Invited members

Dr. Erik Engwall, PhD

RaySearch Laboratories AB

Prof. Thomas Rockwell Mackie, PhD

Department of Medical Physics, University of Wisconsin School of Medicine and Public Health

List of publications

Peer reviewed publications

Treatment planning in arc proton therapy: Comparison of several optimization problem statements and their corresponding solvers

S. Wuyckens, M. Saint-Guillain, G. Janssens, L. Zhao , X. Li, X. Ding , E. Sterpin, J. A. Lee and K. Souris. [Computers in Biology and Medicine](#), 148:105609, 2022

Bi-criteria Pareto optimization to balance irradiation time and dosimetric objectives in proton arc therapy

S. Wuyckens, L. Zhao , M. Saint-Guillain, G. Janssens, E. Sterpin, K. Souris, X. Ding and J. A. Lee. [Physics in Medicine and Biology](#). 67 245017, 2022

Faster proton arc optimization and delivery through energy layer pre-selection and post-filtering

S. Wuyckens, V. Wase, O. Martin, J. Sundström, G. Janssens, E. Borderias-Villarroel, E. Sterpin, K. Souris, E. Engwall and J. A. Lee. Submitted to Medical Physics, 2023.

The proton arc therapy treatment planning problem is NP-hard

V. Wase, S. Wuyckens, J.A. Lee, M. Saint-Guillain. Submitted to Computers in Biology and Medicine, 2023.

Beamlet-free optimization for Monte-Carlo-based treatment planning in proton therapy

D. Pross, S. Wuyckens, S. Deffet, E. Sterpin, J. A. Lee and K. Souris. [Medical Physics](#), 1-9, 2023.

Proton arc therapy for oesophageal cancer: Is the dosimetric benefit worth the technological effort ?

A-C. Wéra, M. Chocan Vera, H. Ozan, E. Engwall , V. Wase, O. Marthin, J. Sundström, S. Wuyckens, K. Haustermans, A. M. Barragán-Montero, K. Souris, J. A. Lee, E. Sterpin. Submitted to Physica Medica, 2023.

Machine-Specific Delivery Sequence Model of Compact Superconducting Synchrocyclotron Proton Therapy Systems – A Multi-Institutional Investigation

L. Zhao, G. Liu, K. Souris, S. Wuyckens, G. Janssens, K. Poels, A. Delor, T. Depuydt, R. L. Deraniyagala Jr, C. W. Stevens, X. Li and X. Ding. [International Journal of Radiation Oncology*Biophysics](#), 114:3, e541, 2022

The first direct method of spot sparsity optimization for proton arc therapy

L. Zhao, G. Liu, K. Souris, S. Wuyckens, G. Janssens, K. Poels, A. Delor, T. Depuydt, R. L. Deraniyagala Jr, C. W. Stevens, X. Li and X. Ding. [Acta Oncologica](#), 62:1, 48-52, 2023.

Preprints

Proton arc plan optimization based on organ at risk sparing and delivery time driven energy layer pre-selection

S. Wuyckens, G. Janssens, M. Chocan Vera, D. Di Perri, E. Sterpin, K. Souris, and J. A. Lee. Manuscript in preparation.

Particle Arc Therapy: status and potential

S. Mein, S. Wuyckens, X. Li, S. Both, A. Caraba, M. Chocan Vera, E. Engwall, F. Francesco, C. Graeff, L. Hong, G. Janssens, B. de Jong,, T. Li, X. Liang, G. Liu, T. Mackie, A. Mairani, A. Mazal, K. Nesteruk, N. Schreuder, M. Soukoup, A. Lomax, H. Paganetti, T. Tessonier, L. Volz, L. Zhao, X. Ding. Manuscript in preparation for submission in Radiotherapy & Oncology, 2023.

OpenTPS – Open-source treatment planning system for research in proton therapy

S. Wuyckens, D. Dasnoy, G. Janssens, V. Hamaide, M. Huet, E. Lojén, G. Rotsart de Hertaing, B. Macq, E. Sterpin, J. A. Lee, K. Souris, S. Deffet. [ArXiv](#), 2022.

International conference communications

S. Wuyckens, G. Janssens, E. Sterpin, J. Lee, K. Souris, *FISTArcPT: a Beamlet-based optimization for Arc Proton Therapy*, Annual meeting of Belgian Hospital Physicists Association (BHPA), 2021, Virtual

S. Wuyckens, G. Janssens, E. Sterpin, J. Lee, K. Souris, [WE-A-TRACK 6-6](#) *A Beamlet-Based Optimization Algorithm for Arc Proton Therapy Planning*, Annual Meeting of the American Association of Physicists in Medicine (AAPM), 2021, Virtual

S. Wuyckens, G. Janssens, E. Sterpin, J. Lee, K. Souris, [OC-0421](#) *Towards a new optimization algorithm for Arc Proton Therapy treatment planning*, Annual Meeting of the European Society for Radiotherapy and Oncology (ESTRO), 2021, Madrid (Spain)

S. Wuyckens, G. Janssens, E. Sterpin, J. Lee, K. Souris, [P138](#) *Development of a new optimization algorithm for Arc Proton Therapy treatment planning*, 59th Annual Conference of the Particle Therapy Cooperative Group (PTCOG 59), 2021, Virtual

S. Wuyckens, M. Saint-Guillain, G. Janssens, E. Sterpin, J. Lee, K. Souris, [P 115](#) *A linear formulation of the Proton Arc Therapy Optimization problem solved with Mixed-Integer Programming*, 60th Annual Conference of the Particle Therapy Cooperative Group (PTCOG 60), 2022, Miami (FL, USA)

S. Wuyckens, M. Saint-Guillain, G. Janssens, E. Sterpin, J. Lee, K. Souris, [MO-H345-IePD-F4-3](#) *ArcPT Plan Optimization: Thinking Out of the Box with Mixed-Integer Programming*, Annual Meeting of the American Association of Physicists in Medicine (AAPM), 2022, Washington DC (USA)

S. Wuyckens, L. Zhao, X. Ding, E. Sterpin, J. A. Lee, K. Souris, *Bicriteria optimization applied to proton arc therapy problem*, Spring Meeting & Young Investigators Symposium (AAPM Great Lakes Chapter), 2022, Southfield (MI, USA)

S. Wuyckens, L. Zhao, X. Ding, E. Sterpin, J. A. Lee, K. Souris, *Bicriteria optimization applied to proton arc therapy problem*, 2022 SIAM Great Lakes Section Annual Meeting (GLSIAM) hosted by Wayne State University, 2022, Detroit (MI, USA)

S. Wuyckens, *Walking on the ridge line between harnessing new degrees of freedom and complying with stringent technical constraints in PAT treatment optimization*, ESTRO Physics workshop: Particle Arc Therapy: from concept to clinical reality, 2022, Lisbon (Portugal)

M. Chocan, H. Ozan, A-C. Wera, S. Wuyckens, E. Engwall, V. Wase, O. Marthin, J. Sundström, A. Barragan, J. A. Lee, E. Sterpin, [PO-1963](#) *Arc proton therapy for esophageal cancer: feasibility and first plan comparison with IMPT*, Annual Meeting of the European Society for Radiotherapy and Oncology (ESTRO), 2023, Vienna (Austria)

S. Wuyckens, G. Janssens, L. Zhao, X. Ding, M. Chocan, H. Ozan, A. Barragan, E. Sterpin, J. A. Lee, K. Souris, [PO-1982](#) *Comparison of SPArc algorithm efficiency using beamlet-based and beamlet-free optimization schemes*, Annual Meeting of the European Society for Radiotherapy and Oncology (ESTRO), 2023, Vienna (Austria)

S. Wuyckens, L. Zhao, M. Saint-Guillain, G. Janssens, A. Barragan, E. Sterpin, K. Souris, X. Ding, J. A. Lee, [PO-1988](#) *Feasibility of bi-criteria optimization applied to proton arc therapy treatment planning*, Annual Meeting of the European Society for Radiotherapy and Oncology (ESTRO), 2023, Vienna (Austria)

List of abbreviations

4D-CT	4 Dimensional computed tomography
ArcPT	Arc proton therapy
BAO	Beam angle optimization
BCO	Bi-criteria optimization
BDT	Beam delivery time
CBCT	Cone beam computed tomography
CI	Conformity index
CT	Computed tomography
CTV	Clinical target volume
D#	Dose deposited in #% of the volume of interest
Dmean	Mean dose deposited in volume of interest
Dp	Prescribed dose
DS	Double scattering
DOF	Degrees of freedom
DVH	Dose volume histogram
EEM	External energy modulator
EL	Energy layer
ELSA	Early layer and spot assignment
ELSP	Energy layer selection problem
ELST	Energy layer switching time
GPU	Graphics processing unit
GTV	Gross tumor volume
HI	Homogeneity index
HU	Hounsfield units
IBA	Ion Beam Applications s.a.
ID	Integral dose

IMPT	Intensity modulated proton therapy
IMRT	Intensity modulated radiotherapy
IP	Integer programming
ITV	Internal target volume
LET	Linear energy transfer
LON	Left optical nerve
LP	Linear programming
MC	Monte Carlo
MCO	Multi-criteria optimization
MIP	Mixed integer programming
ML	Machine learning
MRI	Magnetic resonance image
MU	Monitor unit
NTCP	Normal tissue complication probability
OAR	Organ(s) at risk
PAT	Proton arc therapy
PBS	Pencil beam scanning
PCM	Pharyngeal constrictor muscle
PET	Positron emission tomography
PMAT	Proton monoenergetic arc therapy
PT	Proton therapy
PTV	Planning target volume
QA	Quality assurance
QC	Quality control
QP	Quadratic programming
RBE	Relative biological effectiveness
RON	Right optical nerve
SD	Energy switch-down
SHArc	Spot-scanning hadron arc therapy
SMG	Submandibular gland
SOBP	Spread-out Bragg peak
SPArc	Spot-scanning proton arc therapy
SPR	Stopping power ratio
SU	Energy switch-up
TCP	Tumor control probability

TPS	Treatment planning system
V#	Volume receiving at least #% of dose specified
VMAT	Volumetric modulated arc therapy
WEPL	Water-equivalent path length
WET	Water-equivalent thickness

Contents

1	Overview	1
1.1	Research motivation	2
1.2	Main contributions	5
1.2.1	Combined dose and energy sequence optimization	5
1.2.2	A bi-criteria Pareto optimization	6
1.2.3	A geometrical energy layer pre-selection	6
1.2.4	A dose-based energy layer selection	7
1.2.5	A dose-based energy layer filtering	8
1.2.6	The NP-hardness of proton arc therapy	8
1.3	Thesis outline	9
2	Introduction	13
2.1	Radiation therapy	14
2.1.1	Fighting cancer with radiotherapy	14
2.1.2	The trade-off of radiotherapy	15
2.1.3	Dose sculpting	16
2.1.4	Radiobiology	19
2.1.5	Beam technology	20
2.2	Proton therapy treatment workflow	22
2.2.1	Imaging	22
2.2.2	Contouring and prescription	25
2.2.3	Treatment planning optimization	26
2.2.4	Treatment verification	33
2.2.5	Treatment delivery	34
2.3	Mathematical background	36
2.3.1	Terminology in optimization	36
2.3.2	Optimization problems and methods	37
2.3.3	Multi-criteria optimization	43

2.3.4	Complexity theory	44
2.4	State-Of-The-Art in proton arc therapy	48
2.4.1	Rationale for proton arc therapy	49
2.4.2	Discrete versus Dynamic PAT	50
2.4.3	PAT challenges	51
2.4.4	PAT treatment planning optimization	53
2.4.5	Clinical status	59
3	Exploring treatment planning optimization	65
3.1	Introduction	67
3.2	Background, challenge and related work	67
3.3	Proposed Benchmark	70
3.4	Notations and Problem Definition	72
3.4.1	Notations	72
3.4.2	Problem statement, constraints, and evaluation	73
3.5	State-of-the-art methods	76
3.5.1	SPArc	76
3.5.2	ELO-SPAT	76
3.6	Proposed plan optimization methods	78
3.6.1	FISTArcPT	78
3.6.2	Local Search optimizer	82
3.6.3	Mixed-Integer Programming Formulation	84
3.7	Comparison strategy	90
3.8	Results	92
3.9	Discussion	95
3.10	Conclusions and future research directions	100
4	Striking a balance	103
4.1	Introduction	105
4.2	Materials and methods	107
4.2.1	Problem statement	107
4.2.2	MIP model	108
4.2.3	MIP solver	108
4.2.4	Pareto frontier approximation	109
4.2.5	Evaluation	111
4.3	Results	113
4.3.1	Brain	113
4.3.2	Lung	114
4.3.3	Liver	116
4.4	Discussion	120
4.5	Conclusion	123

5 Streamlining the process	125
5.1 Introduction	127
5.2 Materials and Method	128
5.2.1 Energy layer pre-selection problem	128
5.2.2 Optimization algorithms	132
5.2.3 Patient data	134
5.2.4 Plan evaluation	135
5.3 Results	137
5.4 Discussion	141
5.5 Conclusion	145
6 Bridging the gap	147
6.1 Motivations	149
6.2 Methods	149
6.3 Results	151
6.4 Discussion and conclusion	154
7 Optimizing efficiency	157
7.1 Introduction	159
7.2 Material and methods	160
7.2.1 Energy layer filtering algorithms	160
7.2.2 Patient data and treatment planning	163
7.2.3 Treatment plan evaluation	165
7.3 Results	167
7.3.1 Energy layer filtering algorithms comparison	167
7.3.2 Treatment plan evaluation	172
7.4 Discussion	174
7.5 Conclusion	178
8 Unveiling complexity	179
8.1 Introduction	181
8.2 Background	182
8.3 Main results	185
8.3.1 Hardness of PAT	185
8.3.2 Application to heavy ion arc therapy	190
8.4 Concluding remarks	192
9 Conclusion and outlook	195
A Supplementary material Chapter 3	213
A.1 (Hyper-)parameters	213

A.1.1	FISTArcPT	213
A.1.2	Local Search	215
A.2	Convergence	218
A.3	Problem scaling	218
B	Supplementary material Chapter 4	223
B.1	MIP model	223
B.2	The BCO Pareto approximation algorithm	224
B.3	Benchmark	227
C	Supplementary material Chapter 5	231
C.1	Organs-at-risk dosimetry	231
	References	235

Chapter 1

Overview

1.1 Research motivation

In 2020, the global toll of cancer-related deaths was estimated to be nearly 10 million lives lost. This positioned cancer as the second leading cause of death worldwide [1]. This number is expected to further increase as the world population is growing and aging. The demographic shift not only places a heavier burden on healthcare systems but also underscores the necessity to constantly refine and advance cancer treatment techniques to accommodate the growing number of cases.

The progress made in understanding the biology of cancer and its underlying mechanisms but also technological advances have already led to important breakthroughs in treatment modalities. From surgical interventions and chemotherapy to more sophisticated approaches like targeted therapy, immunotherapy but also radiation therapy (also known as radiotherapy), the medical community has made substantial strides in its arsenal against cancer. External radiotherapy is in fact accounting for about 50% utilization in cancer management [2]. While chemotherapy or other systemic treatments have local and distant action, the largest usage of external radiotherapy is to secure loco-regional control, i.e., targeting and affecting specific body areas requiring treatment.

In the last decades, various radiation therapy techniques have emerged following the discovery of rays or fundamental particles and technological advancements. Photons, discovered by Röntgen in 1895, have been widely used in X-ray radiation therapy since the mid-20th century. Photon radiation therapy rapidly evolved from 2D treatments to “3D conformal radiotherapy” (3D-CRT) in the 1990s and then to intensity modulated radiation therapy (IMRT) in the mid-1990s. These advancements consistently aimed at enhancing dose conformity around tumors [3, 4]. In 1993, a pivotal development occurred when Mackie et al. introduced the first form of rotational IMRT [5]. This technique involve the delivery of a fan-shaped X-ray beam in a ring gantry, that rotates around the patient while they are moved through the machine, a concept known as tomotherapy. Subsequently, other variations of rotational IMRT emerged, streamlining the delivery of IMRT treatments. For example, in 1995, Yu et al. introduced “volumetric modulated arc therapy” (VMAT) [6], where the X-ray beam is continuously shaped and modulated spatially in three dimensions and typically delivered in a single rotation around the patient. Rotational IMRT offers more degrees of freedom to improve plan quality and speed-up the treatment delivery time of complex-shaped dose distributions [7]. Yet, photon therapy’s entrance and exit-dose limits the delivery of necessary curative doses for certain disease sites, as they can lead to excessive irradiation of surrounding normal tissues causing potential unacceptable toxicities

[8, 9, 10].

Protons, discovered in 1911 [11], first gained attention in medical physics due to their unique property of stopping in matter and releasing a sharp energy peak at the end of their range, namely the Bragg Peak [12]. This distinct feature offers advantages in radiotherapy, such as higher dose conformity and better sparing of healthy tissues compared to photon-based treatments. However, proton therapy is more resource-intensive and costly due to the advanced technology and equipment required and the specialized training needed for healthcare professionals. This has limited its adoption, with less than 1% of radiotherapy patients worldwide currently receiving proton therapy, despite potential benefits for a larger population [13].

In terms of technological development, proton therapy has followed a trajectory similar to that of photons as illustrated in Fig. 1.1. It started with “manual” shaping of each proton beam, referred to as the passive-scattering (PS) technique [14]. This technique utilizes scatterers, collimators and compensators to shape the beam and a spinning range modulator to vary the beam energy, resulting in a dose distribution that is a spread-out Bragg Peak (SOBP). Passive scattering proton therapy (PSPT) has progressed to intensity modulated proton therapy (IMPT) delivered through pencil beam scanning (PBS), a method akin to 3D printing that irradiates the tumor spot by spot and layer by layer given the finite range of protons in matter. Just as with photon technology, the desire to deliver proton beams while the gantry rotates became apparent, giving rise to proton arc therapy (PAT), which is the central topic of this thesis.

In PAT, there are two primary challenges that need to be addressed for its implementation into clinical practice. Firstly, the arc delivery itself requires hardware upgrades due to the unique demands of safe continuous gantry rotation synchronized with proton radiation delivery. Secondly, a large effort is invested in treatment simulation, known as treatment planning. Treatment planning is a meticulous process of designing a patient-specific radiation treatment. It aims to determine the optimal arrangement of radiation beams and their intensity to achieve clinical objectives: delivering a sufficiently high radiation dose to eradicate the tumor while minimizing irradiation of surrounding healthy tissue. This is in fact an optimization problem that can be solved with numerical solvers. The PAT treatment planning optimization is the focus of this thesis.

The rationale for PAT mirrors the argument for VMAT, capitalizing on new degrees of freedom for treatment planning optimization combined with the unique depth-dose characteristics of protons. However, these degrees of freedom present an additional challenge – they make the optimization problem more complex.

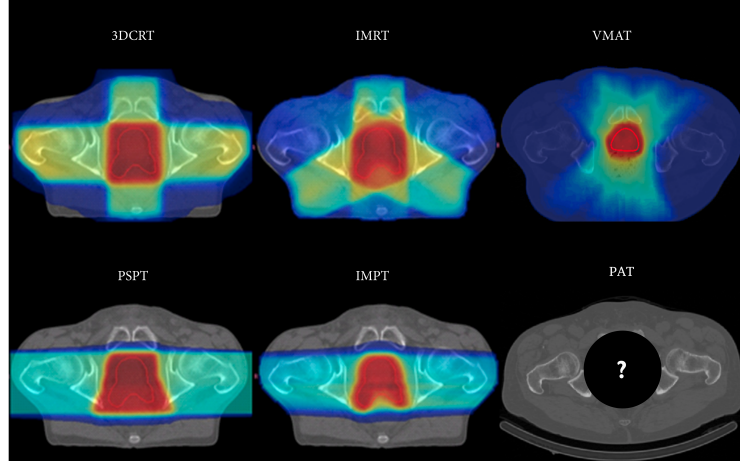


Figure 1.1: Evolution of dose distributions over the years obtained with evolving modalities over time in RT/PT on the same patient. The colored regions scale the dose linearly from low dose (blue area) to high dose (red area). 3D-CRT: 3-dimensional conformal radiotherapy; IMRT: intensity modulated radiotherapy; VMAT: volumetric modulated arc therapy; PSPT: passive scattering proton therapy; IMPT: intensity modulated proton therapy; PAT: proton arc therapy. Adapted from [15].

This challenge originates from the need for the proton delivery system to employ multiple proton energies to fully cover the target volume in-depth, therefore necessitating sequential energy switching. The non-instantaneous nature of energy switching due to technical constraints contributes to total delivery time. While this phenomenon is not a critical concern for IMPT, it has a large impact when planning PAT treatments. In IMPT, each beam is individually delivered at a fixed gantry angle, starting from the highest (distal depth) to the lowest energy (proximal depth). However, in dynamically delivered PAT treatments, it is impractical and excessive to use a decreasing energy sequence at each angle in the arc beam. Additionally, increasing the energy for each next angular window is often even more time-consuming than decreasing it due to a technical limitation coming from the magnetic field regulation. Without efforts to minimize such energy increase occurrences, it would happen between each irradiation direction, thus substantially slowing down the continuous gantry rotation. Consequently, the optimization problem must encompass minimizing delivery time alongside conventional dose-based objectives.

This thesis strives to explore the optimization challenge of PAT, with particular emphasis on delivery time. The research journey embarks with a reflection on

the optimization requirements to harness the full potential of PAT. Initially, an in-depth exploration of the optimization problem is undertaken, where multiple strategies are proposed to derive the optimal solution. The investigation then evolves, suggesting an algorithm that could provide multiple optimal solutions, granting the user the flexibility to select based on the clinical context. The search for innovative algorithms persists, driving towards more computationally efficient solvers. Some of these algorithms are assessed using a commercial treatment planning software, highlighting the potential benefit of PAT for various disease sites. The thesis culminates by presenting a theoretical perspective, emphasizing the inherent complexity of the problem that has been the focal point throughout.

1.2 Main contributions

The main objective of this thesis is to explore the treatment planning methods for PAT, in particular the underlying optimization algorithms tailored for dynamic delivery, i.e. using a continuous gantry rotation. While the concept of PAT is not novel, its clinical translation presents multiple challenges. Before it can be smoothly integrated into clinical practice, these challenges must be addressed. This work does not primarily aim to establish the clinical merits of PAT. Instead, it drives the development of various methodologies aimed at addressing the energy layer selection problem (ELSP), a central component of proton arc planning. Addressing the ELSP is a major concern because it is closely tied to the resulting plan quality, which can ultimately determine the tangible clinical benefits of PAT.

1.2.1 Combined dose and energy sequence optimization

Inspired by Gu et al [16], who introduced a method to address the ELSP integrated within the spot weight optimization framework, we saw an opportunity to explore this area further. This direction seemed promising, especially in maximizing the proton arc's potential. Our team was the first to provide a structured definition of the PAT optimization challenge. We then suggested three distinct solvers, comparing them to conventional IMPT and the state-of-the-art SPArc [17, 18] techniques. Unlike many studies that present a single plan optimization method across various disease sites, our approach offers a more thorough comparison between several existing and new PAT optimization strategies. The implementation of these solvers was carried out within the framework of an in-house open-source Treatment Planning System (TPS) known as OpenTPS¹. This project, initiated and developed by Kevin Souris, received important contributions from the author of this thesis, especially in the optimization modules. Any bench-

¹<http://opentps.org/>

marking conducted for this study has been made accessible online for peers who might want to evaluate their optimization algorithms.

This work led to a publication, which is integrated into Chapter 3. Note that in this chapter, we refer to proton arc therapy as arc proton therapy (ArcPT). As OpenTPS evolved with valuable input from its core developers, it became publicly available in November 2022. An initial draft detailing the tool was shared on ArXiv [19].

1.2.2 A bi-criteria Pareto optimization

While minimizing the delivery time for proton arc therapy already presents a challenge, it is uncertain to what extent this can be balanced with dose fidelity—a competing objective. For instance, a highly delivery-efficient PAT plan may compromise robust target coverage if too few energy layers are used. The difficulties lie therefore in the tuning of the objective weighting parameters in the cost function. Radiation therapy optimization, inherently multi-objective, often requires meticulous manual adjustment of each objective weight until a satisfactory plan that meets clinical goals is found. This can be particularly demanding for cases with complex geometry, where the objectives of target coverage and organs-at-risk (OAR) sparing might already conflict. Several multi-criteria optimization (MCO) strategies have been proposed, aiming to provide planners with a spectrum of Pareto optimal solutions instead of iterative tuning of the weights. This allows the choice of the most appropriate compromise based on the clinical context [20, 21, 22, 23, 24]. Drawing on Craft et al’s method [25], we introduce the first bi-objective optimization study for PAT. This examines the balance between plan quality and delivery time, ultimately enabling clinicians to make an informed decision. Furthermore, the algorithm developed to generate a pool of Pareto optimal plans, was implemented within OpenTPS. This work represents a milestone for PAT planning, laying the groundwork for advanced MCO that may eventually incorporate an actual machine-specific delivery sequence model.

This research has been published and can be found in Chapter 4 of this thesis.

1.2.3 A geometrical energy layer pre-selection

Up until now, all the strategies developed and tested for proton arc planning in this thesis involved optimizing both the energy sequence and spot weights simultaneously. However, this approach substantially slowed down the planning phase, sometimes taking several hours for a simple case, and demanded heavy memory resources, occasionally reaching dozens of gigabytes. It required pre-computing the dose contribution from every possible spot, energy layer, and irradiation an-

gle, and then optimizing the energy sequence along with the massive amount of spot weights candidates. This made it challenging to use robust optimization in OpenTPS, which is primarily written in the Python programming language. To address this issue, we decided to re-evaluate our strategy and explore heuristic methods instead. Our approach involves breaking the problem into two steps. First, we pre-optimize the energy sequence, ensuring that each irradiation angle uses a single energy. The optimization goal is to achieve good geometrical target coverage while penalizing angles that intersect with selected OAR. Additionally, we incorporate a direct approximation of the total delivery time into the objective function. The estimated irradiation time takes into account the gantry's rotation speed, the ELST, and energy layer scanning time. After this pre-optimization step, we only need to optimize the spot weights from the list of pre-selected energies, which could be done using a conventional optimizer from any TPS. Although this methodology is not entirely new and was initiated by the IBA proton R&D team, we decided to further explore their solution by implementing it into our own in-house TPS, OpenTPS. We collaborated closely with RaySearch Laboratories (Stockholm, Sweden), one of the major leader in the TPS industry, to be able to import the energy layer selection we made in-house, into their TPS, RayStation. We then conducted a comparative analysis with their energy layer pre-selection algorithm prototype, which is also heuristic-based.

Currently, we are in the process of preparing a detailed manuscript describing this method, which we plan to submit to a peer-reviewed journal in the near future. The current version of the manuscript is included in Chapter 5.

1.2.4 A dose-based energy layer selection

Splitting the process into two stages, optimizing the energy sequence and then optimizing the dose, appears to be computationally efficient. However, this method may sometimes lack flexibility during the second step because it does not allow for adjustments to the energy sequence when, for instance, target coverage falls short [26]. To address this issue, we collaborated with RaySearch to explore possible improvements of their prototype dynamic arc optimization solution [26], with potential application for other proton arc optimization algorithms. The idea was to expand the set of possible energy layers at each gantry angle during pre-selection and iteratively filter energy layers during the main optimization until a single energy per angle was achieved to reach high delivery efficiency. This process involved using a penalty function and optimization breaks to eliminate the worse-performing energy layers. However, the results showed only marginal improvements compared to the baseline arc plans. Despite the efforts invested in this direction, we did not proceed further with this particular study. A brief sum-

mary of the method developed and the results obtained can be found in Chapter 6.

1.2.5 A dose-based energy layer filtering

Although the results from the previous contribution were not very conclusive, we decided to revisit the energy layer selection concept from a new perspective. We investigated a dose-based energy layer filtering approach, applied as a post-processing step after spot weight optimization, thereby accelerating proton arc delivery. We investigated three distinct filtering methods for this purpose, and they proved to be efficient. These methods were directly integrated into a research version of RayStation 2023B (RaySearch Laboratories AB, Stockholm, Sweden). They were also implemented in OpenTPS for open-source usage. Given the challenge of determining the optimal energy layer selection, these methods serve as post-processing tools for arc optimization, allowing for the removal of energy layers that contribute minimally to treatment and/or negatively affect delivery time. They become especially relevant when used in conjunction with energy layer pre-selection heuristics. To sum up, pre-selection is used to speed-up the optimization and combined to post-filtering to accelerate the delivery of treatment.

The final study has been submitted for publication to Medical Physics journal. Chapter 7 presents the submitted manuscript.

1.2.6 The NP-hardness of proton arc therapy

In the concluding study of this thesis, we returned to fundamental properties, building upon the groundwork laid in earlier chapters. We revisited the PAT problem, which we established as non-convex when addressed as a continuous optimization challenge and raised questions about its intrinsic theoretical complexity. While an initial attempt was made to prove the hardness of the PAT problem in a specific case in the initial publication of this thesis, we aim to revisit and rigorously address this question in Chapter 8. We provide a comprehensive and formal proof that establishes the NP-hardness of the PAT problem.

These findings hold notable implications. Firstly, finding an optimal solution to the PAT problem demands substantial computational resources, often requiring extensive exploration of the solution space. Moreover, as of today, there is not a universally efficient algorithm capable of solving this category of problem for all instances. In practice, obtaining the global optimum may pose difficulties, potentially leading to the acceptance of near-optimal solutions (if any can be found). Instead, we may need to rely on specialized heuristics, meta-heuristics, or approximation methods to achieve practical solutions within reasonable timeframes.

1.3 Thesis outline

The present chapter (**Chapter 1**) outlines the motivations behind this research and provides an overview of the thesis contributions to this evolving field.

Chapter 2 introduces the background for comprehending the research conducted within this thesis. Given the interdisciplinary nature of the work, we start by discussing the medical physics elements relevant to the topic. Subsequently, we transition gradually towards delving into the underlying mathematical theories that form the basis of this study. Following that, we present the current state-of-the-art in PAT.

Chapters 3 to 8 are constructed around the five main author's contributions and organized as follows:

Chapter 3: Exploring treatment planning in proton arc therapy

This chapter marks the author's initial publication and groundwork, titled "Treatment planning in arc proton therapy: Comparison of several optimization problem statements and their corresponding solvers". It establishes the mathematical model behind the optimization problem for treatment planning in PAT and offers strategies on how to tackle it using different optimization schemes.

Chapter 4: Striking a balance

In this chapter, we delve deeper into the balance between irradiation time and dosimetric goals in PAT. Venturing into the realm of multi-criteria optimization, we introduce an algorithm capable of generating a collection of Pareto-optimal solutions for these two objectives. This strategy empowers planners to navigate through options, enabling them to choose the optimal trade-off in line with their experience and clinical constraints. This study led to the second paper "Bi-criteria Pareto optimization to balance irradiation time and dosimetric objectives in proton arc therapy". This achievement is the result of a collaboration with the Proton Therapy Center of Beaumont Health (Royal Oak, MI, USA) during a 6-month secondment period.

Chapter 5: Streamlining the process

Chapter 5 presents an alternative approach to address the optimization problem in PAT treatment planning. This method involves separating the problem into two distinct steps: first, solving the energy layer selection, and then addressing the spot weight optimization problem. While this approach may sacrifice the dose-based energy layer selection, it was chosen for practical reasons, mainly the considerable acceleration and efficiency it brings to the treatment planning process. This

methodology stems from a closely-knit collaboration with Ion Beam Applications S.A. (IBA). The final manuscript is currently being prepared for journal submission.

Chapter 6: Bridging the gap

This chapter focuses on the integration of a dose-based energy layer selection into a commercial treatment planning system. We bring together the insights gained from previous chapters and collaborate with RaySearch Laboratories AB (Stockholm, Sweden). Our objective is to blend geometric energy layer pre-selection with dose-based objectives during the spot weight optimization process to enhance the overall quality of proton arc plans. This approach involves initially expanding the set of energy layers during pre-selection and then retaining only the highest-weighted ones to “sparsify” the energy sequence and keep reasonable delivery times. This research was conducted in the context of a 3-month research stay at RaySearch Laboratories.

Chapter 7: Optimizing efficiency

In this chapter, we combine the fast energy layer pre-selection implemented in a research version of RayStation with post-filtering strategies to optimize both computational and delivery efficiency. A range of methods have been explored to identify and filter energy layers that have a minimal impact on dosimetry. This approach aims to preserve high plan quality, expedite treatment delivery, and complement the optimization speed-up achieved through the use of energy layer pre-selection methods. This work was carried out in collaboration with RaySearch Laboratories. The outcomes of this study are encapsulated in a manuscript titled “Faster proton arc optimization and delivery through energy layer pre-selection and post-filtering”. This paper currently awaits revision in the journal *Medical Physics*.

Chapter 8: Unveiling complexity

This chapter delivers a mathematical proof of the NP-Hardness of PAT, effectively concluding the initial incomplete proof introduced in the first paper (Chapter 3, section 3.6.3). This achievement underscores the inherent computational complexity of the problem, and it plays a pivotal role in directing research endeavors, fostering the design of efficient algorithms, and establishing a theoretical foundation for PAT problem-solving approaches. This work is the result of collaboration with Viktor Wase (RaySearch Laboratories) and Michael Saint-Guillain (CEO & co-founder of Rombio, researcher in ICTeam, UCLouvain), whose collective expertise substantially contributed to this work.

The concluding chapter (**Chapter 9**) reviews and discusses the research pre-

sented in the thesis. It offers insights into the future prospects of proton arc therapy and includes the author's personal perspective on the subject.

Chapter 2

Introduction

2.1 Radiation therapy

This section offers a succinct and general introduction to radiation therapy, its purpose, and the methods used in its application. Experienced readers are invited to start with Section 2.3 where a mathematical description of the concepts used throughout the thesis are provided.

2.1.1 Fighting cancer with radiotherapy

Before we delve into cancer treatment, it is essential to understand the adversary we are up against.

Ordinarily, cell division in the human body is regulated by a complex system of biological signals. Yet, when DNA mutations occur, they can disrupt this division control mechanism. These mutations are frequently caused by various factors like environmental factors, viral infections, errors during DNA replication, and more. While most of these mutations are not inherently harmful, in certain cases, they can lead to the development of cancerous cells. These cells, marked by DNA mutations, proliferate uncontrollably and amass to form tumors, setting the stage for the battle against cancer.

High-energy radiation, often referred to as “ionizing radiation”, can damage the genetic material within living cells through fundamental interactions by directly or indirectly stripping atoms or molecules of their electrons [27]. As soon as genome damage is detected, biochemical processes are activated to expedite repairs. Three main response scenarios emerge. For sub-lethal damage, cells can accurately repair themselves without adverse consequences. Sometimes, cells inadequately repair, surviving with genetic alterations. In cases of extensive DNA damage, ineffective repair can lead to cell inactivation [28]. These ionizing radiations can be photons as well as subatomic particles like electrons, protons, neutrons or even atoms such as light or heavy ions. Unregulated exposure to ionizing radiation can yield serious consequences. Its impact includes adverse health effects such as the development of acute or even late radiation syndromes, often manifesting years or even decades after exposure. For example, ionizing radiation played a major role in contributing to chronic myelogenous leukemia, as demonstrated in a study involving survivors of the nuclear bombings in Hiroshima and Nagasaki [29].

However ionizing radiation can also be used for medical applications to accurately target and eliminate cancer cells, while following safety protocols and using the right equipment. Tumor cells repair DNA damage less efficiently than normal cells and hence tumors can often be eliminated without damage to normal tissues if the dose to sensitive normal cells is not excessive (see section 2.1.2 below). This

is the core idea behind radiation therapy (RT). For certain types of cancer that are confined to a particular area of the body, RT can actually cure the disease. Early stage cancers are therefore usually treated either with surgery or radiotherapy providing a high loco-regional control probability of the disease. RT may also be combined to other cancer treatments, e.g. as part of adjuvant therapy after a surgery to prevent the cancer from coming back. Moreover, RT may also complement chemotherapy for locally-advanced cancers, and, in that case be used before, during, and/or after chemotherapy for certain types of cancer that respond well [30, 31, 32]. Note that about half of RT cases are for palliation in the treatment for pain or other quality of life improvements in the metastatic stage of cancer [33].

Radiotherapy is primarily administered through external beams generated by medical accelerators or by a fixed array of radioactive sources that produce high-energy radiation. An alternative, but much less widespread approach, involves using sealed or unsealed radioactive sources placed directly within the tumor volume or next to it.

Ionizing radiation that is absorbed by an object or person is termed the *absorbed dose*, a unit defined within the International System of Units (SI). This measure is denoted in Gray (Gy), which quantifies the energy in joules (J) deposited per unit of mass (Kg). In RT, the oncologist will therefore set a prescription dose that the tumor must receive in order to cure the disease.

2.1.2 The trade-off of radiotherapy

When we treat a tumor inside a patient using external radiotherapy, the radiation we use also affects the healthy tissues around the tumor. For instance, in conventional radiotherapy with X-rays, the radiation beam goes through the patient's body, affecting all tissues present on the beam path. As a result, the DNA in healthy tissues can also get damaged. However, healthy cells can repair themselves better than corresponding tumor cells whose mechanism are often altered. This phenomenon creates what is known as a "therapeutic window". The challenge of RT for cancer treatment is therefore to give enough radiation to eliminate all the cancer cells and their offspring while causing as little harm as possible to the patient with the lowest dose achievable to healthy tissues. This is equivalent to achieving a high probability of local tumor control (tumor control probability, TCP) at a low risk of normal tissue complications (normal tissue complication probability, NTCP). These probabilities, TCP and NTCP, are derived from in-vitro survival cell rate experiments and are often depicted as curves on a graph against the prescribed dose, as illustrated in Figure 2.1. These curves display a characteristic sigmoid shape, indicating that both very low and very high

doses have a limited impact on dose response. However, the dose response is highly pronounced at intermediate doses. The distance between these two curves illustrates the therapeutic window mentioned earlier.

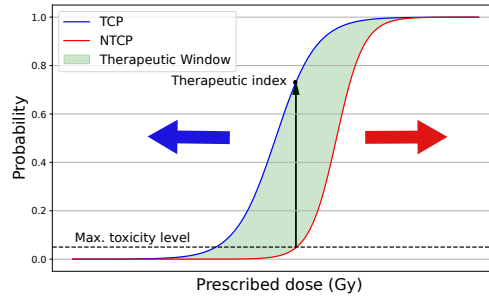


Figure 2.1: Dose-response relationship figuring the deal of RT: delivering enough dose to the tumor volume to ensure high TCP while remaining below toxicity level defined by the NTCP.

Several approaches have been developed to modify the TCP and NTCP curves with the objective to increase the therapeutic window size. This topic will be addressed in the next subsections.

2.1.3 Dose sculpting

The challenge of RT is often analogized to a “ballistic” problem due to its inherent nature. Over the past two decades, technological progress has been geared towards refining the spatial differentiation between normal tissue and tumors to prevent radiation-induced toxicities. One important advancement has been the introduction of multi-leaf collimators, which shape the photon beam by adjusting the positions of individual leaves in the beam’s path. Additionally, it was recognized that using multiple beam directions that converge at the tumor site can improve the sculpting of the radiation dose [34]. Moreover, the implementation of intensity modulation in the 1990s is progressively replacing all classical conformation systems that make use of uniform dose fields. This shift aims to achieve even greater conformity to the tumor’s shape.

When a tumor is located near critical organs, conventional photon-based radiotherapy techniques can sometimes struggle to adhere to dose constraints. This occurs because when a megavoltage photon beam penetrates the body to reach a deep-seated tumor, it releases the most energy close to the surface and subsequently attenuates exponentially due to the photoelectric and the Compton interactions [35]. Photons hence deposit inevitably doses both upstream and downstream of the tumor location.

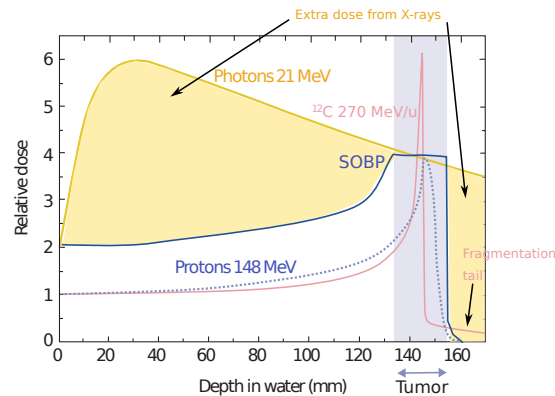


Figure 2.2: Depth-dose curves in water generated by photons (yellow), protons (blue) and carbon ions (red). A spread-out Bragg peak (SOBP), built from the combination of several protons Bragg Peak, is also represented, effectively showing the tumor coverage in-depth with little dose present beyond the distal end of the peak. Adapted from [36].

Charged particles, such as protons, interact differently. They directly collide with other charged particles and nuclei. This interaction pattern causes protons to lose energy in a “quasi-continuous” fashion along their radiation path until they approach the end of their range reaching their maximum dose deposition in what is known as the Bragg Peak, first described in 1946 [37]. Figure 2.2 illustrates the distinct depth-dose curves for photons and protons. By modulating proton energies, it becomes possible to accurately target a specific location within the body, delivering a high dose to it and then stopping with a small further dose distribution downstream of the tumor, as shown in Fig. 2.3. The ability to control the depth of penetration by adjusting the proton energy offers a distinct rationale for proton therapy (PT). It is more likely to spare surrounding healthy tissues while effectively treating the tumor. Since the Bragg peak is narrow, multiple Bragg Peaks with different intensity and energy need to be superimposed in order to cover the full tumor depth and produce the so-called spread-out Bragg Peak (SOBP).

The evolution of proton beam shaping technology has involved improving the delivery systems to enable even finer control over the dose distribution within the tumor. Initially, SOBP were mostly delivered using passively scattered (PS) techniques, involving physical devices like scatterers and compensators to shape the beam. However, technology has advanced and tends to be replaced by pencil beam scanning (PBS). PBS is a more accurate delivery system where narrow proton beams, called “spots”, are scanned across the target volume at different

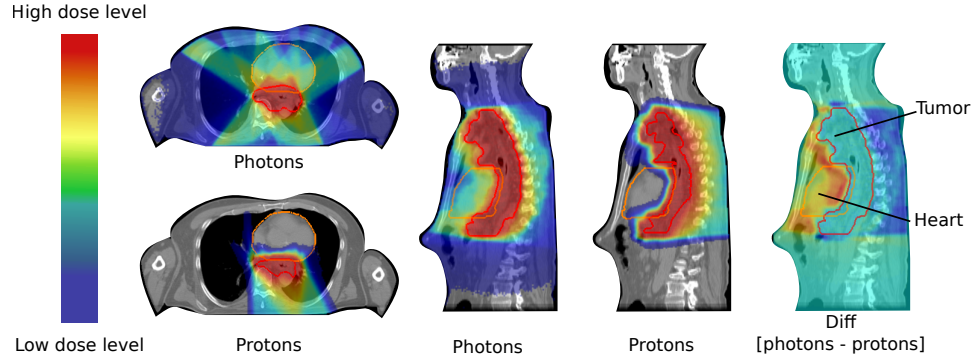


Figure 2.3: Dose distribution comparison between a photon plan and a proton plan for esophagus cancer. While many photons beams are required for dose sculpting, only 2 proton beams are needed. This attest for the high conformity of protons which consequently better spare the heart in this case.

depths, called “energy layers”. It then “paints” the dose sequentially spot by spot, and layer by layer from each irradiation direction [38, 39].

Heavier ions, such as carbon ions, are also represented in Fig. 2.2. The term *heavy ions* is used here for ions heavier than helium ions. Due to their larger mass and charge, they tend to interact more densely with the atoms they encounter along their path. Because of the high density ionization tracks of secondary charged particles that heavy ions leave in matter, they are usually termed as high-LET (linear energy transfer) radiation in contrast to low-LET radiation such as electrons, photons and protons. The LET radiation quality of a beam relates to the biological effect of the ionizing radiation. As for protons, heavy ions stop in matter but they result in a sharper Bragg Peak with a rapid dose fall-off. We also observe a more prominent tail in the depth dose distribution that extends beyond the Bragg peak. This property comes from the increase of nuclear fragmentation processes (secondary low-energy ions). Carbon ions were mostly introduced in radiation therapy for their high biological effect. As a result, they inflict more severe damage to DNA which was proven useful for radioresistant tumors [40, 41]. However, their use should be approached cautiously, as they often lead to substantial damage in normal tissues as well. This topic will be addressed in the next subsection.

Unfortunately, the sharp distal dose gradients inherent to protons and heavy ions also present a major challenge. They can be influenced by daily variations in a patient’s anatomy, leading to uncertainties in how far the particles can travel, a phenomenon commonly referred to as “range uncertainty” [42]. This can cause unintended problems, like exposing healthy tissue to too much radiation or not

giving enough dose to the intended target volume. This dilemma emphasizes the need for a robust treatment planning approach that accounts for uncertainties in particle therapy. This will be further discussed in Section 2.2.3.

2.1.4 Radiobiology

In addition to approaching radiation therapy as a ballistic problem, there are complementary strategies to improve its efficiency and therapeutic ratio, that relate to radiobiology. One such strategy involves the use of fractionation, a technique that divides the total prescribed radiation dose into smaller fractions administered over multiple treatment sessions. This approach takes advantage of the differences in cellular responses to radiation between normal tissues and cancer cells. While cancer cells are often less capable of repairing radiation-induced damage, normal tissues have the opportunity to repair and recover better between fractions, minimizing the risk of severe side effects. Moreover, oxygen levels in tissues substantially influence their sensitivity to ionizing radiation. This is because ionizing radiation interacts with oxygen molecules, leading to the release of free radicals. These free radicals, which constitute the indirect mode of radiation-induced damage, contribute to approximately two-thirds of the overall damage caused by ionizing radiation [43]. Consequently, after irradiation, the initial cells killed by radiation are often the peripheral well-oxygenated ones. Fractionation is therefore employed to address this issue as it allows for reoxygenation between treatment sessions, making the previously resistant tumor layer vulnerable to subsequent radiation therapy treatment fraction. Fractionation relies on other important factors to justify its usage, but this goes beyond the scope of this thesis. Interested readers can refer to the appropriate literature for more in-depth information [36]. Fractionation schemes often involve 5 fractions/week between 1.8 to 6 Gy per fraction delivered over the treatment course extending from 1-7 weeks in duration.

We saw earlier that many types of particles can be used in radiation therapy. However, different particles do not cause the same damage to DNA for the same absorbed dose. To account for these differences in effectiveness of particle radiation (i.e. cell killing), the concept of relative biological effectiveness (RBE) is employed. RBE is defined as the ratio of the dose of reference (often conventional photon radiation) to the dose of the investigated radiation that yields an equivalent biological effect. Protons, for instance, are commonly estimated to be approximately 10% more effective than conventional high-energy photon radiotherapy (uniform $RBE=1.1$) [44]. However, recent experiments [45] and clinical studies [46] indicate that the RBE value for protons is not uniform. Instead, it may increase critically towards the distal dose fall-off of individual Bragg Peaks as LET

increases [47]. Some proton treatments involve SOBP modulated fields, which are a summation of multiple Bragg Peaks of different energy, RBE variations within the SOBP may be expected. This variability could result in a higher-than-expected biologically equivalent dose deposition, which can be problematic in normal tissues located distal to the tumor [48]. To address this aspect in clinical practice, a commonly employed strategy is to carefully select beam angles, ensuring that no more than one beam stops in a sensitive OAR [49]. On the other hand, carbon ion radiation exhibits a RBE much greater than 1 across its entire tissue range which increases with depth [36]. As a result, they lead to greater DNA damage and cell death, providing them an important differential advantage in targeting tumors versus sparing healthy tissue, in particular for radioresistant hypoxic tumors.

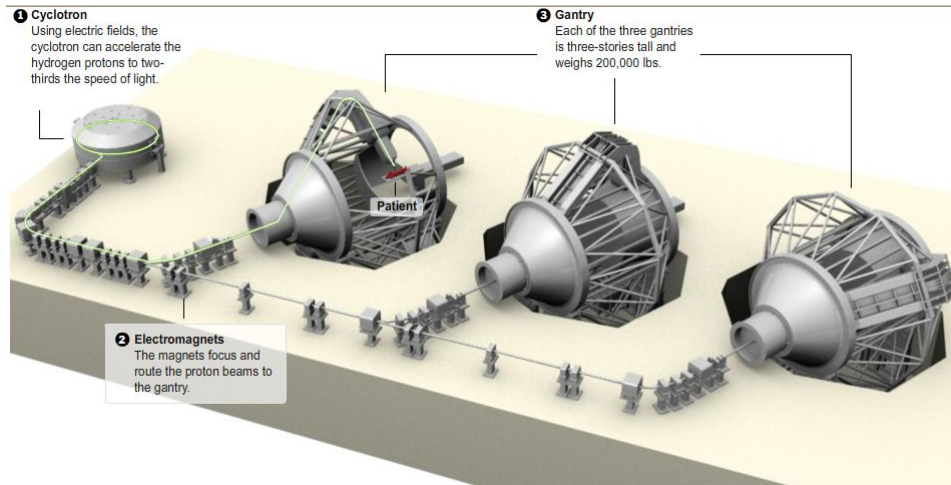
Methods like tumor radiosensitization that improve oxygenation [50, 51] or that employ drugs generally can help to increase loco-regional control of the tumor in concomitant chemotherapy or targeted therapy [52, 53, 54]. Hypoxic regions can also be targeted preferentially. Imaging modalities quantifying hypoxia aid in localizing these areas [55]. Additionally, the development of radioprotector drugs [56] and the emergence of FLASH therapy [57], which involves administering treatment at very high dose rates, demonstrate radio-protective effects (i.e. reduces the normal tissue toxicities), contributing to an improved therapeutic ratio.

2.1.5 Beam technology

In conventional RT, the generation of photon beams is typically done through linear accelerators (linacs). These linacs are situated within the treatment room, often on a fixed beamline or gantry (i.e. a rotating structure that moves the radiation device around the patient). The beam generation process involves the release of electrons from a cathode, which are subsequently accelerated using an electric field toward a target with a high atomic number, such as tungsten or gold. Upon collision, these electrons produce high-energy X-ray photons via bremsstrahlung radiation. The energy spectrum of the produced photon beam is usually fixed for a specific treatment machine. Once generated, the photon beam is ready to be shaped to conform to the tumor volume.

In contrast, the generation of a proton beam necessitates more complex and financially demanding infrastructure, as illustrated in Fig. 2.4. This is achieved using a particle accelerator, either a cyclotron (fixed energy) or, less commonly, a synchrotron (variable energy) [14]. The procedure starts with a hydrogen gas source, and due to the hydrogen atom's composition—a singular proton and electron—the protons can be extracted. These protons are then accelerated within the particle accelerator using an electric field. The accelerator provides the necessary energy to boost the protons to very high speeds, to up to $2/3$ the speed

of light. After the acceleration phase in cyclotron-based facilities, protons are extracted when they have reached their maximal energy. For therapeutic purpose, maximum energies are in the range of 230-250 Megaelectron-volts (MeV) with ranges in water of approximately 33-37 cm respectively. Post-acceleration, an energy layer selection system (ESS) is integrated at the cyclotron's exit, reducing the proton energy to the intended depth in the patient. The ESS consists of an energy degrader of variable thickness, a magnetic analyzer and slits defining the energy width. However, the process is yet to be completed, as the beam must be channeled to the treatment rooms. This transportation is done by a sequence of focusing quadrupole and bending dipole magnets. Upon arrival at one treatment room, the proton beam is typically bent around a cumulative 90° bend around a rotating gantry, also using a set of quadrupole and dipole magnets. The proton therapy gantry is a massive rotating structure, often weighing between 100 to 200 tons and spanning up to 10 meters in diameter, though the patient does not directly see it as it is concealed within the ceiling and floor. At the end of the rotating gantry arm, there is the nozzle that directs and shapes the proton beam onto the patient. Fixed beamline proton therapy (i.e., no gantry) is often used for eye treatments and is starting to be used for other sites by rotating the patient instead of the beam (this paradigm is discussed more later).



Sources: University of Florida Proton Therapy Institute

Vu Nguyen / The New York Times

Figure 2.4: Multi-room proton therapy center at the University of Florida Proton Therapy Institute.

The beam is delivered through one of two main clinical delivery methods: PS or PBS [14]. In PS mode, the narrow beam incident on the treatment nozzle must be shaped to cover the tumor volume. Depending on the initial field size, single or

double scattering (DS) foils are used to broaden the beam. A range modulation wheel adjusts the depth of the proton beam. The field is then shaped using carved hardware devices such as the compensator (distal part of dose distributions) and apertures (lateral variation). In modern PT (PBS delivery), narrow beams, also referred to as spots, are focused and steered magnetically, allowing their Bragg peak to occur at a specific location in space. A dynamic adjustment of each spot contribution to the total dose is then feasible and constitutes the plan optimization step which will be addressed in Section 2.2.3.

The generation of a heavy ion beam introduces further complexities, necessitating bigger and more expensive facilities capable of accelerating the heavy ions to the desired energy and bending them around the necessary beam path to the patient.

The disparities in technological requirements for photon RT and PT, leading to differential cost implications, largely contribute to the limited adoption of proton or heavy ion therapy across all radiotherapy centers [58, 59].

2.2 Proton therapy treatment workflow

This section describes the clinical workflow followed in a PT treatment. While many concepts apply broadly to all external radiotherapy modalities, our emphasis is on PBS-based proton therapy, in alignment with the thesis topic. To help guide our explanations in the following sections, a schematic summary is presented in Fig. 2.5.

2.2.1 Imaging

The proton therapy workflow starts with the acquisition of high-quality images, referred to as pre-treatment imaging. To effectively provide a detailed volumetric representation of the patient, three-dimensional (3D) imaging is employed, combining together 2D cross-sectional images or “slices”. These images are almost always obtained through computed tomography (CT) scans, although magnetic resonance imaging (MRI) or positron emission tomography (PET) scans may also be used to enhance soft tissue contrast, generate reliable pseudo CT images from it [60] or provide physiological complementary information, respectively [61].

Pre-treatment images are usually stored in DICOM format, which stands for Digital Imaging and Communications in Medicine. DICOM is a standardized protocol used globally for storing, transmitting, and handling medical imaging information. It ensures interoperability between different medical systems. Although DICOM

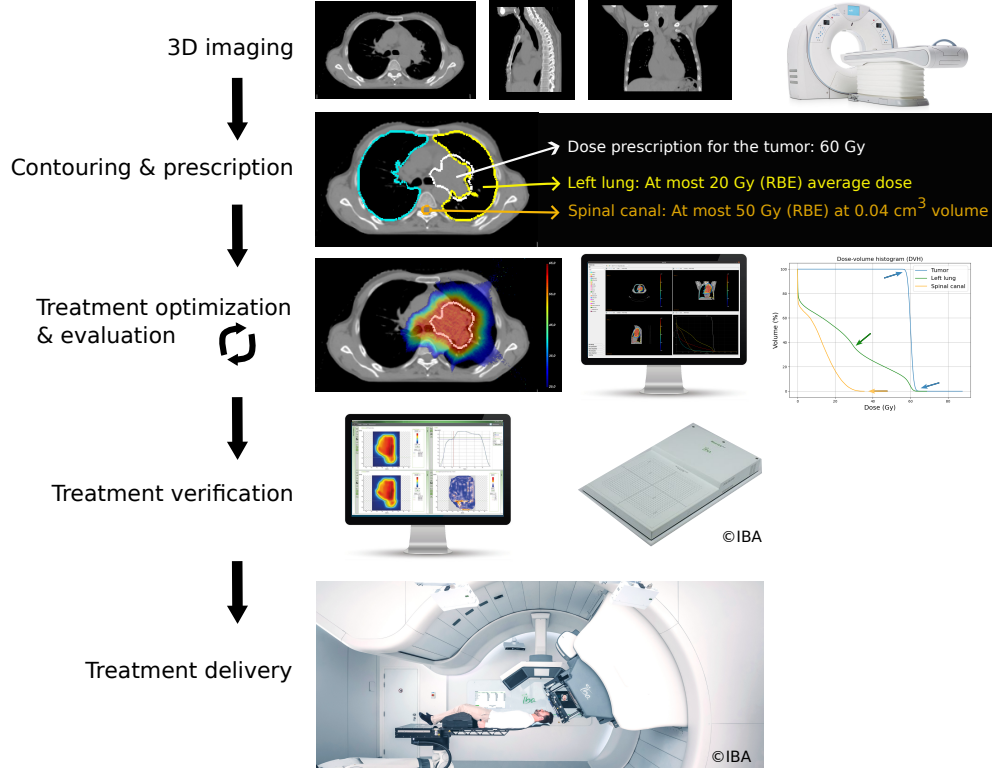


Figure 2.5: Key components of a proton therapy treatment workflow

is not restricted to images, it also encompasses patient data, metadata, and other relevant information, making it integral to modern medical imaging practices.

Coming back to CT imaging, in addition to the anatomical mapping, it also provides physical information necessary to perform proton dose calculation. CT numbers reflect the X-ray attenuation properties of tissues, that is proportional to their electron density information and approximately logarithmically dependent on atomic number. It is important for accurate dose calculations, as the trajectory, range, and energy deposition of protons are influenced by the atomic properties of tissues. Each voxel (3D equivalent of a pixel) in a CT image is therefore assigned a CT number value representing X-ray attenuation [62]. The CT number is generally expressed in Hounsfield Units (HU), which are defined as:

$$HU = 1000 \cdot \frac{\mu_{\text{tissue}} - \mu_{\text{water}}}{\mu_{\text{water}} - \mu_{\text{air}}} \quad (2.1)$$

where, μ_{tissue} denotes the total linear attenuation coefficient of the tissue at a specific x-ray energy. On this scale, water at standard conditions is defined to have a value of 0 HU, while air is approximately -1000 HU. Bone, which is denser and attenuates more X-rays than soft tissues, has a positive HU value, often exceeding +1000 HU. The HU values for bone strongly depend on the peak kilovoltage (kVp) and the filtration of the CT x-ray beam.

The pre-treatment images are obtained in the same conditions as for the treatment. They form the basis for creating personalized treatment plans, guiding the choice of beam angles and the optimization of the proton pencil beam intensities. Consequently, the acquired images must be of high quality to guarantee an accurate treatment. Immobilization devices are frequently used to mitigate patient movement and enhance the reproducibility of patient positioning, particularly given that most treatments are fractionated.

In certain cases, 4D-CT scans are performed, introducing the fourth dimension of time into traditional 3D-CT scans [63]. The primary objective of 4D-CT imaging is to capture the movement of structures, commonly observed in thoracic and abdominal regions due to breathing. Once captured, the images are sorted and grouped based on the breathing phase using a respiratory signal. Subsequently, they are used in treatment planning, aiding in estimating tumor positions and organ movement throughout the breathing cycle. This, in turn, minimizes radiation exposure to healthy organs like the heart and lungs.

Additionally, throughout the treatment course, a preliminary alignment is done prior to each session by means of lower-quality imaging (in comparison to CT imaging) integrated in the treatment room. This aligns the patient with the simulated treatment position based on the pre-treatment images. However, anatomical or physiological changes can be substantial during the course of treatment, such as patient weight loss, tumor shrinkage, or internal organ movement. Differences between the initial treatment plan and the actual patient anatomy can lead to an important divergence between the planned and delivered radiation doses and constitutes a major limitation within current clinical practice in PT [42]. To address this challenge, adaptive radiation therapy strategies have emerged, aiming to dynamically modify treatment plans based on periodic imaging, often referred to as “replanning CTs” or other adaptive strategies during the treatment course [64]. Nevertheless, adaptive treatments are beyond the scope of this thesis and will not be described further.

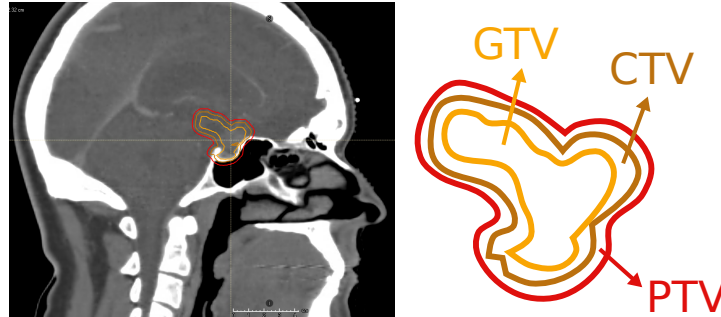


Figure 2.6: Target volume delineation: Example for craniopharyngioma.

2.2.2 Contouring and prescription

After image acquisition, the images are imported into a treatment planning system (TPS). Examples of commercial TPS for protons include Varian's Eclipse™, RaySearch Laboratories' RayStation™ and Elekta's Monaco™. In the TPS, contour delineation is executed on each slice of the CT scan, also called the planning CT in this context. This process involves contouring the target tumor volume, adjacent organs at risk (OAR), and other anatomical structures on the images. Contouring is an essential aspect in treatment planning to ensure that the beam size, number, trajectory and weighting are optimised in relation to these structures.

Depicted in Fig. 2.6, physicians delineate the visible tumor volume, known as the gross tumor volume (GTV). Next, a margin is introduced to obtain the clinical target volume (CTV), which defines the extent of microscopic tumor spread or suspected infiltrated regions by cancer cells. To achieve proper CTV coverage with the prescribed dose, an additional safety margin is routinely drawn, creating the planning target volume (PTV). This accounts for treatment uncertainties, like patient setup errors and internal organ motion. This approach minimizes the risk of underdosing the CTV due to these uncertainties within a certain level of confidence. PTV margin calculation typically follows “recipes” [65] based on population studies and setup uncertainties. As a consequence, in proton therapy, PTV margins are used more for conceptual purposes, in particular due to range uncertainties, which are not accounted for in the PTV margin definition. Nonetheless, PTV margins remain a standard due to tool limitations in commercial TPS. A good practice is often to adjust the PTV margin based on clinical expertise rather than formulas. In any case, the PT community is moving slowly but surely towards robust optimization to address these challenges. Robust optimization will be described in the next section.

Once all regions-of-interest (ROI) are contoured, the radiation oncologist prescribes the dose to the tumor, considering factors like tumor type, location, and treatment objectives and typically based on clinical experience and clinical trials. For example, a prescription could be to deliver 60 Gy in 30 fractions to at least 95% of the PTV volume (conventionally written as $D_{95\%} > 60 \text{ Gy}$). Simultaneously, dose constraints are defined for nearby OAR to limit radiation exposure and potential side effects. The dose constraints are typically set on the (near)-maximum or the mean dose to the OAR. Spinal cords, for instance, are very sensitive to the maximum dose and should not receive more than 50 Gy at 0.04 cm³ of its volume (conventionally written as $V_{50 \text{ Gy}} < 0.04 \text{ cc}$). This step establishes the clinical goals, guiding the subsequent plan optimization carried out by the physicist or dosimetrist.

2.2.3 Treatment planning optimization

The images and clinical goals at hand, the next step is to create and optimize the treatment plan in the TPS. The treatment plan is a list of machine parameters defining how the treatment must be delivered. Plan optimization refers to the process of fine-tuning the parameters of the treatment plan to achieve the best possible outcome for a patient. The ultimate aim is to deliver the prescribed dose to the tumor while minimizing the dose to surrounding healthy tissues and remaining below OAR dose constraints whenever possible.

Plan design

The design of the treatment plan typically includes beam configuration as well as the prescription settings and treatment objectives. Each beam must be defined with a gantry and a couch angle. The gantry angle refers to the rotation of the radiation device around the patient, whereas the couch angle relates to the orientation of the treatment table. These angles are primarily chosen based on clinical experience, but automated methods for beam angle selection is an active research area [66, 67]. For tumors located near the patient's body surface, a range shifter can be used to decrease the proton energy beyond the minimum possible value (usually around 60 MeV, corresponding to a depth of 3 cm in water) by physically positioning it at the nozzle exit. In modern PT, using PBS delivery, a specific set of spots must be defined for each beam, known as the spot placement. This denotes the positioning of individual narrow proton pencil beams to target the tumor. For every irradiation angle, a 3D grid of spots is therefore created around the beam's isocenter. Only the spots whose trajectory reaches the target are retained. Further transportation of the remaining spots within the target define the energy layers, grouping spots of identical energy. This trajectory simulation takes into account the water-equivalent thickness (WET) maps, quantifying the

radiological thickness of a particular tissue that would cause the same amount of protons energy loss as a certain thickness of water. Water equivalence is employed throughout the treatment planning and delivery process to simplify and standardize the deposition of dose in tissues and the determination of beam range. At the end of the spot placement we therefore have a list of spot with their respective position and energy. At this stage, the dose contribution and intensity of each spot is still unknown.

Dose calculation

In the context of PBS delivery, the dose map typically results from irradiation using multiple elementary beams, known as beamlets. Each beamlet corresponds one-to-one with a spot and has a specific weight that determines its contribution to the total dose.

The accurate computation of proton dose distribution in a patient is essential for effective treatment planning. The first step is to establish the relationship between the CT number and proton energy loss, known as the proton stopping powers (SP). These SP values are typically provided by the CT calibration curve of the machine in-use [68]. To standardize the description of tissue properties in PT planning, the stopping power ratio (SPR) is often used. SPR is the ratio of the stopping power of a specific tissue to the stopping power of water:

$$\text{SPR} = \frac{(dE/dx)_{\text{tissue}}}{(dE/dx)_{\text{water}}}$$

where dE/dx is the stopping power, i.e. the mean energy deposited per unit distance. This quantity mainly depends on the materials crossed.

Based on the SPR or SP, a 3D dose distribution can be numerically approximated. Two main types of algorithms can be used and dictate the accuracy with which one wants to compute the dose: analytical and Monte Carlo algorithms.

Analytical algorithms are model-based methods used to approximate the dose distribution within a patient's body. The most common analytical algorithm is the pencil beam (PB) algorithm [69], which divides the proton beam into a grid of narrow pencil beams. The PB algorithm use pre-calculated dose kernels that describe how dose spreads from one point to another in a medium. These kernels are superposed and convolved with the fluence distribution (pencil beam intensities) to compute the total dose distribution. While the position of the kernels depends on the machine parameters, their size is scaled in depth according to the pre-computed SPR maps on the proton pencil central axis. The analytical algo-

gorithms are efficient and provide rapid dose calculations. However, they present limitations in accuracy, especially in heterogeneous structures such as the lungs.

Monte Carlo (MC) algorithms, on the other hand, use statistical methods to simulate the transport of individual protons through the patient's anatomy [70]. These algorithms simulate the multiple fundamental interactions that protons undergo, such as electromagnetic and nuclear reactions (accounting for scattering and energy loss), with high accuracy. By iteratively simulating random physical interactions for a very large number of protons, contingent upon the media's SP and cross-section (i.e. interaction probability), MC algorithms provide a detailed and accurate representation of the dose distribution within the patient. This level of accuracy makes them particularly valuable for situations involving complex geometries, heterogeneities, and sophisticated fields configuration. The drawback of MC simulations is their computational intensity and time-consuming nature, which made them initially better suited solely for final dose calculation and verification in clinical settings. Nevertheless, the field has seen ongoing advancements aimed at accelerating MC dose calculations. These advancements have made MC simulations more accessible and have established them as the gold standard in modern clinical practice [71, 72, 73, 74, 75].

Throughout this thesis, MC algorithms were consistently selected for dose calculations due to their accuracy and availability in the software in use.

In our workflow, MC dose calculation is first used to pre-compute independently the individual dose contributions of each elementary pencil beam, the so-called beamlet, each targeting a specific spot in the tumor. A beamlet is mathematically represented by a vector whose elements are elementary dose contribution to each voxel in the 3D-CT image. Even by reducing the dose grid size and resolution, the number of voxels in a CT image usually reaches hundreds of thousands elements while the number of beamlets, highly dependent on tumor size and number of beams, is generally ranging from hundreds to thousands spots for an IMPT plan. However, the dose contribution from each beamlet over all the voxels in the geometry is mainly zero except in the proton beam path. The beamlets are therefore usually stacked in a compressed sparse matrix format to limit its size for subsequent calculation. The total dose map can be built as a summation of weighted contributions of the elementary beamlets. Formally, in matrix form,

$$\mathbf{d} = \mathcal{A}\mathbf{x}, \quad (2.2)$$

where $\mathbf{d}$ is the vectorized dose, $\mathcal{A}$ is the beamlet matrix, and $\mathbf{x}$ is the weight

vector. In the context of protons and higher LET beams, $\mathbf{d}$ should be a biologically weighted dose in the voxel that includes absorbed dose $\times$ RBE averaged over the pencil beam contributions

Elaborating on all contributions:

$$d_i = \sum_j a_{ij} \cdot x_j$$

Plan optimization

Plan optimization is a major step in PBS-based treatments, as it involves determining the intensity of each pencil beam to create uniform high-dose regions with arbitrary shapes. This process, often referred to as *inverse* planning, takes the desired dose distribution or clinical objectives as input and determines the intensity map $\mathbf{x}$, also referred to as fluence map, needed to achieve it. This stands in contrast to the traditional *forward* planning method, which is a “trial and error” technique. In forward planning, the planner hypothesizes parameters that might yield an optimal dose distribution. This manual method prevailed until the development and widespread adoption of modulated techniques.

Inverse planning operates by iteratively adjusting the spot weight vector, $\mathbf{x}$, which determines the fluence of each pencil beam, aiming for the total dose, $\mathbf{d}$, to closely match the desired dose for the target and to meet the clinical constraints for the OAR. Central to this process is the cost function, $f(\mathbf{d})$, a mathematical representation that combines clinical goals into weighted objective functions, condensing the overall quality of the treatment plan into a scalar value. Often referred to as the “dose fidelity term”, this score measures the difference between the planned and the desired treatment and should be minimized as per Eq. 2.3. The minimum of the function f represents the optimal balance that satisfies the objectives of the treatment. The treatment planning problem can therefore be stated as the following optimization program:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{d}) = \sum_{k=1}^n w_k f_k(\mathbf{d}) \\ \text{subject to} \quad & \mathbf{d} = \mathcal{A}\mathbf{x} \\ & \mathbf{c}(\mathbf{d}) \leq 0, \\ & \mathbf{x} \geq 0, \end{aligned} \tag{2.3}$$

where $c(\mathbf{d})$ are the hard planning constraints, w_k is a weighting factor that represents the relative importance of each objective f_k . Objective functions $f_k(\mathbf{d})$ can also be alternatively swapped with constraints $c(\mathbf{d})$ that impose stringent limitations on the set of feasible solutions and can adopt a nonlinear form. f_k is typically represented as an asymmetric quadratic function that places a higher penalty on outliers, is differentiable, and has a linear derivative. This function penalizes deviations from a reference dose defined for a particular ROI. For instance, minimum or maximum dose fidelity terms are expressed as:

$$f_k(\mathbf{d}) = \begin{cases} \frac{1}{|\text{ROI}_k|} \sum_{i \text{ in } \text{ROI}_k} \min(0, d_i - p_{\min})^2, & \text{or} \\ \frac{1}{|\text{ROI}_k|} \sum_{i \text{ in } \text{ROI}_k} \max(0, d_i - p_{\max})^2 \end{cases} \quad (2.4)$$

with $|\text{ROI}_k|$, the number of voxels within the ROI and $p_{\min/\max}$ the reference dose limits. The objectives can also have different formulations from the quadratic penalty (seen in Eq. 2.4), some of which might not be convex such as dose-volume constraints. Consequently, the selection of an optimizer is influenced by the chosen problem statement. The traditional approach employs convex approximations, where the objective function is quadratic and constraints are linearized. However, this often demands the adoption of comprehensive optimization libraries adept at handling complex, multi-constraint problems [76, 77]. The final solution therefore depends on the cost function guiding the search, the constraints placed on the search and the type of “search engine”, i.e. optimizer, used. The optimization algorithms will be further tackled in section 2.3.

Multi-criteria optimization

Objectives frequently give rise to conflicting goals, particularly when dealing with situations like tumors located adjacent to or within OAR structures. Consequently, the optimizer’s initial convergence may not produce a satisfactory solution, compelling planners to fine-tune individual objectives, such as adjusting relative importance weights (w_k in Eq 2.3), and iterate the optimization process. This trial-and-error procedure can be repeated multiple times until the dose distribution aligns with clinical objectives. However, this manual tuning is often laborious and time-intensive. Another solution is to use a multi-criteria optimization (MCO) tool, that can sometimes be available in TPS. This tool considers objectives on an individual basis and aims to find a range of treatment plans that strike the best balance among them [21, 22, 23]. After generation of the optimal treatment plans, intermediate solutions are intelligently interpolated, and an interactive plan navigation tool is employed to explore the optimal trade-offs between planning objectives for the patient. This process can be completed within minutes

in commercial TPS, in contrast to the conventional method that might take hours for complex treatments. MCO will be further described in section 2.3.

Robust optimization

We saw earlier that PT, due to the nature of proton interactions with matter, has inherent sensitivities to range uncertainties. Generally, geometric treatment uncertainties can arise from various sources, but are usually categorized into:

1. Setup errors: Variations in patient positioning relative to the radiation isocenter from fraction to fraction.
2. Range errors: Uncertainties arising from tissue density variations throughout the treatment course and from inaccuracies in the initial conversion of CT numbers to the stopping power of tissues [78]. Both can impact the depth at which protons deliver their maximum energy deposition.
3. Tumor motion: Tumor motion is often caused by physiological processes like breathing, digestion, weight loss, tumor shrinkage and so on. In particular, when the tumor is located in motion sites such as lung or esophageal cancer.

If these treatment errors are not explicitly considered during treatment planning, this can lead to situations where minor geometric deviations from the *nominal* planned dose, where nothing harmful or unexpected occurs, result in serious differences in the actual dose delivered to the tumor and surrounding normal tissues. Robust optimization aims to address this issue. To introduce robustness into the optimization process, we consider how uncertainties may impact the beamlets and, consequently, the dose. In mathematical terms, denoting these uncertainties as $\mathbf{u}$, we express this relationship as $\mathbf{d}(\mathbf{u}) = \mathcal{A}(\mathbf{u})\mathbf{x}$. Parameter $\mathbf{u}$ is drawn from a random variable U modeling the uncertainty distribution. In practice, the distribution is sampled into discrete scenarios $u_0, u_1, \dots$ from U . Considering that $\mathcal{A}(u_0) = \mathcal{A}$, meaning that u_0 is the *nominal* scenario, other cases u_1, u_2 etc. are often called the *uncertainty* scenarios. The corresponding doses and cost function values can therefore be computed for each uncertainty scenario (requiring beamlet computation for each $\mathcal{A}(u_i)$). The optimization then seeks to find a solution that performs well across all these scenarios, rather than just optimizing for a single, “best-guess” scenario.

From that point, there are two predominant robustness paradigms depending on the discrete scenarios selected from U . The *probabilistic* approach minimizes the expected value of the objective function over the uncertainty distributions U [79]. The second robustness strategy, more conservative, is the *worst-case* scenario

approach whose optimization process focuses on the worst-performing scenarios to ensure that even in the least favorable conditions, the treatment plan remains clinically acceptable. In practice, it evaluates the cost function over all specified scenarios independently, identify the worst case (highest value) and optimize x for that scenario. If, subsequently, another scenario becomes the worst, then it optimizes that one, etc. Because we minimize the spot weights and maximize over the scenarios in an integrated scheme, this is called a ‘minimax’ problem [80],

$$\begin{aligned} \min_x \max_U \quad & f(\mathcal{A}(u)x) \\ \text{subject to} \quad & x \geq 0 \end{aligned} \tag{2.5}$$

The worst-case scenario optimization remains the preferred method in clinical practice as it yields high-quality treatment—a strategy also adopted in this thesis.

Treatment plan evaluation

During and after optimization, it is important for the planner to assess the quality of the proposed solution. This is done using both visual tools and numerical metrics. One of the main methods of evaluation involves dose-volume histograms (DVH). A DVH is a cumulative histogram where the dose (on the x-axis) is plotted against the volume percentage of a structure (shown on the y-axis), like a target or an OAR, that receives at least that dose. An example with three different structures can be seen in Fig. 2.5. By looking at the DVH, one can determine if the treatment objectives, such as the target dose or the maximum dose for an OAR, have been met. DVH offer an overview of how the dose is distributed in and around the tumor, making sure the plan follows clinical guidelines. Metrics extracted from these histograms help check if the treatment meets the clinical goals. For instance, $D_{95\%}$, indicates the dose that 95% of the target volume receives, ensuring most of the tumor gets the prescription dose. $D_{0.03\text{cm}^3}$ measures the dose that a small volume (0.03 cm³) receives. This metric is thus important for very sensitive organs like the spinal cord. Other metrics often reported include:

- Conformity index (CI): A measure of how tightly the prescribed dose conforms to the shape of a tumor. A CI equal to 1 corresponds to the ideal dose coverage or high conformity.
- Homogeneity index (HI): A measure of how evenly the radiation dose is distributed within a tumor, with lower values indicating better dose homogeneity.

- Integral dose (ID): The dose integrated over the body volume, providing insights into the overall exposure of healthy tissues to radiation.

The mathematical definitions of these metrics often vary in the literature, so we will reference the specific definitions used throughout this thesis.

In addition to DVH, 3D dose maps should also be visually inspected to check the dose distribution homogeneity and avoid cold or hot spots (i.e. under- and over-dosage).

Because dose maps can be influenced by uncertainties, it is important to evaluate the treatment plan's robustness against them [81]. This is called robustness evaluation. This means checking how stable the plan is even when patient or situation changes slightly. One common method used in clinics is scenario analysis. This involves recalculating multiple dose maps that consider different potential changes, like a patient moving slightly or uncertainties in calculating proton range. In commercial TPS, the scenarios generation usually follows what is done for the robust optimization. It simulates the dose in the error space for a number of extreme scenarios ("maximum error"). Each perturbed dose map is then checked and we assess the percentage of scenarios covered by a given robustness test (e.g. 98% of CTV volume receives 95% of the prescribed dose). Another method is to use a statistical approach, where many scenarios are randomly sampled from their probability distribution [82, 81]. The analysis can then be done in the dose space, where we evaluate the percentage of dose distributions that meet specific clinical endpoints, such as minimal target coverage. Because many scenarios must be computed to obtain statistical confidence, it is currently not used in clinical routine.

Lastly, we sometimes use the NTCP to provide a quantitative and biological measure of the likelihood of a given side effect occurring in normal tissues due to radiation exposure [83]. By looking at NTCP values for different organs, physicians can balance the goal of treating the tumor with the toxicity risk to other tissues. In general, a plan that has low NTCP values and still treats the tumor effectively is preferred.

2.2.4 Treatment verification

Once the treatment plan has been approved by the treating physician, it can undergo the quality assurance (QA) process. Treatment QA is the process of ensuring that the delivered dose accurately matches the planned dose for a patient's radiation therapy treatment. It guarantees patient safety and treatment efficacy. It usually simulates the entire treatment process, from imaging to dose delivery,

to ensure all components of the treatment workflow are accurate and synchronized. The measured characteristics during the QA of proton therapy treatments typically include: the isocentricity of the proton beam, the size and position of the scanned beams, the fluence of the beams and the energy of the protons.

For patient-specific QA, the planned treatment is first delivered on a device or into a plastic cube, called a phantom, which measures the absorbed dose [84, 85]. In IMPT treatments, we use tools like ionization chambers, films, or 2D/3D detector arrays to do this. One common method is to deliver each beam of the treatment plan on the detector. This detector then records the dose at certain depths, for example, in the middle and distal edge of the SOBP. The comparison method between the planned and measured dose is usually based on a γ -analysis [86] that takes into account both dose and spatial differences to evaluate the plan's accuracy. A higher pass rate in γ -analysis indicates better agreement between planned and delivered doses.

Another QA process is the log-based QA that refers to the use of machine-generated log files for verifying the accuracy and consistency of radiation therapy delivery [87]. During treatment, proton therapy systems can produce log files that record real-time machine parameters such as gantry angles, dose rates, and beam on/off timings. The log-files can then be used for dose recomputation.

2.2.5 Treatment delivery

During each treatment fraction, the patient's positioning on the treatment couch aims to replicate their position from the CT simulation session. This typically involves a supine or prone orientation, although advancements are leading to solutions for upright positioning, with the patient secured to a rotating chair [88, 89]. Accurate patient alignment is important to ensure the proton beam hits its target accurately. Modern proton therapy rooms are equipped with sophisticated imaging devices, including dual orthogonal X-rays, cone-beam CT (CBCT), and in-room CT, to facilitate real-time anatomical visualization. This imaging not only ensures proper alignment with the planning CT but also monitors anatomical variations, potentially prompting adaptive replanning workflows.

Upon satisfactory positioning, the treatment is administered via one of the modalities elaborated in Section 2.1.5: either DS or PBS. With PBS, each spot is associated with a specific intensity found by the optimizer and representing the "spot weight". However, for the actual delivery of radiation, these spot weights are converted into Monitor Units (MU), a standardized measure that ensures the consistent and accurate delivery of the prescribed radiation dose. MU are important because they quantify the machine's output, taking into account factors

like beam energy and intensity. The DICOM standard provides spot weights and a scale factor to accommodate the variations in different proton therapy machines. This scale factor is then used to convert the relative spot weights into actual MU based on the specific calibration of the proton therapy machine in use.

During the actual PBS delivery, the “interplay effect” [90, 91] demands vigilance. This effect arises from the interference between the dynamic beam delivery and the patient’s anatomy movement when the frequencies of beam delivery and movement are similar, potentially introducing dose heterogeneities. To mitigate this effect, rescanning also called repainting can be used. Instead of delivering the entire planned dose to a specific spot in the tumor in one go, the dose is split into multiple smaller doses. The proton beam then “repaints” or revisits each spot multiple times during the treatment session, delivering these smaller doses. By repeatedly scanning over the target volume, the effects of the interplay between the dynamic beam delivery and tumor motion can be averaged out.

Recent technological strides have introduced innovative delivery paradigms like FLASH-PT and proton arc therapy. FLASH therapy seeks to deliver treatment at an ultra-high dose rate, exceeding 40 Gy/s within one to five fractions. This delivery mode exploits a specific radiobiological phenomenon that may reduce normal tissue radiosensitivity while maintaining tumor control [92]. Despite its novelty and yet-to-be-deciphered biological principles, FLASH’s key challenge lies in its demand for ultra-high dose rates that current PBS systems cannot deliver, necessitating a technological shift [93]. One approach employs high-energy “transmission beams” [94], though this risks limiting the inherent dosimetric benefits of protons by delivering excessive dose beyond the tumor. In contrast, conformal FLASH proton therapy employs sophisticated beam-shaping tools (interestingly reminiscent of DS technology) to obtain similar dose distributions to IMPT. This method uses a patient-specific range modulator that is positioned between the machine’s nozzle and the patient. The range modulator adjusts how deep the proton beam goes, but only uses one high energy from the proton system. As a result, it avoids delays associated with energy layer switching and ensures the necessary dose rate. In this context, FLASH necessitates the optimization of the range modulators and the spot delivery pattern to enhance local dose rate [95, 96].

A distinct yet equally (but up for debate [97]) promising modality is proton arc therapy (PAT or ArcPT). Mimicking its photon counterpart (rotational IMRT), PAT involves a continuous beam rotation and intermittent irradiation delivery encircling the patient, hitting the tumor from many angles. This new delivery mode, potentially faster than IMPT delivery, fosters enhanced dose sculpting and improved biological impact of protons. Notwithstanding its potential, PAT’s clinical

integration demands sophisticated hardware and software modifications and rigorous safety validations. A comprehensive discourse on PAT, the centerpiece of this thesis, is provided in Section 2.4.

2.3 Mathematical background

This section, while not providing a fully formalized mathematical exposition of optimization, aims to offer a comprehensive overview of key mathematical concepts and intuitive explanations that support the choices made in this thesis.

2.3.1 Terminology in optimization

Mathematical optimization is a powerful tool that aims to solve and find the best solution possible for real-world problems encountered daily across diverse fields.

Objective Function

The objective function $f(x)$, often referred to as the cost, fitness, loss, or energy function, defines the quantity of interest of an optimization problem. Whether aiming to minimize or maximize, $f(x)$ takes decision variables x , often termed optimization weights or parameters, as input, that returns a scalar value that represents the optimization objective. In this thesis, the formulation we use is to minimize the objective function rather than to maximize it.

Constraints

Constraints are mathematical expressions that impose restrictions on feasible solutions. In an optimization problem, they are often inequalities or equalities. Inequality constraints $g_i(x) \leq b_i$ and equality constraints $h_j(x) = d_j$ define the feasible set of solutions $\mathcal{X}$. While constraints are common, not all optimization problems have them. Some problems are naturally unconstrained, while others become unconstrained through specific problem reformulations.

Decision Variables

Decision variables are the quantities to be determined or optimized, i.e. the unknowns of an optimization problem. Represented as a vector $x = [x_1, x_2, \dots, x_n]$, they are the variables we can adjust to achieve an optimal solution while adhering to constraints. The domain of the decision variables is important: they can be real numbers (often ≥ 0), integers, binaries, or a mix of these.

Mathematically, optimization problems are formulated as follows:

$$\begin{aligned}
 & \text{maximize/minimize} && f(x) \\
 & \text{subject to} && g_i(x) \leq b_i && i = 1, 2, \dots, m \\
 & && h_j(x) = d_j && j = 1, 2, \dots, p \\
 & && x \in \mathcal{D}
 \end{aligned} \tag{2.6}$$

where $\mathcal{D}$ represents the domain of the decision variables, i.e. the allowable range or set of values that these variables can take. In mathematics, a set is a collection of distinct elements. In this context, it is a collection of numbers, grouped together as a single entity, which may be finite or infinite.

The form of $f(x)$, $g_i(x)$ and $h_i(x)$ also dictates the nature of the optimization problem. If they are linear, we are dealing with linear programming (LP). If the decision variables are mixed (i.e., some are integers, some are real numbers), it becomes a mixed integer linear program (MILP). In this thesis, we refer to it as MIP, but it could be more specifically termed MILP. When $g_i(x)$ or $h_i(x)$ are nonlinear, it falls under nonlinear programming. If the functions are quadratic, it is referred to as quadratic programming (QP). Further connections to these concepts can be observed in Figure 2.7 and will be described in broader terms in Section 2.3.2.

Optimum

An optimum x^* (minimum or maximum depending on the problem), refers to one of the best possible solutions, if any, for the given optimization problem. Depending on the form of the objective function and the solver chosen, a *local* or a *global* optimum can be found. A local optimum is the best solution within a certain local neighborhood, while a global optimum is the best solution over the entire feasible set $\mathcal{X}$. Formally, x^* is a global minimum if for all x in $\mathcal{X}$, $f(x^*) \leq f(x)$. While an optimization problem can have multiple local optima, it may also have multiple global optima, i.e. situations where multiple solutions achieve the same best objective value. For instance, in problems defined by linear constraints (even if convex), multiple global optima can be found when the objective function is parallel to one of the constraints within the convex hull of $\mathcal{X}$.

2.3.2 Optimization problems and methods

An essential step in the optimization process is the classification of optimization techniques, as the choice of algorithms depends on the specific problem type.

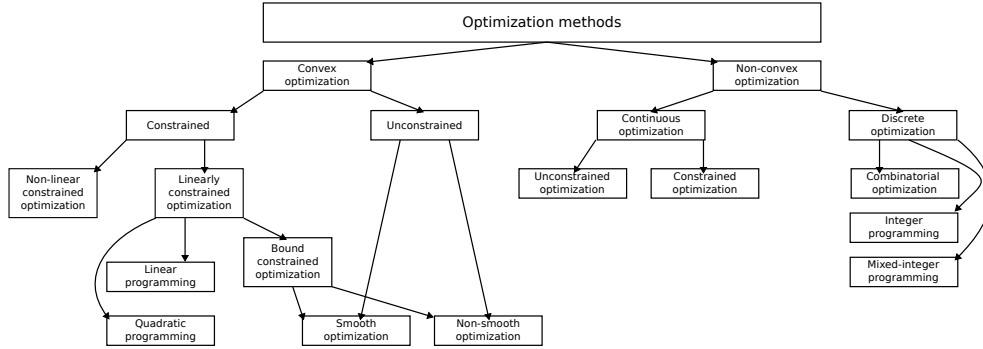


Figure 2.7: Optimization problem types. Adapted from [NEOS](#).

Since this thesis has employed various optimization problem statements, we aim to provide a comprehensive (though not exhaustive) outline of different optimization problem types. We will use the diagram shown in Fig. 2.7 as a guide to explore these categories.

Convex optimization versus non-convex optimization

An optimization can be primarily categorized depending whether it is *convex* or *non-convex*. An optimization problem is convex if it involves minimizing a convex objective function, possibly subject to convex constraints, and the domain of feasible solutions is a convex set. In mathematics, a set is considered convex if, for any two points within the set, the straight line segment connecting those points lies entirely within the set itself. On the other hand, a convex function implies that the curve of the objective function "opens upwards" and remains below any straight line segment connecting two points on the curve as illustrated on Fig. 2.8. Mathematically, a problem is convex if, for all $x, y \in X$ and λ in the interval $[0, 1]$, the following condition holds:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

Convex optimization problems are desirable because it ensures that a local minimum found of a convex function is also the global minimum. This property simplifies the optimization process, as there is a single optimal solution that can be found efficiently using various methods, including gradient-based algorithms. In contrast, a non-convex optimization problem is a problem where the convexity condition is not met. This means that there exist points x and y in the feasible set $\mathcal{X}$ and λ in $[0, 1]$ for which the inequality mentioned above does not hold. Non-convex problems can have multiple local minima, making the search for the global minimum more complex and challenging.

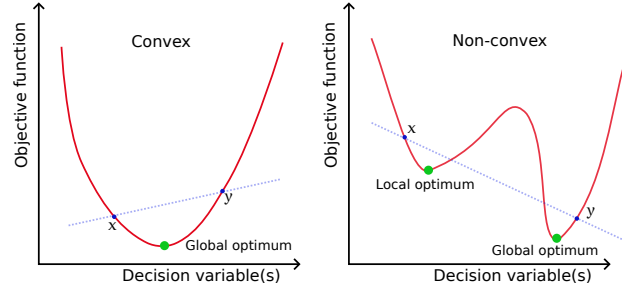


Figure 2.8: Examples of convex and non-convex optimization problems. The function is convex if the dotted line is above all points in the graph between x and y points.

In convex optimization, we often discuss the (non-)smoothness of a function, referring to the number of continuous derivatives the function has. In simple terms, a function could be considered smooth if it is differentiable everywhere and, therefore, continuous. Smooth functions are well-behaved and are preferred in optimization because they allow for efficient algorithms that use gradients or derivatives to find optimal solutions. For instance, the simple gradient descent method uses gradient of the objective function $\nabla f(x)$ to iteratively update decision variables x towards a local minimum or maximum. The update rule is given by:

$$x_{k+1} = x_k - \alpha \nabla f(x_k) ,$$

where α is a step size that controls the size of the update. Taking advantage of second-order derivative information, Newton (Hessian) or quasi-Newton (approximation of the Hessian) methods such as the BFGS (Broyden-Fletcher-Goldfarb-Shanno) [98] algorithm provides faster convergence. The Hessian provides information about the curvature of the function and helps determine the direction and step size for the update. Quasi-Newton methods build an approximation of the Hessian during the optimization and update it based on the changes in gradients. The solution is then iteratively adjusted using the inverse of this approximation.

In contrast, non-smooth functions are functions that lack continuity in their derivatives, making them not differentiable everywhere on their domain. These functions may present abrupt changes, discontinuities, or points where the derivatives are undefined. One common approach to solving convex problems with non-smooth functions is to use subgradient methods [99], which are extensions of gradient-based methods that work with non-smooth functions. Another technique utilizes proximal methods [100], which are especially effective for optimization problems involving non-smooth regularization terms in the objective functions.

For instance, a commonly employed regularization term in optimization is the L1-norm, often referred to as the “Lasso” penalty [101]. This term promotes sparsity in the solution by adding the absolute values of the problem’s vector weight as a penalty, helping select major weight contributions while discarding less important ones. The key idea behind proximal methods is to incorporate a *proximal operator* which is a mathematical function that finds a point that strikes a balance between minimizing the non-smooth function and being close to the current point estimate. This balance is achieved by minimizing the non-smooth function with an added squared distance term, which penalizes deviations from the current estimate and smooths the optimization problem. Actually, for many standard functions used in optimization, this proximal operator can often be analytically derived, facilitating its practical application in various optimization algorithms. Proximal methods typically involve two main steps in each iteration. First, the classical gradient descent step using the smooth part of the objective function. This smooth part usually corresponds to the original function without the non-smooth regularization terms. After the gradient descent step, the proximal operator is applied to the current solution. The proximal operator effectively “pulls” the solution towards the desired properties, such as sparsity or non-negativity. They are widely used in various fields, including signal processing [102], machine learning [103], image reconstruction [104], and more. Common proximal algorithms include the Proximal Gradient Descent (also known as ISTA [105]) and the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [106], used in this thesis, addressing large-scale and high-dimensional optimization problems.

Continuous optimization versus discrete optimization

The distinction of *continuous* and *discrete* optimization emerges next. Continuous optimization deals with problems where the variables can take on any real value within a given range. In contrast, discrete optimization involves problems where the variables are restricted to discrete values, such as integers or a set of pre-defined options. This can substantially impact the complexity of the optimization problem, as the solution space becomes more fragmented. Integer programming (IP) is actually an extension of linear programming (LP) where some or all decision variables are restricted to take only integer values, making it suitable for discrete decision-making scenarios. Techniques like dynamic programming [107], branch and bound [108], or meta-heuristic algorithms (e.g., genetic algorithms) [109] are often applied to solve discrete optimization problems. Mixed-integer programming (MIP) combines both continuous and discrete decision variables. Some variables are allowed to take real values, while others are constrained to be integers. MIP problems encompass a wide range of practical situations where decisions involve both continuous and discrete choices.

Linear versus nonlinear

Linear and *nonlinear* refer to the nature of the relationships between variables in both the objective function and the constraints. Linear programming (LP) problems have linear objective functions and linear constraints and can be represented in standard form [110] as:

$$\begin{aligned}
 &\text{maximize} && \sum_{j=1}^n c_j x_j \\
 &\text{subject to} && \sum_{j=1}^n a_{ij} x_j \leq b_i \quad i = 1, 2, \dots, m \\
 &&& x_j \geq 0 \quad j = 1, 2, \dots, n
 \end{aligned} \tag{2.7}$$

Linear problems are a special case of convex problems, where convex problems represent a broader category that includes both linear and certain types of nonlinear problems. Not all nonlinear problems are convex, but all linear problems are inherently convex. Linear optimization problems can thus often be solved efficiently using methods like the Simplex algorithm or interior-point methods [111]. When introducing nonlinear terms like exponentials, logarithms, trigonometric functions, etc. in the objective function or in the constraints, or both, nonlinear programming must be applied. Nonlinear optimization problems can be more complex to solve and often require specialized algorithms such as gradient-based or meta-heuristic approaches. When the objective function includes quadratic terms involving the decision variables, it becomes quadratic programming (QP). The constraints in QP, however, can still be linear or quadratic. Sequential quadratic programming (SQP)[112] combines quadratic programming with nonlinear optimization by approximating the objective function and constraints using quadratic models.

A few more important details

Optimization methods can be further categorized as either *exact*, *approximate* or *heuristic*. Exact methods aim to find the optimal solution ensuring that the computed solution is mathematically proven to be the best possible. Exact methods like the Simplex algorithm for LP and Branch and Bound for (M)IP problems are designed to find the optimal solution. In particular, Branch and Bound computes two types of bounds: lower and upper bounds. Optimality is then proven whenever both bounds coincide. In contrast, approximate methods generally seek to find good solutions within a reasonable timeframe but typically do not provide

a provable bound on the optimality gap. They may provide solutions that are close to the optimal solution but may not guarantee a strict mathematical optimality. In contrast, heuristics provide no such guarantee at all and prioritize computational efficiency and practicality. Approximate methods and heuristics are particularly useful when dealing with complex, large-scale problems where finding an exact solution is computationally infeasible.

Optimization algorithms can also be classified as *local search* or *global search* methods. Local search algorithms are heuristics that focus on iteratively refining candidate solutions by applying local changes (i.e., exploring nearby solutions) until an optimal solution is found or time runs out. Examples include hill climbing or gradient-based methods. However, they may get stuck in local optima and fail to find the global optimum. Global search algorithms, in contrast, are designed to efficiently navigate the entire solution space, often by balancing exploration (broad search for better solutions) and exploitation (enhancement of known solutions). These algorithms, often utilizing population-based approaches, seek to discover optimal or near-optimal solutions. Inspired by nature, methods like genetic algorithms [113] iteratively improve the population's fitness over multiple generations, selecting and combining solutions, facilitating exploration of diverse areas of the solution space and enhancing the chances of finding optimal solutions. These algorithms are usually grouped under the name "meta-heuristics" which exhibit desirable properties like simplicity, parallelism, and applicability [109].

Finally, optimization methods are usually categorized based on their determinism. *Deterministic* methods follow a predictable path during optimization and produce the same solution for the same initial conditions. They guarantee result reproducibility. *Stochastic* methods, on the other hand, introduce randomness into the optimization process. These methods incorporate probabilistic elements, which can be advantageous when dealing with complex, multimodal (i.e., multiple maxima or minima), or noisy objective functions that might be non-smooth, or time-consuming to evaluate. Stochastic methods offer the potential to escape local optima and explore a wider solution space, but their found solution can vary across different runs due to the random nature of their steps. For instance, simulated annealing (SA) [114, 115] is a global search and stochastic optimization method inspired by the annealing process in metallurgy. It searches for an optimal solution by iteratively exploring the solution space while allowing for occasional "jumps" to less favorable solutions. These jumps decrease in frequency over time, resembling the cooling process in metallurgy, and help the algorithm escape local optima and find a global optimum. Like meta-heuristics, SA are derivative-free algorithms meaning that don't rely on derivatives (such as gradients) of the objective function to find its minimum or maximum.

2.3.3 Multi-criteria optimization

Optimization problems often involve more than a single objective. When several objectives need to be simultaneously maximized or minimized, the optimization problem becomes a multi-criteria (or multi-objective) optimization (MCO) problem. Formally, a MCO problem can be written as:

$$\begin{aligned} &\text{minimize/maximize} && f_1(x), f_2(x), \dots, f_k(x) \\ &\text{subject to} && x \in \mathcal{D}, \end{aligned} \tag{2.8}$$

where $f_1, f_2, \dots, f_k$ are the objective functions. Given the presence of multiple objectives, a single “optimal” solution typically does not exist. Instead, we encounter a set of solutions known as *Pareto optimal* solutions.

A solution is termed Pareto optimal if no other feasible solutions exist that would make one objective function value better without causing a detriment to at least one of the other objective function values. This idea is tied to the principle of *Pareto dominance*, wherein one solution x_A is said to Pareto dominate another solution x_B if x_A is at least as good as x_B in all objectives and strictly better in at least one objective.

Consider Fig. 2.9. Here, we visually represent the optimization problem associated with radiotherapy treatment planning which is by definition a MCO problem [116]. In this case, the objective on the x-axis represents the target coverage while the objective on the y-axis represents the normal tissue sparing. They both need to be maximized. Each dot on the graph represents a unique beam configuration or a specific set of spot weights used to deliver the treatment. Different configurations lead therefore to varying balances between target coverage and normal tissue sparing. Red dots are not Pareto optimal. For every red dot, there is another configuration that performs better in both target coverage and tissue sparing or offers a better performance in one without compromising the other. Green dots stand for the Pareto optimal solutions. These configurations strike the best possible balance between the two objectives. No other configurations outclass them in both objectives simultaneously. The blue curve, known as the Pareto frontier (or Pareto front), connects all the Pareto optimal solutions. It graphically displays the trade-offs a clinician would need to consider: moving along the frontier, improving target coverage might result in reduced tissue sparing and vice-versa.

Obviously, multi-objective optimization problems may involve more than two objectives but adding extra dimensions makes visualization more complex. Nevertheless, the principle remains: solutions on the Pareto front (which could extend to higher dimensions as a surface, volume, or hyper-volume) represent the best trade-offs among the objectives.

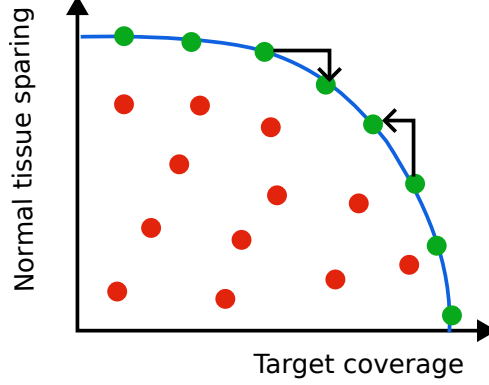


Figure 2.9: Pareto front concept applied to the radiotherapy treatment planning problem where the objective is to maximize the objective function. Red dots are not Pareto optimal, green ones are Pareto optimal and blue curve is the Pareto front.

Multi-criteria optimization involves multiple strategies to find Pareto optimal solutions or approximate the Pareto front. A straightforward approach to multi-objective problems involves combining multiple objectives into a single scalar objective using weighted sums, which can be represented as:

$$f(x) = w_1 f_1(x) + w_2 f_2(x) + \dots + w_k f_k(x) \quad (2.9)$$

Nevertheless, this scalarization technique does not find the whole Pareto front. Moreover, selecting appropriate weights $(w_1, w_2, \dots, w_k)$ is non-trivial and can greatly influence the derived solutions. These weights can be adjusted manually based on domain knowledge or preferences. Alternatively, various dedicated MCO algorithms can be employed to explore different weight combinations, aiming to provide a comprehensive view of the Pareto front. Constraint methods [117] for instance, choose one of the objective as the primary objective to be optimized, while the others are turned into constraints with a threshold. By varying the threshold values, different parts of the Pareto front can be obtained. Goal programming [118] is also a popular approach where each objective is assigned a goal. The optimization process aims to minimize the deviations from these goals for all objectives. Evolutionary Algorithms [119], due to their population-based nature, are well-suited for MCO. They work by evolving a population of individuals (i.e. solutions) over generations to approximate the Pareto front.

2.3.4 Complexity theory

As complexity theory falls outside the scope of this thesis, we aim to keep this introductory section as brief as possible. To do so, we will introduce a more

general terminology that sidesteps the detailed rigor typically found in computer science literature. This will lay the foundation necessary to grasp the complexity proof presented in Chapter 8.

In theoretical computer science, complexity theory [120] is the study of the computational resources required to address specific problems. At its heart, it seeks to classify computational problems based on the amount of a resource, like time, that they necessitate to be solved.

When we speak of time complexity of an algorithm, usually in its worst-case scenario, we are measuring the maximum number of operations this algorithm might take relative to the size of its input. This is important in understanding the applicability of algorithms; some solution methods may work well for small inputs but become intractable (i.e., cannot be efficiently solved) as the input size increases. Formally, for a function $f(n)$ that describes the number of steps an algorithm takes for an input size n , if there exist constants c and n_0 such that $f(n)$ is bounded above by $c \cdot g(n)$ for all $n \geq n_0$, where g is the comparison function, then using the big O notation, we say its worst-case time complexity is $\mathcal{O}(g(n))$. Intuitively, this means that f does not grow faster than g .

Consider the notion of a *polynomial* algorithm. Such an algorithm's time complexity can be expressed as a polynomial function of its input size. Formally, if an algorithm's time requirement for an input of size n can be bounded by some n^k (where k is a constant), we say it has polynomial-time complexity, written as $\mathcal{O}(n^k)$. Such algorithms are viewed favorably in computer science because they scale reasonably well with increasing input size. For example, the selection sort algorithm [121] for sorting an array of numbers is a polynomial algorithm because its worst-case time complexity is $\mathcal{O}(n^2)$. This is due to the fact that for each of the n elements, the algorithm potentially compares it with up to n other elements, resulting in a total of about $n \cdot n$ or n^2 comparisons.

In general, decision problems (having a 'yes' or 'no' answer) can be classified depending on the difficulty it takes to solve them. The problems that can be solved efficiently, i.e., using a polynomial algorithm are said to belong to the set P (for *polynomial*). Then we have problems that fall under *non-deterministic polynomial* (NP) category. A decision problem is in NP if, for every solution, there exists a polynomial-sized certificate (or evidence) that can be checked in polynomial time. For example, in the traveling salesman problem (TSP) [122]: "Is there a route visiting a pre-defined group of cities within a 1000km limit?" For every answer, the certificate would be the specific route taken, which can be verified in polynomial time to ensure it does not exceed 1000km. Then there are more

sophisticated problems within NP that stand out due to their intrinsic difficulty. These are the NP-complete problems for which no efficient solution algorithm has been found, but whose solutions are easy to verify provided a certificate (i.e. in NP). Many important computer-science problems belong to this class — e.g., the TSP mentioned above (decision version). Beyond NP-complete problems lie the NP-hard problems. These are, in some sense, even harder. They include problems that are at least as hard as the hardest problems in NP. However, they do not need to be decision problems, and they might not even belong to NP. A classical representation of problem classification is given in Fig. 2.10.

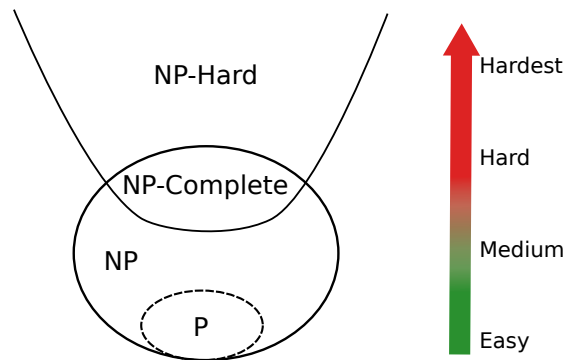


Figure 2.10: Diagram of intersection among classes P, NP, NP-complete and NP-hard problem (assuming $P \neq NP$)

As shown, there exist interconnections between these classes of problems. For instance, every problem in P (those that can be solved quickly) is also in NP (those for which a certificate of the solution exist and can be quickly verified). Yet, an important and unresolved question in computer science is whether P is the same as NP [123]. Resolving this would have profound implications and is one of the seven “Millennium Prize Problems” with a \$1 million reward, though this topic extends beyond the scope of this thesis. Furthermore, a problem is NP-complete if it is both NP-hard and in NP; therefore, it is at least as complicated as any problem in NP, while being itself in NP (assuming $P \neq NP$). This is because verifying a certificate to these problems is easy but finding a solution is hard. Any NP-complete is therefore NP-hard but the reciprocal is false.

In practice, it is not always evident to find the absolute complexity of a problem, researchers therefore show interest in relative complexity from a problem to another. *Reduction* in complexity theory allows to determine the relative hardness of computational problems. Reduction is essentially a way of converting one problem into another, to show that solving the second problem would also solve the first. This implies that if you can reduce problem *A* to problem *B* in polynomial time,

and if an “oracle” or a black-box algorithm can provide a solution for B , then a polynomial-time solution also exists for A . Thus, if A is known to be a hard problem, and can be polynomially reduced to B , denoted as $A \leq_p B$, it suggests that B is at least as hard as A . A standard type of reduction, known as “Karp” reduction, wherein the oracle is invoked once, is depicted in Fig. 2.11.

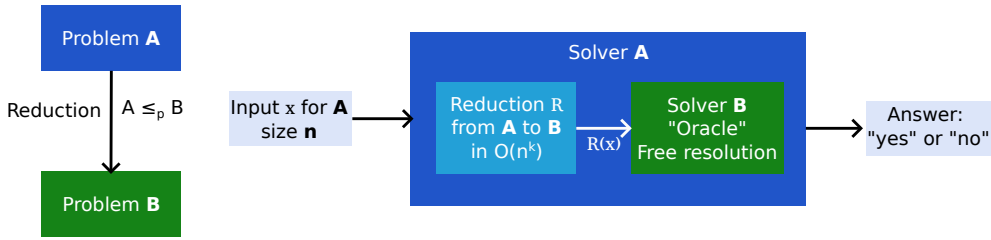


Figure 2.11: Karp reduction: There exists a polynomial time computable function $R: A \rightarrow B$ that transforms inputs to A into inputs to B such that $A(x) = B(R(x))$.

Under the reduction framework, a decision problem is therefore NP-hard if all problems in NP are polynomial time reducible to it. That is, given an efficient algorithm which solves a NP-hard problem, we can construct an efficient algorithm for any problem in NP.

In practice, when attempting to establish, for instance, the NP-hardness of a problem, the first step is to select a problem already formally proven to be NP-complete or NP-hard. Subsequently, one would design a reduction algorithm to transform instances of this known NP-hard problem into instances of the “target” problem, ensuring that this transformation is achievable in polynomial time. If one could efficiently solve the transformed instances of the target problem, it would imply that one could also efficiently solve the instances of the known NP-complete problem. This leads to a contradiction since, by definition, we do not have efficient solutions for NP-complete problems. The target problem is therefore at least as hard as the original NP-complete problem, thereby establishing its hardness.

Knowing the complexity class of a problem is valuable to understand better the computational limits and guide the research efforts. For example, if a problem is proven to be NP-complete, then instead of searching for a polynomial-time algorithm (which might not exist), researchers might focus on approximation algorithms, heuristics, or specialized algorithms that work well for typical instances. Moreover, analyzing and dissecting the problem provides insights into the specific aspects that contribute to its complexity. For instance, in Chapter 3, the proton arc therapy treatment planning problem is analyzed and we establish that the tem-

poral dimension linked to energy layer switching time added profound complexity. This was deduced by demonstrating that, once this temporal aspect was removed, the problem could be distilled down to a tractable LP problem, which is solvable in polynomial time.

2.4 State-Of-The-Art in proton arc therapy

Proton arc therapy (PAT) has emerged as a possible evolution for proton therapy. Drawing inspiration from the arc techniques in photon-based treatments, namely rotational IMRT (IMAT [6] or VMAT [124], TomoTherapy [5], RapidArc [125]), the PAT concept is the delivery of proton radiation from a continuous rotation of the gantry, allowing the patient to be treated from many beam angles. Fig. 2.12 illustrates a comparison to conventional PT which typically deploys just a few fixed-angle beams. This novel delivery technique may offer a higher level of treatment accuracy and effectiveness.

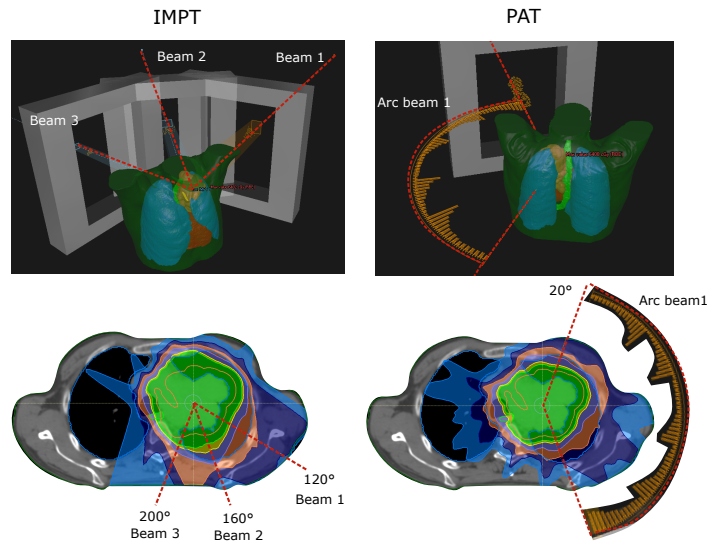


Figure 2.12: Comparison of IMPT versus PAT modality in RayStation TPS.

The origins of PAT can be traced back to 1995 when Deasy et al. proposed the concept of “proton tomotherapy”, where a proton beam rotates in a helical pattern around the patient [126]. They combine the idea with “distal-edge-tracking” (DET), which puts Bragg peaks on the distal edge of the target volume from each irradiation direction. Each individual field is thus highly non-uniform but combining and modulating the intensity of many fields restore the desired dose uniformity across the target volume. This approach aimed to enable the delivery of

highly complex dose distributions, achieving sharp dose gradients. Unfortunately, the lack of available technology and concerns about range uncertainty hindered further development. Shortly after, Sandison et al. demonstrated the feasibility of PAT by delivering a plan using passive scattering to an anthropomorphic phantom that rotated on a platform. Their study showed that PAT could reduce lung doses compared to electron arc therapy when irradiating the phantom chest wall [127]. In 2013, Seco et al. conducted a simulation study comparing VMAT with an in-house-developed arc planning approach for both proton passive scattering and intensity modulated beams in stereotactic body radiation therapy (SBRT) for non-small cell lung cancer. Their findings indicated that PAT had the potential to reduce proton range uncertainties and improve conformality [128]. However, the idea of passive-scattering-based proton arc therapy (PS-PAT) confronted several technical challenges, notably the complexities in changing range modulation wheel settings during gantry rotation and beam delivery. This rendered PS-PAT practically elusive for routine clinical applications. In pursuit of a solution, innovative approaches follow each other for achieving PS-PAT. For instance, in 2014, dynamic beam current modulation combined with a modulation wheel was proposed as a fast range switching technique [129] though it came at the expense of a higher distal dose penumbra and a lower dose rate. Another solution suggested the use of a 3D-printed patient-specific range modifier [130]. Yet, the integration of this technique into clinical settings brought about its own set of uncertainties and demands for quality assurance [131].

Pencil beam scanning's invention [132, 133] was a major breakthrough for PAT when it started to become widely available. This advancement eliminated the need for the patient-specific devices (compensator and aperture), thereby streamlining clinical operations. More importantly, compared to the passive-scattering technique, PBS offers a more conformal dose distribution to target volumes, substantially minimizing the dose exposure to adjacent critical structures. Consequently, the appeal for PAT, utilizing the PBS technique, witnessed a reinvigorated interest. From that moment, research efforts intensified, propelling a renewed exploration into the potentials of PAT but also opening new challenges to face towards its clinical integration.

Initiated in early 2020, this thesis coincides with the development of the proton arc module product by IBA (DynamicARC®). Together, we seek to delve further into the advancements and challenges in the field.

2.4.1 Rationale for proton arc therapy

The motivation to develop PAT stems from the same factors that led to the adoption of rotational IMRT in clinical settings: improved treatment workflow

and larger set of degrees of freedom. For instance, VMAT can adjust field shapes, dose rate, and gantry rotation speed at the same time. This flexibility has shown reductions in monitor units (MU) with subsequent reduction of integral dose and treatment delivery time, offering a potential advantage over conventional fixed-angle field IMRT for several tumor sites [7, 134, 135]. While VMAT usually shows similar outcomes in terms of target volume coverage, dose conformity, and homogeneity, it has often provided better sparing of OAR in specific cases such as prostate and cervical cancers [136, 137, 138, 139, 140, 141, 142].

Extending the arc delivery concept to protons, which are naturally more conformal than photons, PAT holds promise for replicating and perhaps amplifying these advantages, if not for providing even new ones. In pencil-beam scanning proton arc therapy (PBS-PAT), the aim is to provide a proton therapy treatment as the gantry rotates continuously, namely a *dynamic* delivery. This modality might not only simplify the treatment workflow in comparison with standard PBS approach with fixed-angle beams, but it also provides the optimizer with more degrees of freedom. Such flexibility should, in theory, assist the optimizer in identifying a more optimal dose distribution. These extra degrees of freedom may improve the dose distribution and treatment robustness. It may also amplify biological effects by fine-tuning spot placement and weights given protons' reportedly superior radiobiological effect over photons [143]. Consequently, the pursuit for an optimal PAT plan that might further enhance the merits of proton therapy is both logical and promising.

2.4.2 Discrete versus Dynamic PAT

While the *dynamic* approach represents the ultimate objective of PAT, the *discrete* delivery mode was introduced to bridge the technological gap. This approach was designed to enable proton therapy centers to administer PAT treatments with multiple irradiation angles, even if they had not yet acquired the necessary upgrades for the dynamic mode [144]. Both modes share the trait of delivering treatment from numerous directions. However, discrete arcs could be executed either as a single step-and-shoot arc or as individual fixed-angle IMPT beams. As a result, discrete arcs, sometimes termed as “multi-field IMPT” or “static arc fields”, typically have extended treatment delivery time. To maintain clinically acceptable delivery time, we are constrained to use a limited number of irradiation angles, usually ranging from 20 to 40, in contrast to the broader range provided by dynamic arcs. Figure 2.13 offers a comparative illustration of energy layer placements between these two PAT delivery modes.

While both methods offer their unique advantages and considerations, this thesis will predominantly focus on dynamic PAT plans.

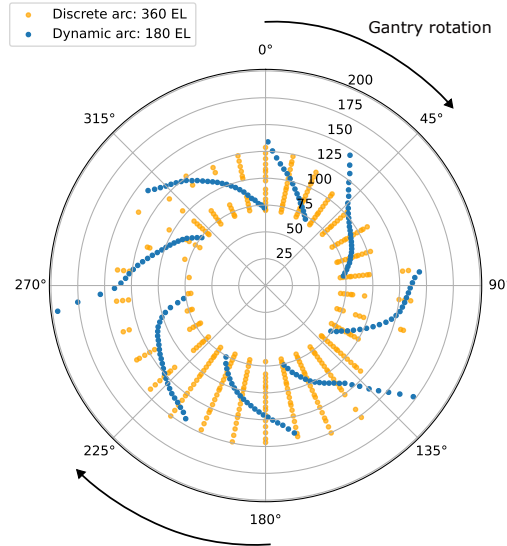


Figure 2.13: Comparison of PAT delivery mode. In this example, *static* arc has 30 irradiation directions and 360 energy layers (EL) while the *dynamic* arc has 180 irradiation directions and 180 EL in total. Radial circles are energy in MeV.

2.4.3 PAT challenges

Bringing proton arc therapy from concept into clinical reality presents several technical challenges that need to be addressed.

Hardware requirements: While discrete PAT might not demand technological upgrades, dynamic PAT certainly does. Presently, IBA leads the effort in advancing a dynamic arc product. The hardware must be upgraded to allow compatibility of PBS with arc delivery and ensure a stable delivery. Essential enhancements include ensuring the PBS controller's ability to process feedbacks efficiently and instruct the gantry controller on arc gantry angles, i.e. synchronize the beam delivery with the continuous gantry rotation. Two main ancillary components have to be developed: the hardware console and an anti-collision system. This console, located in the control room, aims to combine all treatment workflow activities from positioning, position verification (X-ray and CBCT) to irradiation (PAT or IMPT). Safety systems are also being implemented to evaluate clearance before initiating the arc and to halt gantry movement should potential collisions with patients or equipment be detected. The first successful delivery of a dynamic arc, was in fact reported by the Beaumont proton therapy center which received the first IBA prototype upgrade on their proton system to deliver PAT [145].

Quality Assurance (QA): As of today, QA for proton therapy treatments involved a stationary gantry, usually fixed at functional angles like 0° or 90° . However, PAT, with its rotating gantry, brings new parameters such as the beam angle along an arc which should be integrated in the QA procedure. This change necessitates the adaptation of measurement tools and QA protocols suited to a rotating gantry. At Beaumont's proton center, traditional QA tools were mounted on the gantry to measure the performance of the PAT prototype and determine which tools remain relevant for future PAT procedures. However, these setups gave no information towards the irradiation angle uncertainty since they were affixed to the gantry. Another noteworthy development was the exploration of patient QA techniques based on irradiation logs. A dose distribution reconstruction tool has been developed in [OpenReggui \[146\]](#) for the PAT prototype. By extracting the data from irradiation and gantry position logs, this tool can replicate the delivered 3D dose distribution, facilitating a comparison with the planned dose.

Optimization challenge: In radiotherapy, the TPS optimizer is an essential component for generating high-quality plans. The underlying algorithms are commonly regarded as a "black box", allowing the TPS to efficiently run optimizations for various radiotherapy modalities available in the software. However, in the context of PAT, these algorithms cannot be directly applied due to the much larger solution space resulting from the high number of irradiation angles in the arc (typically ranging from a hundred to several hundreds). In the situation where each gantry angle would deliver an IMPT-like beam (each with tens of energy layers to build the SOBP), applying the conventional algorithms may result in an excessively long beam delivery time. This is because of too frequent energy layer and spot switching, which is constrained by technical limitations, thereby introducing delays in delivery. Time is an essential factor in clinical treatment planning, as shorter delivery times enhance patient comfort, minimize treatment uncertainties (e.g. intra-fraction motion), and ultimately increase patient throughput. Another consideration in treatment planning, applicable to all radiotherapy modalities, is the computational power required. Clinicians prefer TPS solutions that are efficient in terms of computing time and memory usage to meet planning demands. However, the large number of beams, energy layers and spots in PAT plans can negatively impact computing efficiency. For PAT to be considered a viable clinical modality, the treatment planning must therefore address three major aspects: (1) the plan quality with respect to important parameters, such as dose and robustness, (2) the delivery time, and (3) the computing efficiency. This can be achieved by reevaluating and refining optimization methods for PAT. Delving deep into these optimization challenges and potential solutions forms the main objective of this thesis.

2.4.4 PAT treatment planning optimization

Beam delivery time models

To harness the degrees of freedom provided by the PAT modality while complying with technical constraints, several optimization strategies have emerged. All these strategies primarily aim to reduce the beam delivery time (BDT) without compromising the quality of the treatment plan. Although the BDT function is a complex function affected by numerous parameters, and differs based on the treatment machine.

Several efforts have been made to address the BDT model for a few PBS-based proton system. Examples include the BDT modeling for the multi-room proton center equipped with HITACHI ProBeat (a synchrotron accelerator) [147] or an IBA ProteusPLUS (a cyclotron accelerator system) [148, 149] and lately an IBA ProteusONE (a compact superconducting synchrocyclotron accelerator (S2C2)) [150]. These machine-specific models usually break the total BDT into three or four main components depending on the accelerator system:

1. Energy layer switching time (ELST): This denotes the time taken to adjust the beam's current energy to the next energy layer. The accumulated times from these switches constitute the total ELST. Due to magnetic hysteresis [151], ELST critically varies depending on the energy step and the energy switching direction (Switch Down (SD) or Switch Up (SU)). It is usually considered as one of the main factor that contributes to the total BDT. It can be modeled as a step function of the energy deviation from its initial value [150]. In this work, we use a constant value of 5.5s for SU and about 0.6 or 0.7s for SD which provides a reasonably accurate estimation of IBA Proteus systems.
2. Spot Switching Time (SWT): This is the time interval required to switch between spots within an energy layer, achieved through the magnetic steering of the pencil beam. Due to varying scanning speeds in different directions and the machine-specific path followed by the pencil beam, the SWT functions can vary but can usually be represented with constant or polynomial piece-wise functions.
3. Spot Scanning Time (SST): SST represents the duration to irradiate a specific spot within an energy layer. Depending on whether the beam is pulsed or not, the fitted function can vary. Nevertheless, nonlinear functions offer the most efficient data fitting

4. Burst Switching Time (BWT): Exclusive to the IBA S2C2, the system divides each spot MU usually into three radiation bursts. This approach can be seen as an adaptive layer repainting technique that relies on feedback from the previous burst. BST then denotes the time interval between each burst. It is modeled as a linear function of the transmission data summation between the two radiation bursts [150].

The total BDT can therefore be obtained by summing all these contributions defined as,

$$\text{BDT} = \text{ELST} + \text{SWT} + \text{SST} (+ \text{BWT}) \quad (2.10)$$

The expression presented in Eq. 2.10 relates solely to the irradiation parameters, the so-called *static* delivery time. This works for IMPT plans but not for PAT plans, which can use a *dynamic* delivery mode. In dynamic PAT, irradiation occurs simultaneously with gantry rotation, introducing several additional mechanical parameters influencing the treatment delivery. These include the gantry's minimum and maximum velocity, acceleration and jerk constraints to approach the next irradiation window and the angular tolerance necessary for spots delivery within the irradiation window. The modeling of the actual delivery sequence results in a highly non-convex objective function, making it resistant to traditional optimization methods employed for dosimetric quantities, mainly suited for convex problems. For these reasons, appropriate surrogate of the BDT is employed, typically a convex approximation of the count of upward energy switches (as discussed in more detail later). In PAT problem statements, the objective is thus to indirectly minimize the BDT through this surrogate.

Energy layer selection problem

When designing a conventional PBS plan, each fixed-angle beam is set up with a SOBP formed by an energy-decreasing sequence of stacked energy layers (EL) and the sum of the beams ensures an uniform dose distribution. Depending on the tumor's specific characteristics, such as size and position, the sequence might consist of anywhere between 20 to 40 EL. Most conventional PBS plans utilize 2 to 4 fixed-angle beams, which cumulatively require around 40 to 200 EL that are available for the treatment planning optimization. When initializing a PAT plan, a similar count of EL per angle can be determined, but this time it is considered for many more directions. This approach linearly increases the number of EL, potentially entering into the thousands. Obviously, delivering all these EL for a treatment would not be time-efficient given the large contribution of ELST to the total BDT. It is good practice to limit the number of layers to be delivered per angle span, and identify a good opening window size for a single layer delivery.

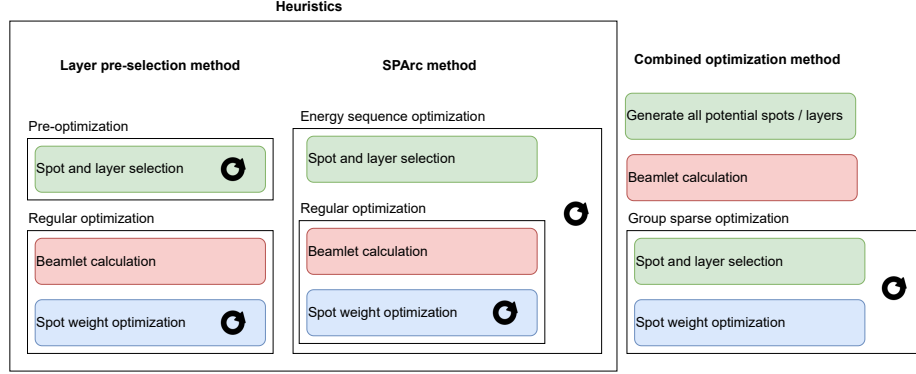


Figure 2.14: PAT optimization methods. Adapted from Kevin Souris.

Moreover, there is an added challenge: initializing the proton arc plan with SOBPs means that energy may increase between each beam delivery, which is more costly than decreasing it. Energy layer sequencing that favors SD over SU is therefore also preferable.

Thus, the goal for PAT planning is to provide the most effective energy layer selection such that it will keep a reasonable BDT and achieve the clinical goals. This is the energy layer selection problem (ELSP) and it constitutes the core of the work in this thesis.

Optimization approaches

At this stage, mainly two types of optimization approaches are under investigation for PAT: *combined optimization* and *heuristic* approaches. Both have been investigated in this thesis. The combined optimization method aims to handle the ELSP together with the spot weight optimization in a single-integrated step handled by a single scalarized function. In contrast, the heuristic method starts with a pre-selection of the EL either before diving into the spot weight optimization or at intervals during the optimization process. This difference in approaches is schematized in Fig. 2.14.

The Spot-scanning Proton Arc Therapy (SPArc) algorithm, as introduced by Ding et al. [17], uses an iterative heuristic approach to produce robust PAT plans. Starting with an IMPT plan that has a low number of discrete beams, it then splits each beam into two sub-beams, and energy layers are redistributed to these new sub-beams. This process repeats until the energy layers have the desired spacing. The energy layers are, in a sense, pre-selected from the initial IMPT plan. Every time a beam splits into two new irradiation directions, a dose

calculation is required, and spots need to be reoptimized. An iterative filtering step is applied after the last beam splitting to remove low MU weighted layers, necessitating spot-weight optimization after each filtering process. A more recent version of SPARc include energy sequencing at each beam splitting step to improve PAT delivery efficiency [18]. In this algorithm version, the initial number of “IMPT” beams to split sets the number of SU in the arc. SPARc was actually implemented using the scripting interface of RayStation v.4.8 (RaySearch Laboratories, Stockholm, Sweden) allowing for straightforward incorporation of robust optimization into the framework. This approach is valued for its simplicity and has demonstrated effectiveness for various tumor sites (see Section 2.4.5). However, beamlet computation is repeated several times, increasing the computational load and there is no guarantee of optimality.

Several heuristic strategies distinctly separate the energy layer selection from the spot weight optimization, saving on computer resources. Early methods suggested techniques such as DET that place more of the Bragg-peak buildup within the tumor, thereby enhancing target conformity [126], or mono- or bi-energetic arcs to improve delivery efficiency by avoiding the energy layer switching time issue [152, 153, 154]. Commercial TPS, such as RayStation or Monaco, have also ventured into robust PAT optimization using heuristics methods. For instance, Engwall et al. [26] developed an early energy layer selection and spot assignment (ELSA), a fast method integrated in a research version of RayStation that involves picking energy layers based on a geometry analysis of the target volume. Following this pre-selection, the regular spot weight optimization occurs. Drawbacks are the fact that no dosimetric information is available at this pre-selection stage and consequently it may present a lack of flexibility in the subsequent dose optimization. Chapter 5 presents the in-house development of a pre-selection method for PAT planning.

The combined optimization method does not prioritize any spots or energy layers initially. Instead, it balances spot weight optimization with energy layer selection, steered by dose-based objectives, in an integrated framework. It potentially converges faster towards an optimal plan while providing certain optimality guarantees. Several variations of this approach were explored in this thesis, leading to its initial publication in Chapter 3. Before the optimization, initial spot and energy selection is done for each discrete direction to cover the target volume in-depth in the same way as in IMPT, thereby pre-computing a large quantity of beamlets. After this, the optimization process includes multiple new objective terms in its cost function, striving to: (1) distribute energy layers throughout the full arc, (2) minimize the number of energy layers per irradiation direction, and (3) order the energy layers to minimize SU occurrences. By setting up the opti-

mization this way, we aim for a solution that results in a sparse and ordered energy sequence, saving on delivery time. The oncology team at UCLA first showcased this simultaneous approach [16]. Building on this foundation, Zhang et al. made further improvements to the process [155, 156]. However, despite its thorough mathematical foundation and flexibility, this method has certain limitations which are detailed in Chapter 3.

In discrete PAT, one faces the challenge of optimizing the total number of energy layers and spots in the treatment plan. The goal is to include enough angles to maximize the arc's therapeutic advantages without adding so many that delivery times become impractical, potentially affecting the overall treatment efficiency. Battinelli et al. presented the energy layer reduction (ELR) method during the spot weight optimization [144, 157]. Their approach included several variations: gradually adding layers based on a gradient score, methodically removing layers considering their spot weights, a hybrid strategy, and a method that integrated a layer-sparsity term into the treatment plan's optimization process. Following this, RaySearch has since incorporated a similar, but less complex, energy layer filtering (ELF) mechanism into the latest released version of RayStation [158]. ELF reduces the count of energy layers from an initial set at regular intervals based on the user-defined number of filtering steps. Cao et al. [159] introduced an alternative, geometry-based method, that selects a subset of energy layers at each gantry angle to ensure adequate coverage of each target voxel, with dose contributions above a specified threshold. These methods are specifically designed to address the challenges of discrete arc optimization, making it more clinically realistic than an IMPT plan with numerous beam angles.

Biologically driven arc planning:

The increased degree of freedom provided by PAT may also enhance the LET distribution, potentially improving the radiobiological effectiveness of protons compared to conventional IMPT planning [160]. This is because protons become what we would consider 'high-LET' particles at their Bragg Peak's distal end. The position and/or intensity of this high-LET components can then be adjusted to enhance the biological impact of protons. Several groups have started to explore this research direction for PAT planning. For instance, Bertolet and Carabe presented the proton monoenergetic arc therapy (PMAT) [161, 152] to enhance the high dose-averaged LET (LET_d) within the target. Their heuristic algorithm, in a pre-optimization step, selects monoenergetic partial arcs during a full gantry rotation to position spots at the middle of a target. The monoenergetic partial arcs are selected in steps. First, the algorithm computes for each beam angle the water-equivalent path length (WEPL) of each target voxel. The WEPL represents

the distance a proton beam would need to travel through homogeneous liquid water to reach the particular voxel in question. This step gives a WEPL list for each angle. The method subsequently determines a range of suitable WEPL values that aim at target voxels with WEPL ranging within $55\% \pm 5\%$ of the WEPL list (to focus on the Bragg's peaks at the central part of the target). The next step is finding which paths or trajectories will successively pass through most of these selected regions. These strategically chosen paths constitutes the final energy layer selection and constitutes the “monoenergetic partial arcs”. The entire process is illustrated in Fig. 2.15. As a result of the energy layer selection criterion, LETd increases within the target, which may translate into a more biologically effective treatment. However, challenges remain, such as ensuring the dosimetric robustness of PMAT plans across various diseases and patient geometries. On the same line, a group recently suggested a blind approach that simply prune high energy spots [162] which could increase both LET and RBE in the target volume and at the same time decreased values in healthy tissue, without affecting the target volume dose coverage. Combined dose and LET optimization in proton arc plans is another avenue explored and has yielded promising outcomes [163, 164]. Through this approach, it is possible to decrease high LET in OAR while preserving robust target coverage. The studies indicate that dynamic arc plans, in comparison to IMPT plans with fewer beams, appear more effective at minimizing LET in OAR. This is because they can employ several beam angles to prevent protons from stopping close to OAR. Still, adjusting the LET in specific areas may impact dose and LET elsewhere, warranting careful clinical evaluation.

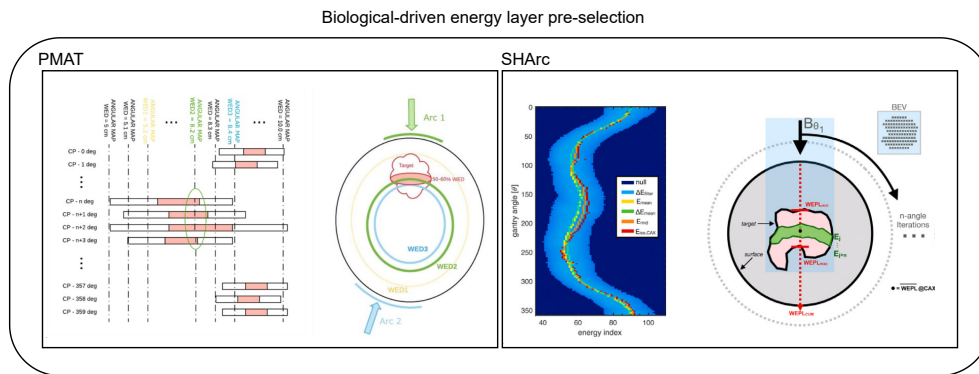


Figure 2.15: PMAT (left) and SHArc (right) energy layer selection workflow. In both cases, the selection is made to place the particle Bragg’s peaks in the middle of the target, thereby increasing LET within the target volume.

Recent research have extended arc therapy to other ion species like helium, carbon, and neon in Spot-Scanning Hadron Arc (SHArc) therapy [165, 166]. In

line with PMAT, SHArc uses criteria for energy selection but does not limit itself to monoenergetic arcs. It evaluates a spectrum of energies, based on the average WEPL at a beam angle's central axis, for every beam angle as illustrated on right side in Fig. 2.15. Energy layers are retained if they intersect the target volume and meet a predefined spot count threshold. This method achieves impressive dose conformity and unprecedented LET boosts within targets. Such LET escalation can potentially tackle radioresistant tumors by overcoming hypoxia [167]. While the potential benefit is substantial, SHArc delivery presents its own set of challenges. This is particularly true in existing clinical facilities, which are all synchrotron-based and feature distinct extraction and acceleration phases of ions, each of a few seconds duration. These limitations can hinder the efficient switching of energy layers to allow a continuous delivery.

2.4.5 Clinical status

Over the past two decades, both in-house and commercial PAT optimization algorithms have been applied in numerous *in silico* planning studies. These investigations aim to understand the merits and drawbacks of PAT, in discrete or dynamic delivery mode, for different treatment sites. Metrics evaluated include robust target coverage, CI, HI, ID, OAR sparing, NTCP, and BDT. The interplay effect, arising from interference between the intra-fraction target motion and active pencil beam scanning, is occasionally analyzed in relevant disease sites.

In the early development stages of PAT, initial clinical planning studies focused on simplistic arc designs, employing passive scattering and intensity modulated proton beams. This was achieved using an in-house-developed arc planning approach, often by selecting an arbitrarily large number of IMPT fields to approximate an arc together with standard clinical TPS tools. These studies presented a more favorable longitudinal dose profile than their photon arc counterpart reducing normal tissue integral dose [127, 168] and predicted risk of radiogenic second cancer following prostate radiotherapy [169]. Further studies indicated that increasing the range of irradiation angles could improve the dose distribution. Benefits included improved target volume coverage, increased conformity, and greater efficiency in sparing nearby critical structures, as opposed to standard methods [170, 128, 171]. However, a common point among these studies was their limitation to *in silico* dose calculations, either in patients or phantoms. Notably absent from the discourse was an exploration into the practicality and feasibility of these proposed solutions, especially challenges like the arc delivery time.

Post-2015, there was a notable trend towards advanced proton arc planning, with a shift in the global market moving towards PBS-based solutions. Sánchez-Parcerisa et al, already recognized for their contributions regarding the feasibility

of PS-PAT [129], further delved into PBS-PAT, exploring both single- and dual-energetic proton arcs, the so-called PMAT approach. They showed that PAT could achieve similar dose coverage and OAR capabilities as full-coverage 2-field and 4-field IMPT plans for a brain tumor and a thoracic phantom [153]. PMAT's potential was also evaluated in a study involving six patients diagnosed with skull base chordoma [172]. In comparison with SFUD and IMPT methods, PMAT seemed to offer relatively similar target CI and HI. Additionally, there appeared to be a reduction in dose for certain DVH metrics. However, the extent to which PMAT planning could be applied to more complex target geometries and ensure proper plan robustness remains to be studied. There is also a scope to investigate deeper into the potential biological advantages associated with PMAT, as hinted in section 2.4.4.

A pivotal development during this period was the SPArc algorithm developed by Beaumont Health group (Royal Oak, MI), permitting dynamic arc robust optimization. They made important contributions in recent years, contrasting the clinical advantages of PAT across various treatment sites through the use of SPArc algorithm. Their exhaustive research covers a spectrum of tumor sites, including prostate [173], non-small lung cancer [174, 175], breast [176], brain [145, 177], spine [178], and head and neck [179]. Table 2.1 summarizes the most relevant data from these studies. Overall, all SPArc studies published to date by the Beaumont team consistently indicates that target coverage, target CI and HI are maintained or sometimes slightly improved compared to IMPT. Additionally, these studies highlight substantial sparing of OAR and clear potential to reduce integral dose without foreseeably extending the treatment delivery time.

Another study comparing PAT to interstitial high-dose-rate (iHDR) brachytherapy for 10 gynecologic cancer patients revealed that PAT could offer a noninvasive alternative with superior dosimetric outcomes [180]. The planning strategy followed began with a multi-field plan involving 9 to 11 IMPT fields which was subsequently converted into an arc plan, similar to SPArc workflow. The method bypassed the energy sequencing, a constraint in SPArc plans design to achieve delivery-efficient plans [18], claiming to use the same incident energy from the accelerator and an external energy modulator [154, 181]. On the downside, this might lead to a slightly larger spot size and possibly a reduction in target CI, even if it wasn't specifically highlighted in the study.

Tumor site	N patients	Modalities compared	Dose metrics	$T_{\text{SPArc}}/T_{\text{IMPT}}^*$	NTCP calculation	Reference
Whole brain	8	IMPT, VMAT	OAR dose +, target HI +, TVC robustness ~, OAR robustness ~	0.9	/	[177]
Head & neck	14	IMPT	OAR dose +, target CI/HI +, ID +, TVC robustness, ~OAR robustness +	1.1	Salivary flow	[179]
Breast	8	IMPT	OAR dose +, target CI ~, HI +, TVC robustness ~, OAR robustness +, interplay effect +	1.1	Heart, lung, skin, ascending artery	[176]
Lung	14	IMPT	OAR dose +, target CI ~, ID +, TVC robustness ~, OAR robustness +, interplay effect +	1.2	/	[174]
Lung SBRT	10	IMPT, VMAT	OAR dose +, target CI/HI +, ID +, robustness +, interplay effect +	0.2**	Chest wall, ipsilateral lung pneumonitis	[175]
Prostate	9	IMPT	OAR dose +, target CI/HI ~, TVC robustness ~, ID +	2.0	/	[173]
Spine SBRS	7	IMPT, VMAT	OAR dose +, ID +	1.0	Radiation myelopathy	[178]

Table 2.1: Clinical studies using SPARC algorithm. (*) Ratio of treatment delivery time assuming a value of 1s for energy layer switching time. (**) SPARC is effective as 5 volumetric repainting IMPT. TVC: Target volume coverage, CI: Conformity index, HI: Homogeneity index, ID: Integral dose, (+): improved; (~): equal. Adapted from Mazal et al [182] and Toussaint L. (ESTRO2021)

As the field has progressed, the industry has begun to invest in the development and evaluation of its own solutions. For example, RaySearch Laboratories introduced ELSA, which offers a different approach to dynamic PAT planning [26]. ELSA pre-selects energy layers before the conventional spot weight optimization. Preliminary results utilizing ELSA delivered mixed outcomes, with certain tumor sites evidently benefiting more than others. For instance, in their introductory proof-of-concept study, ELSA was applied to three prostate cancer patients. While the resultant PAT plans managed to decrease the integral dose in comparison to 2-beam IMPT plans, there were challenges in achieving consistent robust target coverage. Nevertheless, the study's primary objective was to reduce both the optimization and delivery times of PAT plans, ensuring decent plan quality. Our team conducted research using the ELSA prototype with a sample of 17 esophageal cancer patients [183]. By utilizing two partial arcs (with identical yet opposing rotations) and comparing them with 2-beam IMPT plans, our findings suggested that, for this specific tumor site, the PAT approach did not particularly stand out in comparison to standard IMPT, only offering marginal improvements. Nevertheless, the advantages of rotational IMRT plans over conventional 3D conformal radiotherapy were mainly observed in the context of complex-shaped targets. PAT's effectiveness may thus vary depending on the characteristics of the treatment site. For instance, targets in the head and neck region with complex CTV and the need

for conformal avoidance, such as sparing the hippocampus in the brain or addressing multiple spinal metastases, may benefit more from PAT. Conversely, for cases involving circular or simple convex shapes, such as single metastases or esophageal tumors, the advantages of PAT may be less pronounced.

While there is growing accessibility to dynamic arc planning methods, discrete approaches suitable for static gantry delivery continue to be pertinent, especially as we await the full development of dynamic arc in clinical settings. Given this context, several researchers have investigated the clinical merits of discrete arcs across various treatment sites. Toussaint et al, in their 2019 investigation [184], formulated multi-field pseudo PAT plans for four pediatric brain cases, focusing on preserving cognitive function by sparing the temporal lobes. The use of arc beams made this feasible. Although concerns about the increase of the low-dose bath radiation and its implications for secondary primary cancers (SPC), especially in pediatric cases, arose. Building upon their earlier work, the team assessed the radiation-induced SPC risks associated with PAT in five pediatric brain tumor cases [185], sticking to their initial planning approach. Their findings highlighted a noticeable reduction in the integral dose for PAT when compared to VMAT and IMPT, yielding a consistently lower risk estimation for discrete PAT plans. While these results are encouraging, they disregarded technical aspects of PAT optimization and delivery using multi-field IMPT beams with PTV-based optimization instead of robust optimization and should therefore be interpreted with caution. Another investigation, focused on sinonasal cancer, studied a cohort of 24 patients [186]. The study, which employed an 11-field pseudo-arc configuration, indicated noticeable dosimetric advantages for adjacent OAR compared to conventional IMPT with 3–4 beams. Yet, the robustness of the target coverage suffered toward daily anatomical variations. In a different study targeting a group of 10 oropharyngeal cancer patients [157], discrete PAT was found to minimize toxicity when measured against IMPT and VMAT. This research utilized a specialized discrete PAT optimization algorithm, specifically the ELR method [144]. This enabled the selection of the energy layers during the dose optimization, ensuring clinically viable delivery times, provided a step-and-shoot arc delivery. Further research by the same team on a larger cohort of 223 oropharyngeal cancer patients [187] revealed that PAT outperformed IMPT and VMAT. This translated to a statistically significant reduction in NTCP values, thereby increasing the percentage of oropharyngeal patients selected for proton therapy according to the Dutch model based approach for patient radiation therapy treatment referral [188].

In summary, while current findings do not firmly support superior target CI or HI using PAT, there is consistent indication of reduced body integral dose and improved OAR sparing, potentially leading to NTCP reduction. It also shows

potential for improved robustness (range and interplay effects) although this is very algorithm-dependant and not systematically observed. Given the low number of patients in most studies, results should be approached with caution.

Chapter 3

Exploring treatment planning optimization

The content presented in this chapter corresponds to the published article: [Treatment planning in arc proton therapy: comparison of several optimization problem statements and their corresponding solvers](#). Comput Biol Med. 2022 Sep;148:105609.

Authors: Sophie Wuyckens¹, Michael Saint-Guillain², Guillaume Janssens³, Lewei Zhao⁴, Xiaoqiang Li⁴, Xuanfeng Ding⁴, Edmond Sterpin^{1,5}, John A. Lee¹, Kevin Souris^{1,4}

¹ UCLouvain, Institut de recherche expérimentale et clinique, Molecular Imaging and Radiation Oncology Laboratory, Brussels, Belgium. ² UCLouvain, ICTEAM, Louvain-La-Neuve, Belgium. ³ Ion Beam Applications SA, Louvain-La-Neuve, Belgium. ⁴ Department of Radiation Oncology, Beaumont Health, Royal Oak, MI, USA. ⁵ KULeuven, Department of Oncology, Leuven, Belgium.

Summary

Arc proton therapy (ArcPT) is an emerging modality in cancer treatments. It delivers the proton beams following a sequence of irradiation angles while the gantry is continuously rotating around the patient. Compared to conventional proton treatments (intensity modulated proton therapy, IMPT), the number of beams is significantly increased bringing new degrees of freedom that leads to potentially better cancer care. However, the optimization of such treatment plans becomes more complex and several alternative statements of the problem can be considered and compared in order to solve the ArcPT problem. Three such problem statements, distinct in their mathematical formulation and properties, are investigated and applied to solving the ArcPT optimization problem. They make use of (i) fast iterative shrinkage-thresholding algorithm (FISTA), (ii) local search (LS) and (iii) mixed-integer programming (MIP). The treatment plans obtained with those methods are compared among them, but also with IMPT and an existing state-of-the-art method: Spot-Scanning Proton Arc (SPArc). MIP stands out at low scale problems both in terms of dose quality and time delivery efficiency. FISTA shows high dose quality but experiences difficulty to optimize the energy sequence while LS is mostly the antagonist. This detailed study describes independent approaches to solve the ArcPT problem and depending on the clinical case, one should be cautiously picked rather than the other. This paper gives the first formal definition of the problem at stake, as well as a first reference benchmark. Finally, empirical conclusions are drawn, based on realistic assumptions.

Highlights

- Arc proton therapy is a new modality of cancer treatment.
- Arc treatment plans involve long and highly constrained sequences of beams.
- Several statements and solvers of the arc optimization problem are compared.
- The best statement and solver depend on the clinical parameters at stake.
- A benchmark is accessible online to test arc optimization algorithms.

3.1 Introduction

At this time, arc proton therapy (ArcPT) has not yet any commercial available software solution to optimize treatment plans. Moreover, due to the increased degrees of freedom and complexity of optimization along the arc trajectory, generating an ArcPT treatment plan efficiently is one of the hindrances for a future clinical implementation. A treatment plan normally includes all parameters that allow the calculation of radiation dose to be delivered to the patient. A prototype of the ArcPT delivery machine, built by the IBA company, exists and is being studied at the Beaumont clinic in the United States [145] (Fig. 3.2). The finalized product is still under development by IBA. In this study we have developed and investigated three optimization techniques, namely, fast iterative shrinkage-thresholding algorithm (FISTA), local search (LS), and mixed-integer programming (MIP). We compared them with the existing state-of-the-art iterative SPARC optimization framework [17].

Contributions

The main contributions of this work are: **(i)** We introduce the first reference benchmarks for this problem, generated through several processing steps performed on a *realistic* medical image set. The benchmarks are accessible from Open Source Dataverse [189] and can be used freely for testing. **(ii)** Three new optimization algorithms to generate ArcPT treatments plans are proposed. **(iii)** Whereas most existing studies showcase a single plan optimization method in several different disease sites, this is the first comprehensive and quantitative comparison across several existing and new ArcPT optimization techniques.

3.2 Background, challenge and related work

Background

Radiotherapy (RT) treats localized cancer with ionizing radiation, mostly x-ray photons. However, protons offer unique physical properties by stopping in the tumor and releasing most of their energy at the end of their track, known as the Bragg Peak [37]. The interest in proton therapy (PT) delivery has therefore increased. This modality has a clinical potential over RT because radiation dose deposited beyond the tumor can be avoided while at the same time the proximal dose remains low, thereby reducing the radiation toxicity in healthy tissues. Nowadays, state-of-the-art PT treatments are delivered through intensity modulated proton therapy (IMPT) using pencil beam scanning (PBS)[190]. Typically

made up of 2 to 4 *fixed-angle* beams, as illustrated in Fig. 3.1 (left), IMPT delivers each beam with multiple smaller intensity modulated beams [191], called spots, each reaching a specific location of the target. Moreover, to irradiate the target in depth, multiple energy layers per beam angle (also referred to as incidence angle or irradiation angle) are needed. Therefore, an IMPT treatment is delivered spot-by-spot within a same energy layer, layer-by-layer within a beam, and then beam-by-beam.

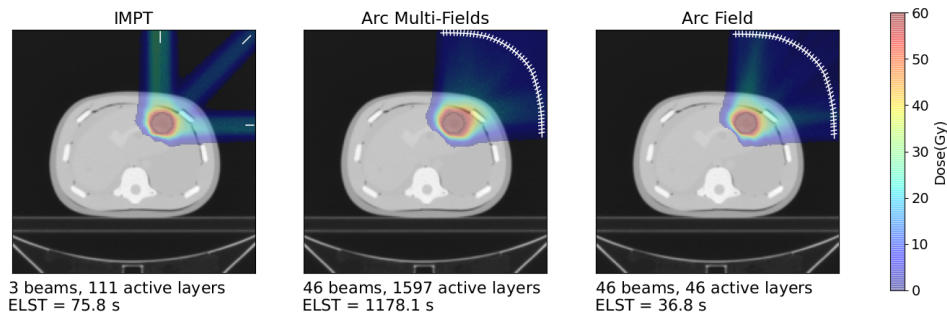


Figure 3.1: PT modality and delivery comparison

ArcPT builds on top of IMPT in the sense that, instead of delivering the treatment with a few *fixed-angle* beams, it uses a gantry that *rotates continuously* around the patient with a variable speed and delivers the treatment following a predefined discrete series of irradiation angles, called the angular control points. An arc is hence defined by a start, stop, and step angles and by specifying the gantry rotation direction. This is depicted in Fig. 3.3. ArcPT is expected to increase treatment conformity and robustness [128] and to achieve better clinical goals in terms of tumor coverage and organs at risk sparing [127].

Challenge

When the number of beam angles is increased (e.g. from 3 to 46) to form an arc and the amount of energy layers in each beam of the arc is kept as numerous as in the IMPT treatment (Fig. 3.1, left), usually in the order of tens, a so-called arc multi-field plan is obtained (Fig. 3.1, center). Unfortunately, this high number of beams (and energy layers) results in a significant increase in the complexity of the optimization problem, computationally demanding in time and memory. Moreover, although conducting to potentially better treatments, the plan obtained on Fig. 3.1 (center) would, above all, not be delivered in practice due to the huge delivery time required to deliver each set of energy layers in each beam which affects the

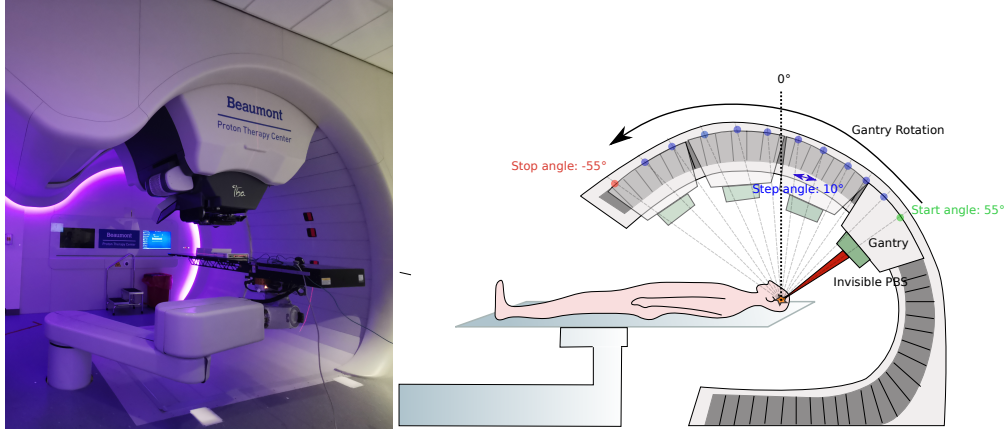


Figure 3.2: Proton therapy Room at Beaumont Hospital with IBA arc prototype

patient's comfort and motion effects over time and directly relates to the number of patients that can be treated in a day. Indeed, the PT system spends a large part of the delivery time to switch the energy between layers, namely, the energy layer switching time (ELST). Therefore, the challenge of ArcPT plan optimization lies in selecting as few energy layers as possible, ideally only one for each beam direction to reduce the ELST and obtain deliverable ArcPT plans (Fig. 3.1, right). Moreover, due to the magnetic hysteresis effect, the delivery system takes more time to increase the energy than decreasing it, which means that the energy switch-downs should be favored over the switch-ups, to further reduce the ELST. The key to obtain such sparsity in energy is thus the selection and the sequence of the energy layers while keeping sight of the dosimetric objectives to get a fully-optimized treatment plan.

State-of-the-art

Today, only a few research groups have investigated this research topic and published methods of optimizing ArcPT plans. For instance, RaySearch (RaySearch Laboratories AB, Stockholm, Sweden), has recently published a proof-of-concept method of energy layer and spot selection, coined ELSA (Early Layer and Spot Assignment), prior to spot weight optimization [26]. They were able to produce high-quality ArcPT plans in a fast optimization process at the cost of losing degrees of freedom in the solution space, caused by these well-separated steps for ELSA and spot weight optimization. Furthermore, an oncology group from UCLA has published some results on a parsimonious algorithm named ELO-SPAT [16],

minimizing an objective function made of several terms. The terms control the selection and sequence of the energy layers, therefore including a delivery time criterion. However, they resort to a non-convex term that can degrade the algorithm performances. Additionally, a group from Beaumont clinic in the US, developed the Spot-Scanning Proton Arc (SPArc) algorithm [17, 192], an iterative greedy approach based on robust optimization (what-if analysis). At the moment, this is the most popular and practical method for its simplicity and efficiency. It has been applied to various disease sites and has shown remarkable results [178, 176, 173, 179, 177, 174]. The major drawback in the SPArc is the greediness of its iterative approach, with an optimization process that can spend much time in refining a solution that might eventually prove to be suboptimal in terms of plan quality. In this paper, three new algorithms that can tackle the ArcPT challenges are presented and compared with each other, as well as with existing state-of-the-art methods, in terms of plan quality, energy layer sparsity, delivery efficiency and, lastly, optimization and dose calculation speed.

3.3 Proposed Benchmark

This section describes the different steps that were performed to generate the ArcPT treatment plans based on a primary image set acquired with Computed Tomography (CT). For this study, a [public, open-access, computed tomography \(CT\) phantom image set](#) collected at MAASTRO Clinic (Maastricht, NL) [193] in DICOM format was elected. In the field of medical imaging, a phantom is a specially designed object or synthetic image created to experimentally evaluate and tune the treatment or imaging workflow. Here, the phantom simulates the abdomen of a small adult.

Generalities about proton therapy Treatment Planning

The design of plans follows a well established procedure made up of three typical stages applicable to all PBS-based PT treatments. They are illustrated in Fig. 3.4 and described hereafter:

- (1) **Spot placement.** This first stage determines, for each beam angle chosen beforehand, a list of energy layers (EL) to cover the target in depth (the higher the proton energy the deeper they penetrate in the target). Then for each EL, a grid of irradiation points (namely spots) is generated. At this step, the plan is thus already defined by the list of spots, with their x and y positions, their energy, and angular incidence.
- (2) **Beamlet calculation.** The patient geometry is discretized in small three

dimensional volume elements called *voxels*. With a particular spot (*i.e.*, within a specific energy layer) is associated a so-called *beamlet*. As depicted in Fig. 3.4, a beamlet describes how the irradiation (elementary dose) from the spot is distributed inside the patient, that is, over the voxels. These beamlets are obtained by simulating the underlying proton physics, using a dose calculation algorithm based on a Monte Carlo engine, MCsquare, developed in-house [72]. This type of algorithm is valued for its accuracy although it can be slow.

- (3) **Spot weight optimization.** This is the core of this work, the spot weights (also referred to as spot intensities) optimization for ArcPT modality. The total dose delivered to the patient can be computed as the sum of all beamlets weighted by the intensity of the corresponding spots (See Section 3.4 Eq. 3.2). Therefore, we only need to optimize the weights of these spots so that the total dose map meets the clinical objectives and constraints. Compared to conventional IMPT plans, new degrees of freedom (the numerous angular positions) and new constraints (treatment delivery time) arise and make the ArcPT optimization problem more complex. New plan optimization methods to determine ArcPT spot weight are implemented in *OpenTPS*, an in-house treatment planning system (TPS).

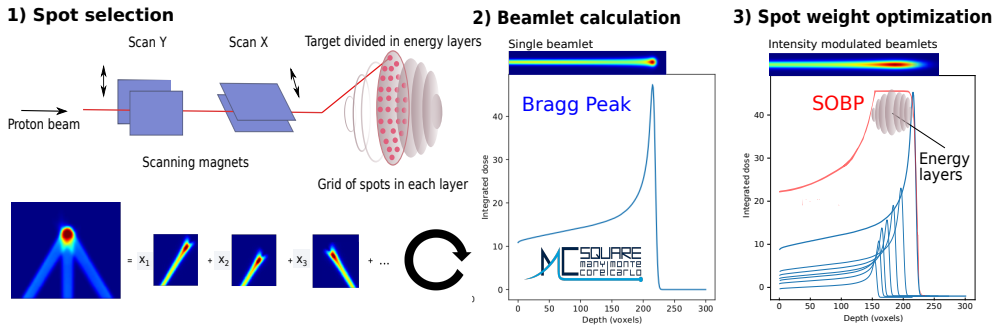


Figure 3.4: Typical PT treatment plan generation steps.

ArcPT benchmark

In our test sets, the beam angles chosen in step (1) were sampled along a quarter of arc (90°) set with a start angle (90°), stop angle (0°), and a patient couch angle (0°), yielding a counter-clock wise gantry rotation. We decided to produce arc plans with 5 different step angles 1° , 2° , 3° , 5° , and 10° , yielding plans with 91, 46, 31, 19, and 10 beams, respectively. The spot placement was then performed for each beam direction [194] by generating multiple hexagonal grids stacked in

depth to fully cover the 3D target volume with spots separated by a pre-defined layer spacing and spot spacing. Prior to computing the adequate spot intensities (*i.e.*, solving the optimization problem), the beamlets were calculated (step (2)) based on the spot placement using the MCsquare dose engine [72]. Benchmarks of different sizes and accuracy were eventually obtained by computing these beamlets under 5 different image resolutions (from 1 to 5mm) using a resampling operation of the DICOM files. A summary of the benchmarks with the time and memory allocated to produce the beamlet files is reported in Table 3.1. All the datasets we produced are publicly available from Open Source Dataverse [189].

CT resolution		1 mm	2 mm	3 mm	4 mm	5 mm
Number of voxels		20,010,375	2,469,888	720,896	304,128	150,176
Step angle	Number of spots	Beamlet computation time and memory				
1°: 91 beams	13,105	5373.3 s 10.5 GB	600.1 s 1.3 GB	185.4 s 395 MB	78.9 s 169 MB	41.5 s 88 MB
2°: 46 beams	6,635	2429.8 s 5.3 GB	297.0 s 668 MB	91.3 s 200 MB	38.5 s 86 MB	20.8 s 45 MB
3° : 31 beams	4,469	1770.8 s 3.6 GB	204.6 s 450 MB	62.0 s 134 MB	27.3 s 58 MB	13.3 s 30 MB
5°: 19 beams	2,622	992.2 s 2.1 GB	117.5 268 MB	36.8 s 80 MB	16.7 s 34 MB	8.4 s 18 MB
10° : 10 beams	1,453	549.8 s 1.2 GB	64.5 s 147 MB	20.1 s 44 MB	8.7 s 19 MB	4.3 s 10 MB

Table 3.1: Benchmarks characterization of beamlets computation time and file size associated with different CT resolution and step angles

3.4 Notations and Problem Definition

3.4.1 Notations

We can now introduce notations that will be useful within our optimization framework. In this paper, the notation used by Gu *et al.* [16] has been slightly modified for the sake of clarity and ease of reading.

- B is the sequence of beams forming an arc, defining the angular control points of the gantry and spaced by a pre-defined step angle; the sequence is ordered by irradiation angle, following the gantry rotation;
- E_b is the sequence of candidate energy layers for beam $b \in B$, ordered by descending energy value;
- S is the set of all the spots defined in the plan;

- S_e is the set of candidate spots for layer $e \in E$;
- $V = T \cup R$ are the sets of target (T) and OAR (R) voxels defined in the patient geometry;
- $\mathcal{A}$ is the beamlet matrix, containing $|S| \times |V|$ elementary dose values (where $|\cdot|$ is the cardinality operator);
- $\mathbf{x}$ is the column vector of spot weights, containing $|S|$ values.

Therefore, the dose map is computed formally as:

$$\mathbf{d} = \mathcal{A}\mathbf{x} = \sum_{j \in S} \mathcal{A}_j x_j, \quad (3.1)$$

$$d_i = \sum_{j \in S} \mathcal{A}_{ij} x_j, \quad \mathcal{A}_{ij} \geq 0, \quad (3.2)$$

where $\mathbf{d} \in \mathbb{R}_{\geq 0}^{|V|}$ is the vectorized dose and d_i is the elementary dose delivered to voxel i . Eq. 3.1 encodes the physics of the treatment, *i.e.*, the description of the dose contribution for each spot, as a linear combination of the beamlets $\mathcal{A}$ (input data) and the spot intensities $\mathbf{x}$ (decision variables).

For practical reasons, we will mostly use a grouped version of $\mathbf{x}$ where each vector component represents the spot intensities within beam b , $\mathbf{x}_b$. The latter is itself a vector where each component represents the spot weights within beam b and layer e , $\mathbf{x}_e$. We also use a compact representation of $\mathbf{x}$ and $\mathbf{x}_b$, $\mathbf{y}$ and $\mathbf{y}_b$ that are vectors with each element representing the sum of weights in each beam and each layer, respectively:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_b \\ \vdots \\ \mathbf{x}_{b'} \end{bmatrix}, \quad \mathbf{x}_b = \begin{bmatrix} \mathbf{x}_e \\ \vdots \\ \mathbf{x}_{e'} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_b \\ \vdots \\ \mathbf{y}_{b'} \end{bmatrix}, \quad \mathbf{y}_b = \begin{bmatrix} y_e \\ \vdots \\ y_{e'} \end{bmatrix}, \quad y_e = \sum_{j \in S_e} x_j, \quad (3.3)$$

where $b, b' \in B$ and $e, e' \in E_b$ and eq. 3.3 can be written as a matrix multiplication in a more compact way:

$$\mathbf{y} = W\mathbf{x}$$

with W , the summation matrix along the spot dimension.

3.4.2 Problem statement, constraints, and evaluation

The problem faced by the physicists in ArcPT treatment planning can be ideally formulated by Program (3.4-3.8). The solution is a vector of spot weights $\mathbf{x}$ that

minimizes the beam delivery time $BDT(\mathbf{x})$ subject to medical constraints, *i.e.* target volumes (TV) coverage and organs at risk (OAR) sparing with minimal, maximal and/or mean prescription and dose limits respectively,

$$\min_{\mathbf{x}} BDT(\mathbf{x}) \quad (3.4)$$

$$\text{s.t.} \quad \forall k \in \text{TV}, \forall i \in T_k : \quad p_{\min T_k} \leq d_i \leq p_{\max T_k} \quad (3.5)$$

$$\forall k \in \text{OAR}, \forall i \in R_k : \quad d_i \leq p_{\max R_k} \quad (3.6)$$

$$\forall k \in \text{OAR} : \quad \frac{1}{|R_k|} \sum_{i \in R_k} d_i \leq p_{\max \text{mean} R_k} \quad (3.7)$$

$$\mathbf{x} \in \mathbb{R}_{\geq 0}^S \quad (3.8)$$

with $|T_k|$ and $|R_k|$, the total number of voxels associated with TV k and OAR k and $p_{\max R_k}$ stating a maximal irradiation for any voxel i in a given OAR k , for example the maximal dose in the spinal cord must not exceed 45 Gy [195].

However, plan optimization methods in the current literature fail at strictly considering clinical constraints (3.5)-(3.7), leading to OAR being irradiated over the acceptable maximal (resp. average) dose constraint (3.6) (resp. constraint (3.7)) if, for example, they are in direct contact with the tumor. Instead, existing methods, usually based on gradient-descent (continuous) optimization, approximate them with soft constraints $f(\mathbf{d})$ that penalize an overall (alternative) objective function:

$$\min_{\mathbf{x}} BDT(\mathbf{x}) + f(\mathbf{d}) \quad (3.9)$$

$$\text{s.t.} \quad \mathbf{d} = \mathcal{A}\mathbf{x} \quad (3.10)$$

$$\mathbf{x} \in \mathbb{R}_{\geq 0}^{|S|} , \quad (3.11)$$

where $f(\mathbf{d})$ is a penalty function evaluating the quality of the simulated treatment dose *w.r.t.* constraints (3.5)-(3.7), formulated as a weighted sum of multiple clinical goals depending on the priorities of the regions of interests (ROI, composed of both TV and OAR volumes):

$$f(\mathbf{d}) = \sum_{k \in \text{ROI}} w_k f_k(\mathbf{d}) . \quad (3.12)$$

Depending on the clinical objective, $f_k(\mathbf{d})$ is typically formulated as a squared l_2 -norm of the difference between the dose $\mathcal{A}\mathbf{x}$ and prescription $\mathbf{p}$. In practice, multiple clinical goals are encoded using weighted soft objectives to constrain the minimum, maximum, or mean doses. This way two slightly offset one-sided quadratic functions can be used to constraint the dose within a band to which we

can give either low or large importance weight. The total cost function can then explicitly written as:

$$f(\mathbf{d}) = \sum_k \frac{w_k}{|T_k|} \sum_{i \text{ in } T_k} \min(0, d_i - p_{\min T_k})^2 \quad (3.13)$$

$$+ \sum_k \frac{w_k}{|T_k|} \sum_{i \text{ in } T_k} \max(0, d_i - p_{\max T_k})^2 \quad (3.14)$$

$$+ \sum_k w_k \max(0, \frac{1}{|R_k|} \sum_{i \text{ in } R_k} d_i - p_{\max \text{mean } R_k})^2 \quad (3.15)$$

$$+ \dots \quad (3.16)$$

Program (3.9-3.11) now properly describes the ArcPT problem and shows a real trade-off between the smooth part dealing with the clinical goals $f(\mathbf{d})$ and the non-smooth part minimizing the beam delivery time $\text{BDT}(\mathbf{x})$. Because a real delivery time model will be certainly not convex and hardly differentiable, we choose to optimize the energy sequence in the absence of a better model of the BDT. Indeed, the BDT is predominantly driven by the energy layer switching time (ELST) in ArcPT therefore we use the ELST as a surrogate to the BDT in our optimization problem. In order to simplify the model, we assume that it takes 5.5 seconds to switch the energy up versus 0.6 seconds to switch it down [192]. In this study, we thus assume the ELST to be the sum of all individual switching times between the active energy layers (layers with non-null spot weights) at each angular control point along the arc.

Evaluation

To assess a plan and score its quality, the physicists compute dose-volume histograms (DVH) and the associated DVH metrics for both TV and OAR. A DVH (see for example Fig. 3.10) shows a decumulative distribution of dose within some given ROI. A point on the curve represents the percentage of the considered ROI that receives a dose less than or equal to the dose value corresponding. As we want to irradiate the tumor volume, the ideal curve is a step from 100% to 0% at the prescribed dose to avoid under- and over-dosage. On the other hand, we want to spare healthy organs. This translates as a ideal step from 100% to 0% at null dose. Obviously, in practice, actual DVH curves generally look like smoothed steps. In addition to DVH curves, DVH metrics further summarize the dose map information into a few scalar indicators. For instance, $D_{95\%}$ represents the minimal dose received by at least 95% of the considered volume (usually a target volume).

3.5 State-of-the-art methods

3.5.1 SPArc

Developed by Ding *et al.* [17], SPArc is a greedy heuristic planning optimization algorithm that uses two nested iterative loops to optimize the dose distribution while increasing the number of control points to make it denser until a desired arc delivery sampling frequency is obtained. It starts with a sparse arc, *i.e.*, with very few irradiating directions. Subsequently, the nested loop optimization algorithm can start. The spot weights are optimized for the current beam-layer configuration. Four steps follow:

1. Control point resampling: each beam is split into two or more new irradiation directions
2. Energy layer redistribution between the new sub-control points
3. Energy layer filtration to get rid of low weighting energy layers

At each iteration of optimization, steps (1) to (3) are carried out until the angle sampling frequency reaches the user-defined parameter. At this stage, SPArc ensures acceptable delivery time by designing plans with 1 to 3 active layers per control point. However, the energy switching pattern is still arbitrary. SPArc has therefore been improved [192] to re-distribute the energy layers in a more efficient way, namely, in descending order to mitigate the impact from energy-switch upward on the delivery time.

Because of its simplicity and proven potential in several studies [176, 173, 179, 174] this algorithm is appreciated in the community. However, its heuristic and greedy character makes it less efficient in the treatment plan optimization process. Additionally, the spot weights and the energy sequence are not optimized jointly but as separate steps and this may result in sub-optimal delivery efficiency and plan quality. Finally, improvement of plan quality or delivery efficiency of the final treatment plan may require a reset of the initial sampling process.

3.5.2 ELO-SPAT

Gu *et al.* developed the algorithm ELO-SPAT [16]. It solves the optimization problem by including a regularization strategy, to address the ArcPT non-smooth problem stated in Eq. 3.9-3.11. In particular, it uses the proximal gradient method (PGM)[196], a popular technique to solve non-smooth optimization problems for image reconstruction. However, traditional PGM methods tend to be quite slow

on convergence, so an accelerated proximal gradient scheme is implemented in the computation. The algorithm chosen is the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [106]. The ELO-SPAT framework is formulated with an objective function made of several terms to meet all ArcPT constraints. Based on notations introduced in Section 3.4, Eq. 3.17 gives an intuition on the ELO-SPAT objective function:

$$\begin{aligned} \arg \min_{\mathbf{x}} \|\mathcal{A}\mathbf{x} - \mathbf{p}\|^2 + \sum_{b \in B} \sum_{e \in E_b} \alpha_e \|\mathbf{x}_e\|_2^{1/2} - \beta \sum_{b \in B} \log \left(\sum_{e \in E_b} y_e \right) + \gamma \sum_{b \in B} h(\tilde{\mathbf{y}}_b) \\ \text{s.t. } \mathbf{x} \geq 0, \mathbf{y} = W\mathbf{x} \end{aligned} \quad (3.17)$$

The first term, namely, the dose fidelity term, encodes the objectives of the treatment. It is a soft-constraint version of constraints (3.5)-(3.7), where the OAR volumes are naturally constrained with a zero dose. It implicitly contains a weighted sum of several objectives (Eq. 3.12), each being a squared l_2 -norm between the prescription to tumor (and dose limits to OAR) and voxels of the total dose distribution estimate. The second term is a group sparsity regularization defined by a non-convex $l_{2,1/2}$ -norm that induces a sparsity at the level of the energy layers, *i.e.* to obtain as few energy layers per beam as possible resulting in a reduced ELST. This term is weighted by the hyperparameter α_e and is responsible to turn off non-interesting layers (low weight in layer). The third term represented by a log barrier behaves as a counter-acting term to the sparsity term in order to distribute the selected layers across the whole gantry rotating range, a heuristic for obtaining better arc-shaped treatments. Finally, in the fourth term, a regularization function is used to sequence the energies, in a way that tends to minimize the delivery time (Eq. 3.4). This function asymmetrically penalizes energy switching low-to-high harder than high-to-low, using a one-sided quadratic cost function h that acts on the energy differences between adjacent beams where the layer with the maximal spot intensity $\tilde{\mathbf{y}}_b$ has been selected beforehand. In terms of workflow, the algorithm starts from all possible EL's candidates, select as few as possible and then encourage energy switch from high to low.

This algorithm raised a lot of attention because it is written in a very well defined mathematical framework. ELO-SPAT converges towards the global or local minimum (due to non-convex term) faster compared to SPARC from a mathematical point of view. However, there is still room for improvement, in particular, in the formulation of some objectives. Firstly using a non-convex group sparsity term together with FISTA is not proper given that FISTA has been designed to work with convex objective functions. Next, the log barrier is a really strong term for the intended goal since it uses a continuous approximation of a penalty that should

be null on a large domain. The penalty starts to increase too early and might shift the optimum where this is not necessary. Finally, for the last term, they use specially designed discrete gradient operators that need to be updated after every iteration which is again not ideal for FISTA convex optimization. We worked to improve the formulation of the ArcPT cost function based on ELO-SPAT.

3.6 Proposed plan optimization methods

In this section, we present three new, fundamentally different plan optimization methods. One is an improved variant of ELO-SPAT, thus based on gradient descent. The second method is a local search approach, based on simple neighborhood operators and a Simulated Annealing meta-heuristic. The third method is a Mixed-Integer Programming formulation. As discussed later in Section 3.9, these three methods come with significantly different mathematical properties in terms of optimality, treatment guarantees, and computational effort.

3.6.1 FISTArcPT

FISTArcPT (Eq. 3.18) is a modified version of the ELO-SPAT objective function formulation:

$$\begin{aligned} \arg \min_{\mathbf{x}} f(\mathbf{d}) + \lambda \sum_{b \in B} \sum_{e \in E_b} \alpha_e \|\mathbf{x}_e\|_2^1 + \sum_{b \in B} I_C^b(\mathbf{x}) \\ + \gamma \sum_{b \in B^-} \text{LeakyReLU}(D_b^B \text{LSE}(\mathbf{z}_b \cdot \tanh(a \cdot \mathbf{y}_b))) , \\ \text{s.t. } \mathbf{x} \geq 0, \mathbf{y} = W\mathbf{x} , \end{aligned} \quad (3.18)$$

where B^- represents the set B of beams, without the last beam of the sequence B . Each term of Eq. 3.18 is described by order of appearance, from left to right, below in details.

1. Dose fidelity

The dose fidelity term $f(\mathbf{d})$ packs multiple objectives into a sum of weighted soft objectives as described in Eq. 3.14-3.16. This term is the core of the treatment plan optimization and occurs in any RT or PT problem formulation.

2. Group sparsity

For the second term of Eq. 3.18, we use the group LASSO [197], that is, the sum of squares of coefficients belonging to the same group and, in this case, the weights belonging to the same energy layer. It is embodied

in a weighted $l_{2,1}$ -norm, contrasting with the non-convex $l_{2,1/2}$ -norm used in the original ELO-SPAT method. The main reason is that the $l_{2,1}$ is convex (Fig. 3.5), which fulfills the requirements of the FISTA algorithm. Moreover, the proximal operator is simpler to implement and still efficient for our purpose, the deactivation of layers. The inner l_2 -norm of $\|x_e\|_2^1$ enforces selection of all the weights within an energy layer while the outer l_1 -norm promotes sparsity in the number of selected energy layers. It is controlled by two parameters, namely, the constant λ set at the beginning of the optimization and α_e that is updated every ten iterations to follow the spot weights update, and further accelerate the layer sparsity and defined as:

$$\alpha_e = \begin{cases} \frac{1}{y_e} = \frac{1}{\sum_{j \in S_e} x_j} & \text{if } y_e \geq 0, \\ 1 & \text{otherwise} \end{cases} . \quad (3.19)$$

With proper tuning of λ , this term selects fewer layers out of all candidate layers.

3. Indicator function

The group sparsity term, if used solely, can result in a solution with some beams having multiple layers activated and some with none, which reduces the arc geometry benefits. To counteract this effect, we implemented an indicator function I_C defined in Eq. 3.20, as an alternative to the log barrier in ELO-SPAT:

$$I_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{if } x \notin C \end{cases} , \quad (3.20)$$

where the admissible domain C is defined as

$$C = \{y_e : \sum_{e \in E_b} y_e \geq \epsilon\} .$$

In practice, the function operates if the sum of weights in a beam is less than the tunable threshold, ϵ . If so, it projects the weights of active layers belonging to the beam under consideration such that the sum of weights in these layers is exactly ϵ . If no layer is active in the beam, it projects directly the weights of the layer with the minimal gradient on ϵ . This is a smooth way to ensure that no zero intensity beam remains and that a certain homogeneity in the arc is kept. By combining the group sparsity term with the indicator function, on average one layer per beam is selected. However, the energy sequence is not yet optimized.

4. Energy sequencing

To encourage energy switches from high to low and, therefore, to minimize the treatment delivery time, we devised a differentiable regularization term controlled by the weighting parameter γ . A hyperbolic tangent is first used to map the sum of intensities of each layer between 0 and 1 with a a scaling factor tuned for that purpose. Next, this vector is multiplied element-wise by z_b , the vector of energies corresponding to each layer. A log-sum-exp operator is then used as a soft maximum to select the most eligible energy layer in each beam. After this, a discrete gradient operator D_b^B is applied to each beam. For example, $D_b^B \mathbf{y}_b = \mathbf{y}_{b+1} - \mathbf{y}_b$, i.e., the vector of the intensity difference between the two beams considered. Finally a leaky ReLU function is used to asymmetrically penalize the energy switching pattern in the following way:

$$f(x) = \begin{cases} 0.01x & \text{if } x < 0 \text{ (switch-down)} \\ x & \text{else (switch-up).} \end{cases} \quad (3.21)$$

Optimizer

The cost function is non-differentiable due to $l_{2,1}$ -norm and indicator function. For this reason, we decided to use the FISTA optimizer, as in the original ELO-SPAT method, which converge theoretically at an optimal rate of $\mathcal{O}(1/k^2)$ in the case of a single non-smooth function¹[106]. The FISTA optimizer engine was implemented in *OpenTPS* with the Python package PyUNLocBoX [198] that is based on the original FISTA algorithm [106]. The pseudo code of FISTArcPT is described in Algorithm 1 and mainly consists in the forward step performing the gradient descent (dose fidelity and energy sequencing) followed by the backward step achieved through the proximal operator of the non smooth part of the objective function, i.e., the group sparsity term. While proximal operators are applied sequentially in Algorithm 1, they should be implemented slightly differently to deal with more than $n = 1$ non-smooth function [199]¹.

¹Edited from the published version

Algorithm 1 Pseudo code for FISTArcPT algorithm

```

1: Input: Initial weights  $x_0$ , step  $\alpha \geq 0$ ,  $\lambda$ ,  $\epsilon$ ,  $\gamma$ 
2: for  $i = 1$  to  $N$  do
3:    $x_k = \text{update\_sol}(x_k)$  // FISTA acceleration
4:    $\alpha = \text{update\_step}(x_k)$  // Backtracking
5:    $x_{k+1} \leftarrow x_k - \alpha \cdot \nabla f(x_k)$  // Forward step
6:    $x_{k+1} \leftarrow I_{\geq 0}(x_{k+1})$  // Positive projection
7:    $x_{k+1} \leftarrow \text{prox}_{\lambda \|\cdot\|_2^1}(x_{k+1})$  // Backward step
8:    $x_{k+1} \leftarrow I_C(x_{k+1})$  // Indicator II step
9: end for
    
```

Convexity

As mentioned in Subsection 3.5.2, FISTA has been developed to solve convex optimization problems. By analyzing each term of the FISTArcPT problem statement (Eq. 3.18), we can check the convexity of our objective function. Firstly, the fidelity term involves soft objectives in the form of half-quadratics that are convex functions. By definition, the sum of convex functions is also convex, so is our total dose fidelity term. The second term is the group sparsity $l_{2,1}$ -norm purposely chosen for its convex nature, illustrated in Fig. 3.5. The third term, the indicator function, is convex by observation. Fig. 3.6 exposes in a simplified schema the convexity of the set C .

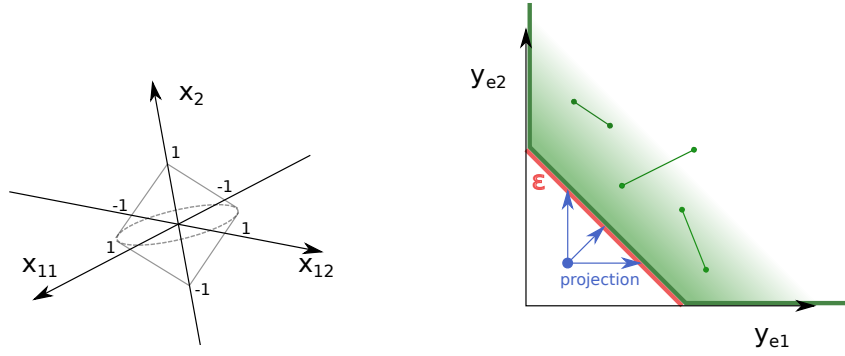


Figure 3.5: Group Lasso penalty: $\|x_1\| + |x_2| = 1$

Figure 3.6: Simplified representation of set C and its convexity. Any pair of points (green) $\in C$, can be connected by a line that lies in C .

We were not able to analytically determine the convexity of the last objective, the energy sequencing (ES) function. Instead, a Monte Carlo simulation suffices

to prove the non-convexity of the ES function. Given that the energy sequencing term is a sum of functions over the different beams of the plan, the non-convexity of only one term in the sum needed to be proved. For some beam b , we evaluate the function $\text{LeakyReLU}(D_b \text{LSE}(z_b \cdot \tanh(a \cdot \mathbf{y}_b)))$. For one evaluation of this function, at minimum four variables were needed due to the discrete nature of the gradient that requires two beams (b and $b + 1$) to operate and also to the fact that $\mathbf{y}_b$ is a vector containing at least two variables. Jensen's inequality [200] was used as convexity criterion requiring the function to be evaluated twice. The Monte Carlo simulation therefore needs $(N^4)^2 = N^8$ input variables. Running the simulation during one hour that generates $N = 1000$ points per iteration adding up to 23673 iterations, we computed the proportion of non-convex points to be about 35.7%. Unfortunately, a convex form of the energy sequencing term could not eventually be found. This conclusion leads to the fact that the optimization problem as stated in Eq. 1 is non-convex. Therefore, the optimization algorithm FISTA is likely to converge to local minima and suffers in its performances.

3.6.2 Local Search optimizer

Alternatively to the gradient-based methods, we propose a simpler method based on the local search (LS) algorithm, described in Algorithm 2. It implements a Simulated Annealing [114] meta-heuristic for approximating the optimal solution x^* , minimizing

$$f(x) = f(\mathbf{d}) + \alpha \text{ sparsity} + \gamma \text{ ELST} . \quad (3.22)$$

Starting from an initial incumbent (*i.e.*, current) solution x , Algorithm 2 iteratively modifies it by using a set of five neighborhood operators. At each iteration, it randomly chooses a solution x' in the current neighborhood $\mathcal{N}(x)$ (line 4) and either accepts it and resets the distance counter d to zero (line 6) or rejects it and increments d (line 11). Whether a candidate solution is accepted as the new incumbent solution or not depends on a predefined *acceptance criterion* function, described below, which implements the so-called *Simulated Annealing*. A restart strategy is also implemented in line 14: whenever the current incumbent solution becomes too distant (in terms of local changes) from the last best solution x^* , the search simply restart from x^* . At the end, the algorithm simply returns the best solution x^* encountered so far, which is stored at line 8.

Initial solution

If not specified, the initial solution is constructed by simply assigning a zero value to every spots. Otherwise, any possible value assignment of the spots can be considered as an initial incumbent solution for the LS algorithm.

Algorithm 2 Local Search algorithm

```

  Let  $x = \{x_s : s \in S\}$  be an initial solution.
2: Set  $x^* \leftarrow x$ ; Set the distance counter  $d$  to 0
   while some stopping criterion is not met do
4:   Select a solution  $x'$  uniformly at random in  $\mathcal{N}(x)$ 
     if some acceptance criterion is met on  $x'$  then
6:       set  $x$  to  $x'$  and  $d$  to 1;
       if  $f(x') < f(x^*)$  then
8:          $x^* \leftarrow x'$ 
       end if
10:    else
         $d \leftarrow d + 1$ 
12:    end if
     if  $d = d^{\max}$  then
14:       restart: set  $x \leftarrow x^*$ 
     end if
16: end while
   return  $x^*$ , the best solution found

```

Stopping criterion

The stopping criterion designed here simply depends on the computational time budget that the user allocates to the algorithm. Once the allotted time elapsed, the criterion is met. Notice that alternative stopping criteria may be devised, based on convergence indicators, for example: relative quality of the last improvements, delays between previous improvements, etc.

Neighborhood operators

We consider five neighborhood operators in $\mathcal{N}(x)$: **i)** select a spot at random and set its value to zero; **ii)** select a spot at random and assign a random positive value in $]0, M]$, where M is the highest possible intensity; **iii)** select an activated layer at random and set all its spots to value zero; **iv)** select a layer at random and reassign random positive values to all its spots; **v)** find the target voxel for which the total dose is the farthest from the prescription, select one of its contributing spots at random and assign it the (non-negative) value that minimizes the dose difference from that voxel to the prescription. At each iteration, the local search algorithm simply selects an operator at random. We notice that classical improvement avenues for such neighborhood based optimization method include an adaptive mechanism that selects the current operator probabilistically, according

to its success rate observed so far, preferring operators that contributed at generating relatively many interesting solutions, that is, that passed the acceptance criterion. Such a mechanism is, however, typically implemented whenever the neighborhood operators include more complex and computationally demanding operations (e.g, large neighborhood search, [201]).

Incremental computation

An important property of the proposed set of neighborhood operators is that given a current solution x , any neighboring solution $x' \in \mathcal{N}(x)$ is actually constructed incrementally from x . In particular, the new objective value $f(x')$ is also incrementally computed provided $f(x)$. In all our operators (**i** to **v**), both computations of x' and $f(x')$ are done in $\mathcal{O}(|S|)$ linear time, relatively to the number $|S|$ of spots. Incremental computation and evaluation is in fact a necessary property of any efficient local search approach. On the contrary, every iteration of gradient based methods such as SPArc, ELO-SPAT or FISTArcPT, involve the heavy matrix-vector product Ax of worst-case $\mathcal{O}(VS)$ complexity.

Acceptance criterion

We use a Simulated Annealing (SA) acceptance criterion [114]. Improved solutions are always accepted, while degraded solutions are accepted with a probability that depends on the degradation magnitude and a temperature parameter, *i.e.* the probability of accepting x' is $e^{-\frac{1-f(x)/f(x')}{T}}$. The temperature T is updated by a *cooling factor* $0 < f_T < 1$ at each iteration of Algorithm 2: $T \leftarrow f_T \cdot T$. During the search process, T gradually evolves from an initial temperature T_{init} to nearly zero. We also implement a *restart strategy*: each time the current temperature T decreases below a fixed limit T_{min} , T is reset to T_{init} . In all experiments, SA parameters were set to $T_{\text{init}} = 5$, $T_{\text{min}} = 10^{-6}$, and $f_T = 0.8$.

3.6.3 Mixed-Integer Programming Formulation

An exact alternative to approximate optimization methods as presented above is to use a MIP formulation and an appropriate solver, allowing a complete search of the solution space to be conducted in order to prove optimality. Furthermore, the treatment requirements can be guaranteed, as hard constraints, instead of using soft constraints with minimized violations, as it is often the case with gradient-based methods, hence approximating the initial (*i.e.*, ideal) treatment problem stated in (3.4)-(3.8).

Energy sequencing

We represent a specific sequence of energy layers (EL) as a path in a directed graph, from a source node 0^- which has outgoing directed edges $(0^-, n)$ to every EL $n \in E$, to a sink node 0^+ which has incoming edges from every EL, passing by a network of edges that link the EL altogether: there is an edge (m, n) for each pair of EL such that n belongs to a strictly higher beam than m . We note E^{edges} the set of all edges that constitute the graph, whereas $E^- = E \cup \{0^-\}$, $E^+ = E \cup \{0^+\}$, $E^\pm = E \cup \{0^-, 0^+\}$ denotes the nodes. Remark that the digraph resulting from E^{edges} allows for valid energy sequencing only. In particular, a unique path from 0^- to 0^+ enforces that *at most one* EL is activated in each beam; otherwise, the path would contain forks. Finally, a non-negative value c_{mn} gives the cost of using edge $(m, n) \in E^{\text{edges}}$ in the path, that is, the cost of switching from layer m to layer n , as illustrated in Fig. 3.7.

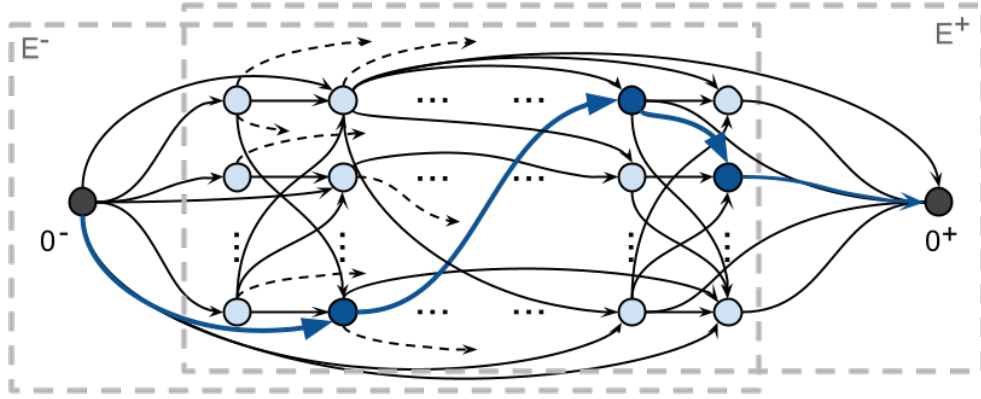


Figure 3.7: The directed graph resulting from the E^{edges} edges construction. Source and sink nodes are represented as the left and right circles. A path in the digraph is necessarily a valid sequence of activated energy layers. In this simplified example, bold edges represent a possible unique path (from source node 0^- to sink node 0^+). in which the dark blue nodes are the activated ELs, and where the first beam has no activated EL.

Linear model

Let x_j and e_{mn} be decision variables such that

- $\forall j \in S$: x_j determines the intensity attributed to a spot $j \in S$,
- e_{mn} are redundant variables determining whether some EL switch happens:
 $\forall m, n \in E^\pm$: $e_{mn} = 1$ if $(m, n) \in E^{\text{edges}}$ and layer $n \in E^+$ is activated

directly after layer $m \in E^-$, otherwise $e_{mn} = 0$.

Then, the arc proton therapy treatment optimization problem can be modelled as a mixed-integer two-index flow formulation

$$\min_{x_1, \dots, x_{|S|}} \sum_{i \in R, j \in S} \mathcal{A}_{ij} x_j \quad (3.23)$$

$$\text{s.t.: } \forall i \in T \quad p_{\min T} \leq \sum_{j \in S} \mathcal{A}_{ij} x_j \leq p_{\max T} \quad (3.24)$$

$$\sum_{n \in E^+} e_{0^-n} = \sum_{n \in E^-} e_{n0^+} = 1 \quad (3.25)$$

$$\forall m \in E \quad \sum_{l \in E^-} e_{lm} = \sum_{n \in E^+} e_{mn} \quad (3.26)$$

$$\sum_{(m,n) \in E^{\text{edges}}} c_{mn} e_{mn} \leq U \quad (3.27)$$

$$\forall m \in E, \forall j \in S_m \quad x_j \leq I \sum_{l \in E^-} e_{lm} \quad (3.28)$$

$$\forall j \in S \quad 0 \leq x_j \leq I \quad (3.29)$$

$$\forall (m, n) \in E^{\text{edges}} \quad e_{mn} \in \{0, 1\} \quad (3.30)$$

The objective function (3.23) minimizes the total dose on the OAR voxels, *provided* (3.24) each target voxel $i \in T$ receives its prescribed treatment, here in the range $[p_{\min T}, p_{\max T}]$, *and* (3.27) the total energy sequencing cost does not exceed a predefined bound U . In order to stick with the previous methods, we assume $c_{mn} = 1$ when m is of lower energy than n , 0 otherwise, so that the total summation simply gives the number of EL switch-ups. The set of constraints (3.25)-(3.26) are classical flow conservation constraints, ensuring that the selected edges from E^{edges} form a unique path (without fork) from source 0^- to sink 0^+ .

The relation between each energy layer and its associated spots is defined by constraints (3.28). A positive constant I represents the highest possible intensity assigned to any spot. In fact, a spot j for which the energy layer m is not activated (*i.e.*, is not part of the flow) must necessarily be assigned $x_j = 0$. Finally, the aforementioned decision variables are defined in (3.29)-(3.30), where we notice that contrarily to x 's, the flow variables e_{mn} are binary integers.

Benefits and lessons from the MIP problem statement

In contrast to the previous methods, the exact MIP statement provides *guarantees* on the optimized treatment plan that will fulfill the constraints. In fact, any feasible solution for constraints (3.24-3.30) guarantees fulfilling the treatment requirements (3.5) on the target voxels (here that every target voxels receives a dose between p_{minT} and p_{maxT}). Otherwise, alternative statements can easily be formulated in order to enforce (3.6) and/or (3.7), while minimizing the energy sequencing cost. Furthermore, any optimal solution for program (3.23-3.30) guarantees the best possible treatment. For instance, one can set $U = 5$ and, therefore, determine whether a treatment that satisfies the $[p_{minT}, p_{maxT}]$ target dose prescription bounds actually exists when limiting to 5 EL switch-ups and what is its side effect cost in terms of OAR irradiation. Therefore, varying the EL bound U eventually provides a Pareto curve, describing the possible compromises between treatment quality and duration, for a particular use case. Finally, from a mathematical point of view, the problem statement as a linear model tells us that the energy sequencing can be represented by a convex function, not only for 0-1 switch up counts, but also for any possible cost functions to measure duration, as a simple pre-computed c_{mn} table.

Computational complexity

The inherent complexity of the ArcPT treatment problem, stated in (3.4)-(3.8), has not been addressed yet in the literature. Is the problem *NP-hard*? It could be that the problem is in fact polynomial, but remains intractable at this time, due to the typically huge size of the beamlet matrix $\mathcal{A}$. This question will likely remain open for now, although we propose a preliminary analysis, where we draw the first basic complexity boundaries by determining two variants of our problem: a special case that is polynomial and a generalization that we prove to be NP-complete.

Energy sequencing is hard

Unlike linear programming, which is known to be solvable in polynomial time, (e.g., using interior point methods [202]), integer and mixed-integer programming are naturally NP-hard in general. It is easy to see that program (3.23)-(3.30), without energy sequencing, that is, when removing constraints (3.25)-(3.28) and

(3.30), leads to a linear program:

$$\begin{aligned}
 & \min_{x_1, \dots, x_{|S|}} && \sum_{i \in R, j \in S} \mathcal{A}_{ij} x_j \\
 \text{s.t.: } & \forall i \in T && p_{\min T} \leq \sum_{j \in S} \mathcal{A}_{ij} x_j \leq p_{\max T} \\
 & \forall j \in S && 0 \leq x_j \leq I,
 \end{aligned}$$

which is solved in polynomial time. Therefore, if the overall treatment problem is NP-hard, the complicated part of the problem actually lies in the energy sequencing. By setting a value U that is large enough to never be reached by the summation (3.27), that is, *the problem becomes polynomial if one does not care about the beam delivery duration*.

NP-hardness: the case of the lower bounded energy sequencing variant

A direct generalization of the problem can be obtained by considering a lower bound L on the energy sequencing time, in addition to the upper bound U , hence replacing (3.27) with

$$L \leq \sum_{(m,n) \in E^{\text{edges}}} c_{mn} e_{mn} \leq U.$$

In order to prove the hardness of this generalization, we show how to achieve the polynomial reduction from the well-known NP-complete subset sum problem (SSP): given a multiset K and a target value Y , decide whether some subset of the items of K sums exactly to Y . Let us construct our EL directed graph (that of Fig. 3.7), such that there is one column per item $k \in K$, only one node in each column (representing the value of k), and finally c_{mk} equals the value of the target node k , whatever the source node m . Clearly, a path in the constructed directed graph represents a particular subset selection with its associated total value. Any decision algorithm for our generalized treatment problem then decides the SSP, when $L = Y = U$. Our generalized problem is then NP-complete, as any given solution can be verified in polynomial time (*i.e.*, the problem is in NP).

Large Neighborhood Search.

NP-hard or not, preliminary experiments showed that the ArcPT treatment optimization problem remains complicated to solve in practice. In fact, the disadvantage of problem statement (3.23)-(3.30), common to all complete optimization methods, is that it hardly scales up to realistically-sized problems. In this section, we propose a basic workaround, by exploiting a well-known technique: large neighborhood search (LNS).

LNS [201] relieves the optimization process by considering, iteratively, sub-parts of the entire optimization problem. It works as follows. Given a current (sub-optimal) solution x , select a subset $x_i, \dots, x_j$ of decision variables and fix their current values in x . This results in a sub-problem in which part of the variables are fixed, whereas the remaining variables could be optimized. By solving this sub-problem to optimality, one obtains a new current solution x' , which is necessarily as good as x . Starting from an initial solution x_{init} , this process can be repeated as long as needed.

LNS can be seen as a variant of our local search method defined in Algorithm 2, in which the only neighborhood operator $\mathcal{N}$ consists in this variable-selection and optimization process. Because $\mathcal{N}$ may not deteriorate the current solution x , $\mathcal{N}(x)$ is always accepted. LNS is therefore a “hill-climbing” heuristic method.

Variable selection

As the energy layers form a partition of the decision variables, it seems natural that the variable decision mechanism follows these energy layers. At each iteration of the LNS and given the current solution x , the optimization is then performed by restricting the treatment to use only a subset of the energy layers (EL). This subset is composed of all the EL that are currently activated in x , as well as a percentage (5% in our experiments) of randomly selected EL. Any EL that does not belong to the selection is necessarily not activated in current x , and therefore all the associated decision variables are set to zero in the MIP model for the upcoming optimization.

Initial solution and bi-objective optimization

Finding an initial feasible solution, which satisfies constraints (3.24-3.30), is already really challenging. Instead, it would be easier to start from an empty solution, thus no activated EL, and iteratively improve the solution by activating EL. Some iterations would then activate some ELs, whilst deactivating some of those EL that were previously activated (this is why currently activated EL are always part of the variable selection), and so on, hence continuously improving the current solution through the LNS iterations. The workaround we propose is then to relax constraints (3.24) so that an empty initial solution becomes feasible:

$$\forall i \in T : \sum_{j \in S} \mathcal{A}_{ij} x_j \leq p_{\max T}.$$

At each LNS iteration, the current MIP model (*i.e.*, having only subset of the decision variable available for optimization) is then optimized based on the following

lexicographical bi-objective function

$$\max_{x_1, \dots, x_{|S|}} \sum_{i \in T, j \in S} \mathcal{A}_{ij} x_j, \quad \min_{x_1, \dots, x_{|S|}} \sum_{i \in R, j \in S} \mathcal{A}_{ij} x_j,$$

which maximizes the total treatment on the target voxels T first and minimizes the irradiation on OAR voxels R second. In other words, a soft constraint substitutes for the hard constraint on the lower bound of the target voxels' prescribed dose.

LNS: pros and cons

The direct advantage of the LNS approach is that each iteration actually ends up with a solution. In fact, for large instances, the complete MIP approach may require hours (or even not terminate) before providing a feasible, even sub-optimal, solution. By sacrificing the optimality guarantee, *i.e.*, by renouncing the model completeness, the MIP model can still be exploited to construct and improve solutions, yet sub-optimal, scaling to larger problem instances. However, in addition to optimality, with the relaxation of lower bounding constraints in Eq. 3.24, part of the treatment guarantees are lost with the LNS approach.

3.7 Comparison strategy

Plan optimization methods

From the proposed benchmark (see Section 3.3), ArcPT plans have been generated and optimized, using the three algorithms proposed in Section 3.6. To serve as first reference point, IMPT plans (fixed-angle beams) were also produced. Another reference point was given by SPArc plans (described in Section 3.5.1), obtained with a Python implementation of the original SPArc algorithm in our in-house TPS. All these plan optimization methods were empirically compared to each others.

Key performance indicators (KPIs)

The KPIs used for this study are divided into four main categories:

1. The first KPIs in importance relate to the planned dose quality: the dose volume histograms (DVHs), re-normalized to D50 (dose $\leftarrow$ dose $\cdot p/D50$) for an easier comparison. Quantitative KPIs are also extracted from the DVH, such as the absolute difference between D95 and D5, written as $|\Delta(D95, D5)|$, characterizing the uniformity of dose distribution within the

target volume, as well as the conformity index (CI_{RTOG} [203]) that describes how tightly the prescription dose is conforming to the target.

2. The next KPIs relate to the time delivery efficiency of the plan, which depends on the sparsity of energy layers. Sparsity is quantified by computing the percentage of beam angles that ultimately have only one active energy layer remaining after the optimization process². The time delivery efficiency also depends on the number of energy switches upwards or downwards, further summarized in the ELST metric used as a surrogate of the delivery time and given in the tables.
3. The third KPIs are the final numbers of spots (nonzero spot weights) and layers (*activated* layers with non-null intensity), which are provided as proxies for the degrees of freedom available to the spot weight optimizer.
4. The Last KPIs are the optimization times, which were also reported in the comparison tables. Although all the algorithms ran on the same machine, execution times should be regarded carefully because each algorithm makes use of different programming languages. FISTArcPT is written in *Python3*, whereas the local search and MIP were implemented in *C++*. For the last two, the optimization time depends on the time limit set in the input parameters. Furthermore, whenever the MIP found the optimal solution, it would necessarily stop before reaching the time limit.

Benchmarks

For the patient data, we are using the benchmark defined in Section 3.3 with clinically acceptable dose resolutions of 2 mm and 1 mm. In addition, to assess the quality of the different plans, the algorithms were executed on two target sizes: 1.47 cm^3 (*small target*) and the same but dilated using 1 cm margin (*expanded target*). The initial degrees of freedom available to the spot weight optimizer are reported in Table 3.2 for both modalities, *i.e.*, IMPT and ArcPT (across all methods). ArcPT plans (FISTArcPT, LS, MIP) were initialized with a pre-defined arc set between 0° and 90° with a counter-clock rotation and a step angle set to 2° . The SPArc plans were initialized with three beams (77.0° , 45.0° and 13.0°) with a maximum number of splits set to 4 so that the same arc design would be obtained. The IMPT plans were designed by initializing 3 beams at equal angular distances, at 0° , 45° and 90° , respectively.

²Edited from the published version to provide clarity towards quantification of the sparsity metric.

		IMPT	ArcPT
Small target	#EL	39	361
	#spots	1805	6635
Expanded target	#EL	25	558
	#spots	400	29269

Table 3.2: Degrees of freedom available to optimizer

Experimental settings

The proposed methods come with many hyper-parameters. They were tuned in a preliminary study (Appendix A.1) and they remained fixed in the next experiments. Notice that a convergence report of each proposed algorithm is also provided in Appendix A.2, as well as a description of the algorithm behaviour, for various problem scales in Appendix A.3. The dose prescription to the target is set to 60 Gy and no objectives nor constraints are set on the OAR because they are not included in the analysis. The maximum number of iterations, set to 4000, was used as stopping criterion for all the FISTArcPT plans. IMPT plans were optimized by applying a SciPy implementation of L-BFGS [204] in OpenTPS to a beamlet decomposition of the dose. SPArc also relies on the L-BFGS algorithm for each step of optimization. The stopping criteria used for SciPy L-BFGS optimizations are the maximum number of iterations, set to 100, or whenever the convergence was acceptable with a relative error in function values not exceeding 0.001.

3.8 Results

The optimization methods presented in Sections 3.5-3.6 are intrinsically different but they all aim at designing ArcPT plans. This section demonstrates the proof of concept of each proposed method and also identifies the reasons for which one outperforms some others for specific key performance indicators.

Results on the small target

The different final dose distributions obtained with each algorithm are shown in Fig. 3.8 (2 mm resolution) and Fig. 3.9 (1 mm resolution), as well as the DVH comparison in Fig. 3.10. The KPIs can be compared in Table 3.3. In general, all the algorithms perform well, showing results that compare with the reference, that is, the IMPT treatment. FISTA struggles the hardest to reduce the number of switch-ups and to obtain a good layer sparsity with at most one energy layer active in each beam. For the lowest resolution instance (2mm), MIP performed the best

and was able to reduce the number of switch-ups to zero, eventually yielding the smallest ELST, while still covering the target homogeneously. For arc plans with 1 mm of resolution, the dose conformity to the target is poor, especially for those obtained with the LS method. Hot spots appear indeed in the dose distributions (Fig. 3.9), because no objective in the cost function prevented the optimizer from depositing dose in any locations outside the target. Anyway, SPArc seems to give the best performances at high resolution with a very good homogeneity of the dose in the target and an acceptable time for energy layer switching.

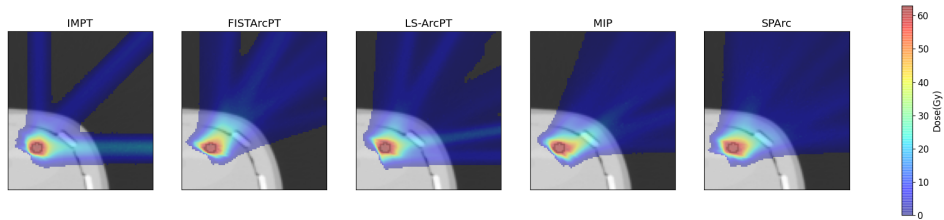


Figure 3.8: Optimized dose distributions (2mm resolution) obtained with the different modalities/algorithms - Small target

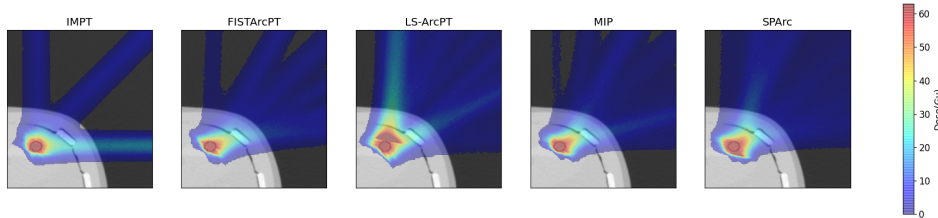


Figure 3.9: Optimized dose distributions (1mm resolution) obtained with the different modalities/algorithms - Small target

Results on the expanded target

Dilating the target by 1 cm happens to complicate the problem considerably and the algorithms faced more difficulties to solve it, as can be observed in Table 3.4 and in the DVH plotted in Fig. 3.13, especially when the CT resolution is thinner (1 mm). Despite a good ELST, the DVH obtained with the LS algorithm is off the 5% tolerance around the prescription with a large D5. Similarly, MIP can deal with the 2-mm resolution data set but failed to produce a decent solution for the 1 mm resolution with major deviations observed in its DVH. FISTA struggles again to reduce the ELST compared to the other optimizers. In terms of dose

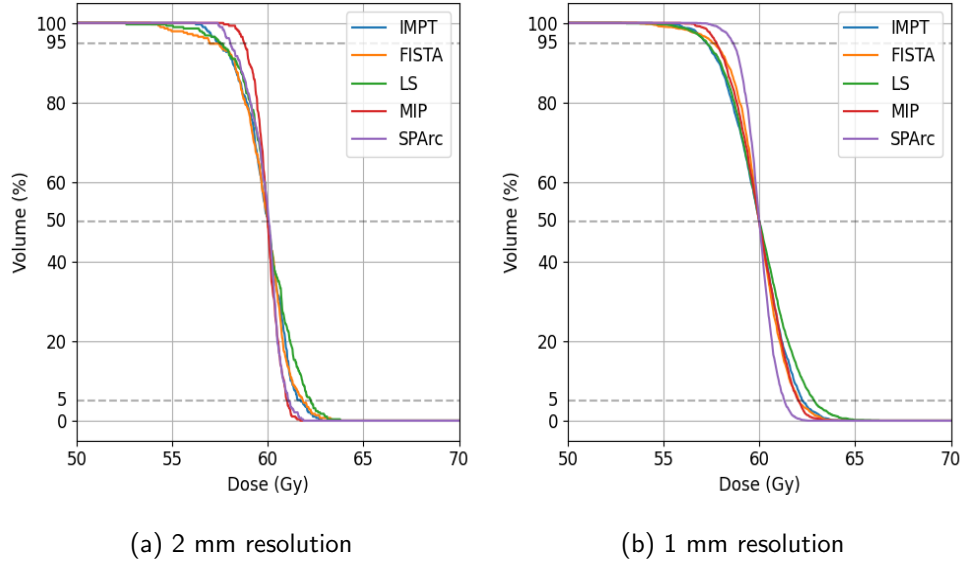


Figure 3.10: DVH comparison between the different algorithms used for the small target

maps (Fig. 3.11), the solution obtained with FISTA displays aggregated beams on the dose map despite the indicator function meant to address this kind of issue. Moreover, the energy layers selected by each algorithm are visible on the dose distribution plots. This could be explained by the fact that the layer spacing chosen is too large or there are not enough beams to cover the energy space smoothly. Surprisingly, IMPT outperforms arc plans. Possible explanations could be that the degrees of freedom (DOF) available in the arc modality are not exploited properly by the various optimizers. As MIP starts from a solution with null spot weights and solves the problem progressively by selecting a subset of the layers (LNS approach, see Section 3.6.3), it struggles to activate spots and layers resulting in a very small final number of activated spots and layers. The same observation holds for the LS method, even though it is caused by the neighborhoods operators that (de)activate layers and spots at random. Finally, FISTA and SPArc seem to use the available DOFs in the arc better than LS or MIP, although they cannot outperform the IMPT plans in terms of dose quality. A reason could be that the optimizer needs to run longer in order to deal with these many DOFs and obtain a better solution eventually.

Metric/Method	IMPT		FISTA		LS		MIP		SPArc	
Resolution (mm)	2	1	2	1	2	1	2	1	2	1
Planned dose quality										
$ \Delta(D95,D5) $ (Gy)	4.2	5.0	4.0	4.6	4.8	5.5	2.2	4.0	2.9	2.7
D98 (Gy)	56.8	56.7	55.0	56.3	56.5	56.5	58.5	57.3	57.7	58.0
CI_{RTOG}	1.4	1.1	1.2	1.4	1.3	3.7	1.1	1.2	1.3	1.9
Delivery efficiency										
#switch-ups	2	2	4	5	2	2	0	0	2	2
#switch-down	22	22	10	9	14	9	8	6	13	12
ELST (s)	24.2	24.2	28	32.9	19.4	16.4	4.8	3.6	18.8	18.2
Sparsity (%)	0	0	95.6	93.5	100	100	100	100	100	100
Degrees of freedom										
#EL activated	25	25	48	48	42	42	40	39	46	46
Non zero spots	400	400	1198	1214	131	228	171	217	1163	1225
Computational performance										
Optimization time (s)	2	12	2000	14000	300	3600	3600	3600	714	4423

Table 3.3: Metrics comparison between the different methods used for small target

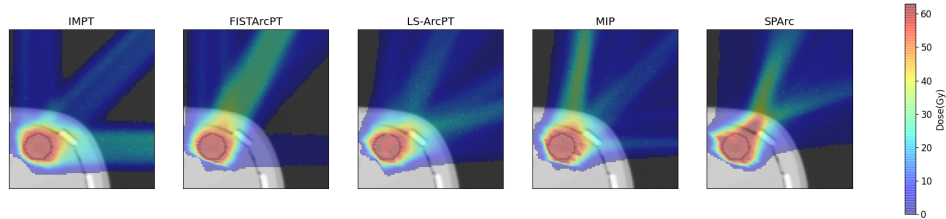


Figure 3.11: Optimized dose distributions (2mm resolution) obtained with the different modalities/algorithms - Expanded target

3.9 Discussion

We presented three new methods of optimizing treatment plans for arc proton therapy. Theoretical analysis and experimental evidences revealed the strengths and weaknesses of each method.

Firstly, FISTArcPT, which extends a classical gradient descent algorithm with regularization terms to control the number and the sequence of energy layers in each beam belonging to the arc, is the ideal method in the sense that it fully benefits from the arc geometry and its degrees of freedom and finds locally optimal solution with rapid convergence of objective function values. This method was able to produce good quality plans in terms of target homogeneity and conformity.

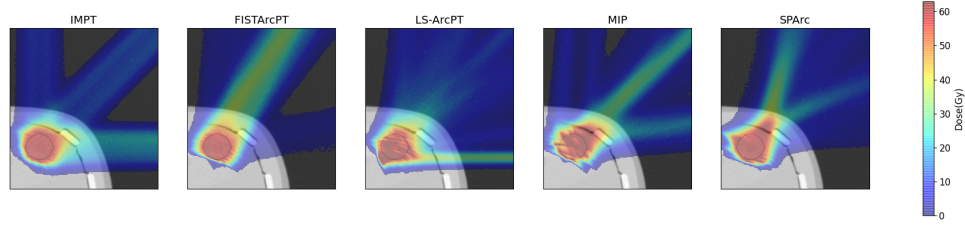


Figure 3.12: Optimized dose distributions (1mm resolution) obtained with the different modalities/algorithms - Expanded target

Metric/Method	IMPT		FISTA		LS		MIP		SPArc	
Resolution (mm)	2	1	2	1	2	1	2	1	2	1
Planned dose quality										
$ \Delta(D_{95}, D_5) $ (Gy)	3.3	3.7	6.0	5.8	8.4	18.4	6.4	23.5	6.5	7.5
D98 (Gy)	57.7	57.5	55.5	55.3	53.8	48.8	53.8	43.6	54.1	52.6
Cl_{RTOG}	1.3	1.4	1.1	1.1	1.2	1.5	1.3	2.3	1.7	1.7
Delivery efficiency										
#switch-ups	2	2	11	14	2	2	3	2	2	2
#switch-down	35	36	25	24	22	18	18	7	20	20
ELST (s)	32.0	32.6	75.5	91.4	24.2	21.8	27.3	15.2	23.0	23.0
Sparsity (%)	0	0	84.8	80.4	100	100	100	100	100	100
Degrees of freedom										
#EL activated	38	39	60	60	46	46	29	12	46	46
Non zero spots	1897	1805	2949	3139	457	511	433	272	2539	2474
Computational performance										
Optimization time (s)	19	181	20000	76000	3600	3600	12000	12000	2273	14423

Table 3.4: Metrics comparison between the different methods used for expanded target

However, the sensitivity to the hyper-parameters in the objective function and the need to tune them finely on case-by-case basis makes it inconvenient, especially considering the potential future use in clinic. Also, the inter-dependence between those parameters makes the algorithm biased and could lead to rapid local minima, without any means to get out of them, especially because of the definition of the shrinkage group parameter α_e , that depends on the intensity of each layer. In addition, because the objective function is not formulated to fulfill all the FISTA assumptions as defined by Beck *et al.* [106], this convex optimization iterative approach could underperform and also be trapped in local minima. Recall that we found, in Section 3.6, that the energy sequencing term of the objective function is actually not convex. Aside from those theoretical properties, the matrix-vector

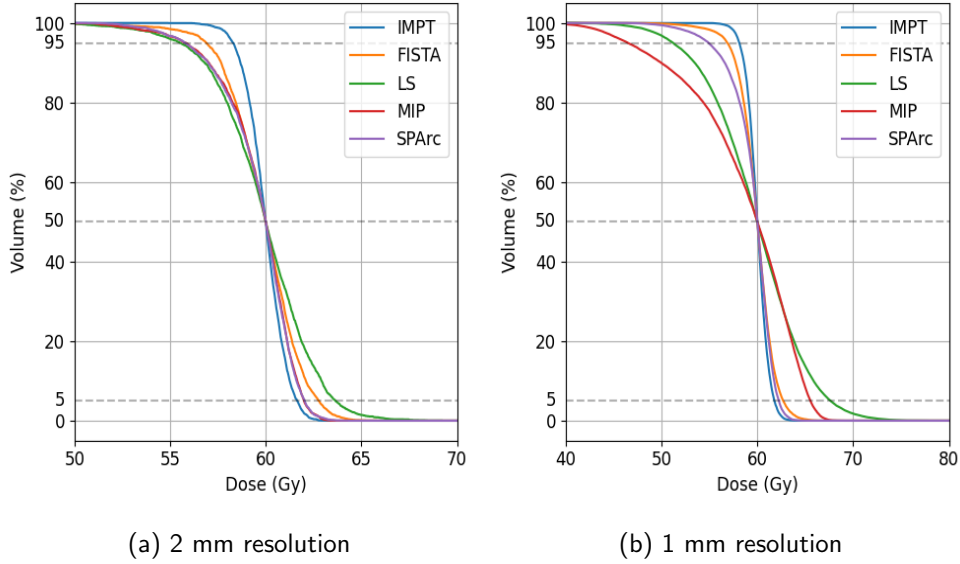


Figure 3.13: DVH comparison between the different algorithms used for the expanded target

product $\mathcal{A}x$ involved in the dose fidelity term is in practice the heaviest operation in the algorithm. It is clear through the examples that the larger the involved matrices are, with increases in resolution and/or target size, the longer the optimization time will be, and this linearly. Lastly, in addition to give an approximate solution (no treatment guarantees with soft objectives), FISTA is primarily an incomplete method, like all gradient-descent-like algorithms, because it does not consider the entirety of the solutions space and there is therefore no optimality guarantee (unless the objective function is convex, which is not our case).

Secondly, a local search approach with simulated annealing has been tested. In contrast to FISTArcPT, LS uses a randomized approach and will therefore probably never yield the same solution twice. Also, the simulated annealing meta-heuristics allows for escaping from local minima. LS was able to give solutions with high delivery efficiency (in terms of ELST) but is mostly discarded due to poor dose quality, especially for large-scale datasets. Regarding the optimization time, defined by the user time limit, it is less impacted by the size of the problem compared to FISTArcPT. In fact, in LS the computationally demanding matrix-vector product (in the gradient step of FISTArcPT) is carried out incrementally through the neighborhood operators. Besides, those neighbourhood operators remain very simple, and they could even be improved to get better results for larger-scale problems. Another point to consider is that any kind of objectives

and/or constraints, including more realistic ones, such as the BDT, can be easily implemented with such algorithms, without any concerns in terms of convexity or similar properties. However, it is noteworthy that the local search algorithm is a heuristic method with no optimality guarantee and, like FISTA, it yields an approximate solution with no actual guarantee of treatment feasibility or acceptability. In any case, if the dose quality does not improve in the future upgrades of the algorithm, it cannot be used for treatment planning. Nevertheless, LS could be used as a fast pre-optimization step to obtain an initial solution with an arc configuration, such as a specific number of upwards energy switching which would then ease the search for the main, subsequent algorithm.

Thirdly, Mixed Integer Programming is a method with good potential: given enough time, it ultimately finds the optimal solution to the problem at hand, thereby guaranteeing the best possible treatment. New objectives and constraints can be easily implemented with this problem statement as long as they remain linear. With MIP, the user can, for example, set the number of switch-ups allowed in the arc at the beginning of the optimization, as well as the range of dose to the target voxels, at the expense of losing degrees of freedom of the arc method, like SPArc. However, this type of algorithm, even though giving excellent solutions for small data sets, can rapidly struggle hard once the problem reaches a certain size, requiring hours before providing a feasible, sometimes sub-optimal solution if only it ever terminates. Several attempts have been made to address this problem scale issue. In particular, we have exploited a well-known technique: large neighborhood search (LNS) at the cost of losing part of the treatment guarantees. Another idea could be a multi-resolution approach, with a first optimization at a very low resolution to obtain a good result quickly and to input it as an initialization for a subsequent optimization at higher resolution and so on until the desired dose resolution is obtained.

Energy sequencing

The common point of all the three proposed problem statements and associated solvers (mathematically speaking, completely different from each other) is the difficulty to optimize the energy sequence as observed in the results. Stating the problem with continuous functions and using gradient-based methods such as FISTA [106], the convexity of the objective function is important to guarantee optimal performances of the solver. In the context of ArcPT, devising such functions is complex because of the discrete nature of arc therapy, leading to a delivery time estimator that is neither convex nor differentiable. Even though we simplified the real estimator by optimizing the energy sequence as a proxy of the BDT, we were not able to define a convex formulation that can be the Achilles heel of our

optimization. In contrast, MIP allowed us to formulate a convex estimator for the energy sequencing, at the expense, however, of introducing integrality constraints, which complicate the problem considerably. Regarding the LS algorithm, we assume that, if the energy sequencing was not in the objectives, the cost function would be convex and a solution would be more easily found and this, even without the simulated annealing meta-heuristic. The reason is that the energy sequencing term is the cause of the local minima in which we are trapped. Yet, when we will implement a real model of the delivery time, the local search algorithm combined with the simulated annealing could become the most suitable method of beam delivery time optimization.

SPArc

Compared to SPArc, FISTArcPT was able to give similar dose quality in all cases. Only MIP could outperform SPArc in the small target cases, which indicated its high potential for the ArcPT problem. SPArc provided a reproducible plan solution (though not necessarily optimal) in both small and expanded target with different resolutions. Compared to clinical IMPT, SPArc is an advanced IMPT optimization framework that implements a greedy, iterative re-sampling process. It is noteworthy, though, that the limitations of SPArc (or other similar methods) could only appear, from a theoretical standpoint, for more complex cases with OAR sparing for example where the heuristic greediness of SPArc and its restriction to a limited number of degrees of freedom could lead to under-performance compared to the proposed solution methods. It is therefore still difficult to demonstrate the superiority of the proposed methods with this simple test case. Yet, all the reported results should be analyzed cautiously since the benchmarks are not obtained from real patients.

Treatment uncertainties

No uncertainties have been taken into account neither in the optimization process nor in the treatment evaluation. In a clinical environment, the treatment plan has to be robust against various type of uncertainties before it can be delivered to the patient. They are usually divided into two main categories, namely, aiming errors, relatively to the tumor position, and range errors, concerned by the nature of the tissues the protons must traverse to reach a targeted depth. Integrating the robustness into each optimization schemes presented in this paper can be complex and will impact the quality of the results. The usual way to do it is to compute N scenarios that takes into account all types of errors that can occur before and during treatment, e.g., the patient set up errors are modeled as a shift of the beamlets, giving N sets of beamlets. The next step is to determine

the worst-case scenario and to minimize the corresponding objective function; repeating these two steps iteratively allows the performance to be lower-bounded. In our in-house TPS, *OpenTPS*, we have implemented this worst-case robust optimization. FISTArcPT could then be able to produce robust-optimized plans already. A robust version of our local search algorithm can be obtained by considering worst-case scenarios in a sampling-based approach. In fact, as long as the sample remains of reasonable size, our objective function can still be incrementally computed on each scenario independently. Finally, the method of mixed-integer programming will require more elaborated mathematical techniques in order to obtain robust treatment plans. Given a set of sampled scenarios, the L-shaped method [205] can be exploited to minimize expected penalties. Dealing with the inherent uncertainties of our treatment process is definitively a challenging issue, yet necessary. This is, however, out of our current scope, and left for further work.

3.10 Conclusions and future research directions

Arc proton therapy is an emerging cancer treatment modality, showing great potential and significant advantages over radiotherapy with x-rays and even over conventional proton therapy treatments. However, designing a valid treatment plan for this new modality defines a complex optimization problem, for which very few plan optimization methods exist in the particle therapy community. In practice, at the time of writing, there is still no proven or approved software solution available for the clinicians and further investigations is needed to develop the treatment planning system for ArcPT.

Contributions

This paper first reviews the literature about the existing approaches to ArcPT that raise interest in the proton therapy community, namely, SPArc to which we were able to compare our results and ELO-SPAT from which we based the formulation of one of our proposed method. Next, a first formal statement of the constrained optimization problem at hand is presented. Cues in this formalization guide the search for adequate optimization algorithms. Accordingly, we propose three different approaches to solving the problem, each with a specific problem restatement and an optimization algorithm. Our new methods are tested on different realistic benchmarks and compared to the existing state-of-the-art method SPArc as well as to a conventional IMPT method. After analysis of the influence of hyper-parameters in each method, our algorithms were able to produce acceptable ArcPT treatment plans. These plans achieve good dosimetric results and delivery efficiency in general. MIP has shown its superiority compared to the

other methods for small targets, as it can give guarantees on the constraints and objectives defined in the statement of a small-scale problem. However, for bigger targets, the IMPT plans and SPArc plans are still ahead which leaves some space to improve our methods depending on the direction we are taking. Though, each clinical case can be different and it is understandable that a method could be more appropriate than another for a specific case but improper for another case. Finally, we introduce the first reference benchmark to design ArcPT plans, obtained from realistic image data sets. This benchmark is made available to the research community.

Further work

The algorithms both optimize the clinical goals and minimize the delivery time in a single integrated process, thereby contrasting with SPArc. This followed path clearly suggests the use of a Multi-Criteria Optimization (MCO) that will be investigated into the future. Robustness is key in proton therapy due to its dosimetric sensitivity compared to the photon radiotherapy. Implementation of robust optimization in our ArcPT framework is therefore considered in the near future in order to make plans robust against uncertainties. Finally, this paper only considered an artificial image, *i.e.*, a patient phantom, as a proof of concept. We will apply our methods to real patients to further investigate the potential of ArcPT.

Chapter 4

Striking a balance

The content presented here comes from the published article: [Bi-criteria Pareto optimization to balance irradiation time and dosimetric objectives in proton arc therapy](#). Phys. Med. Biol. 67 245017, 2022.

This publication is the result of a collaboration with the Beaumont proton therapy center (MI,USA) in the context of 6-month secondment.

Authors: Sophie Wuyckens¹, Lewei Zhao², Michael Saint-Guillain³, Guillaume Janssens⁴, Edmond Sterpin^{1,5}, Kevin Souris^{1,4}, Xuanfeng Ding², John A. Lee¹

¹ UCLouvain, Institut de recherche expérimentale et clinique, Molecular Imaging and Radiation Oncology Laboratory, Brussels, Belgium. ² Department of Radiation Oncology, Beaumont Health, Royal Oak, MI, USA. ³ UCLouvain, ICTEAM, Louvain-La-Neuve, Belgium. ⁴ Ion Beam Applications SA, Louvain-La-Neuve, Belgium. ⁵ KULeuven, Department of Oncology, Leuven, Belgium.

Summary

Background and purpose: Proton arc therapy (PAT) is a new delivery technique that exploits the continuous rotation of the gantry to distribute the therapeutic dose over many angular windows instead of using a few static fields, as in conventional (intensity modulated) proton therapy. Although coming along with many potential clinical and dosimetric benefits, PAT has also raised a new optimization challenge. In addition to the dosimetric goals, the beam delivery time (BDT) needs to be considered in the objective function. Considering this bi-objective formulation, the task of finding a good compromise with appropriate weighting factors can turn out to be cumbersome.

Methodology: We have computed Pareto-optimal plans for three disease sites: a brain, a lung, and a liver, following a method of iteratively choosing weight vectors to approximate the Pareto front with few points. Mixed-integer programming (MIP) was selected to state the bi-criteria PAT problem and to find Pareto optimal points with a suited solver.

Results: The trade-offs between plan quality and beam irradiation time (*static* BDT) are investigated by inspecting three plans from the Pareto front. The latter are carefully picked to demonstrate significant differences in dose distribution and delivery time depending on their location on the frontier. The results were benchmarked against IMPT and SPArc plans showing the strength of degrees of freedom coming along with MIP optimization.

Conclusion: This paper presents for the first time the application of bi-criteria optimization to the PAT problem, which eventually permits the planners to select the best treatment strategy according to the patient conditions and clinical resources available.

4.1 Introduction

A common problem in radiotherapy planning is to find the best possible trade-offs between target coverage and normal tissue sparing. This is usually done by compounding a multi-objective function and manually tuning the weighting factors of each objective until the plan meets the clinical goals based on the planner's insights and experience [20, 24]. This iterative trial-and-error process may turn treatment planning into a long process and sometimes it can become tedious, especially for the patients with complicated geometry with conflicting clinical goals in terms of target coverage and OAR sparing. Another popular approach is to use multi-criteria optimization (MCO), which automatically generates a set of Pareto-optimal plans for different (well-chosen) patterns of objectives weights [206, 207, 21, 116].

Pareto optimality and front.

A solution (*i.e.* a treatment plan) is called *Pareto-optimal* if, by further improving one of the objectives, at least one of the other objectives gets worsened. In fact, there are as many Pareto-optimal solutions than there are possible such compromises between all the objectives at stake. By computing a set of such solutions, a so-called *Pareto front* (a.k.a. Pareto set or frontier) is obtained. In our case, the Pareto frontier represents the set of all Pareto-optimal treatment plans, over which the clinicians can efficiently navigate, and from which they can eventually pick the best trade-off depending on the clinical constraints at stake. This problem has already been extensively described in the literature for photon radiotherapy optimization such as VMAT and IMRT [25, 208].

Proton arc therapy.

In this article, we apply bi-criteria optimization to a recent innovation in proton therapy delivery, namely, proton arc therapy (PAT). PAT differs from conventional proton therapy in how the treatment is delivered to the patient, involving a sequence of many incidence angles along a rotating gantry arc, instead of using only a few fixed-angle beams. The price to pay for the larger number of irradiation angles and the increased flexibility that comes with them is that the beam delivery time (BDT) also increases, in spite of the continuous rotation of the gantry. The underlying challenge raised by this new modality is therefore to minimize the BDT concurrently with the traditional clinical or dosimetric goals. Because a true delivery time model will be certainly not convex, involving a *dynamic* delivery with mechanical parameters, we choose to rather optimize the *static* beam delivery time approximated by a simplified irradiation time, denoted $BDT^{\approx}$. We

are investigating a bi-objective formulation where we optimize simultaneously an estimation of the irradiation time, and the dose fidelity term grouping a fixed weighted combination of objectives to treat the tumor and to penalize dose in various healthy organs.

The PAT problem.

We refer to the problem of finding the best PAT treatment plan (*i.e.*, optimal dose distribution with minimal delivery time) as the *PAT problem*. A solution (*i.e.* a treatment plan) is composed of a sequence of irradiation angles. To each angle is associated a sequence of energy layers. Finally, each energy layer is composed of a sequence of spots, each spot is associated with an intensity used to modulate the dose delivered to the patient. In other words, a solution is a big sequence of spot intensities with their x-y coordinates. Informally (more formally defined in Sec. 4.2.1), our PAT problem consists in finding a combination of such intensity values, such that a set of dosimetric and logistic objectives are optimized.

Related work and contributions.

To the best of our knowledge, this is the first time bi-criteria optimization is applied to proton arc therapy in order to optimize the treatment delivery time and plan quality simultaneously. Among the few published PAT optimizers currently available, the oncology group from UCLA was the first one presenting a simultaneous approach for energy delivery sequence and plan quality optimization [16]. Based on this paper, Zhang et al. [155] have speed up the optimization and improved both delivery efficiency and plan quality using energy matrix regularization. The other previous published paper, however, do not directly optimize the energy sequence simultaneously with the spot intensities yet. For instance, RaySearch (RaySearch Laboratories AB, Stockholm, Sweden) has recently published a method of selecting energy layers and spots, as a proof of concept coined ELSA (Early Layer and Spot Assignment), prior to spot weight optimization [26]. Additionally, Ding et al. [17] developed the Spot-Scanning Proton Arc (SPArc) algorithm, an iterative greedy approach based on robust optimization. Bertolet & Carabe [161] have also developed the so-called proton monoenergetic arc therapy (PMAT) technique in which, as its name suggests, monoenergetic partial arcs are pre-selected and thus the energy sequencing optimization is a by-product of the application of the energy selection criteria. However, their research goal was different and the selection was carefully designed to enhance the dose-averaged linear energy transfer distribution within the target. As a consequence, we present the first bi-objective optimization study applied to proton arc therapy. The proposed framework is based on Pareto bi-criteria optimization (BCO) theory, using

an original algorithm for computing the Pareto frontier, inspired from the method introduced in Craft et al.'s [25].

Organization.

The rest of the paper is as follows. Section 4.2 provides a detailed description of the proton arc therapy bi-criteria problem and modeling as well as a sketch of the method applied to solve this problem incorporated in an innovative binary tree search technique. Section 4.3 demonstrates the method on three patients with different tumor sites: a brain, a lung, and a liver. Sections 4.4 and 4.5 discuss and conclude the paper. Appendices are also provided to define the mathematical model on which the PAT optimization is based (Appendix B.1) and to help the reader to understand the BCO Pareto approximation algorithm in more detail (Appendix B.2).

4.2 Materials and methods

This section defines the necessary mathematical notations, states the bi-criteria PAT problem within a Mixed-Integer Programming (MIP) framework. A binary tree search technique is proposed to construct an approximation of the Pareto front.

4.2.1 Problem statement

The PAT problem pools two different types of objectives. The first objective, namely, the dose fidelity term $f(\mathbf{d})$, reflects the standard clinical goals and it is formulated with the deviation between the delivered and prescribed doses of each voxel. The second objective measures the *static* beam delivery time $\text{BDT}(\mathbf{x})$, *i.e.* the irradiation time, that is minimized to obtain efficient PAT plans. This term depends on the plan delivery sequence which includes irradiation parameters such as spot scanning speed, energy layer switching time, and burst switching time and was modeled by Zhao et al. [150]. Given the set of spot S , we consider the following bi-objective optimization problem.

$$\min_{\mathbf{x}} \quad \{f(\mathbf{d}), \text{BDT}(\mathbf{x})\} \quad (4.1)$$

$$\text{s.t.} \quad \mathbf{d} = \mathcal{A}\mathbf{x} \quad (4.2)$$

$$\mathbf{x} \in \mathbb{R}_{\geq 0}^{|S|}, \quad (4.3)$$

where $\mathbf{d} \in \mathbb{R}_{\geq 0}^{|V|}$ is the vectorized dose on the $|V|$ voxels ($|\cdot|$ is the cardinality operator), $\mathcal{A}$ is the beamlet matrix, containing $|S| \times |V|$ elementary dose values,

i.e., the dose deposited by each spot to each voxel computed with an in-house Monte Carlo engine, MCsquare [72]. $\mathbf{x}$ is the column vector of spot weights, containing $|S|$ values (namely, our set of decision variables).

There is a conflicting interplay between these two main objectives. Reducing the BDT improves the patient's comfort, mitigates motion effects over time (and therefore treatment errors), and increases the number of patients who can be treated in a day. Yet, reducing too much the BDT can also significantly degrade the dose quality.

4.2.2 MIP model

The fidelity term can be modeled as a linear cost function (See equation B.8 in Appendix B.1) with weighted linear objective functions for penalizing over-dosing and under-dosing on target, OAR doses over the goal. This term alone leads to a linear programming (LP) problem that already demonstrated its ability for dose optimization in radiotherapy as early as in the 60's with Bahr et al. [209] and, more recently, with Bortfeld [210] or Romeijn et al. [211].

The real BDT objective, naturally *dynamic*, is not convex and this leads to difficulties for any optimizer. We choose to model and minimize the irradiation time ($\text{BDT}^\approx$) using the reasonable approximation that it takes 1 s to deliver one energy layer, 5.5 s to increase the energy, 0.6 s to decrease it and 0.2 s when same energy is kept to deliver the next layer [192]. In a previous work [212], we showed that this approximation of irradiation time can be modeled as a convex linear objective provided integrality constraints resulting in a mixed-integer programming (MIP) problem statement. The MIP statement has been extended to frame the bi-objective optimization problem 4.1, the specific mathematical expression of which is B.1-B.6 in Appendix B.1.

It is worth noticing that the integrality of the PAT problem is due to the $\text{BDT}^\approx$ (as well as the BDT in general). In fact, it mainly depends on whether each irradiation angle and energy layer is activated or not, introducing this temporal dimension turns our PAT problem into a discrete, integer problem. In a recent study [212], we proved in that the irradiation time only is responsible for the inherently complexity (i.e., the NP-hardness) of the PAT problem.

4.2.3 MIP solver

Optimization is carried out by Gurobi 9.5 [213], through its Python API, for mixed-integer programming using tuned parameters to speed up the algorithm. These tuned parameters include the specification of the algorithm used to solve the

continuous model, the presolving level and number of pass limits, etc. They were obtained using the Gurobi tuning tool that automates the search for parameter settings. In short, it runs the solver multiple times on the model, choosing different initial parameter settings for each call, in a search for settings that improve run time. The tuning tool parameter was run 24 hours on a model built from a phantom case with a small target. At the end, it provided a summary showing the number of parameter sets it tried and details on a few of the best parameter sets it found. On this basis, we use the parameter set with the shortest running time for the subsequent optimizations. Gurobi solves the MIP model using a linear-programming (LP) based branch-and-bound algorithm [108, 214, 215]. Starting from the original MIP, it removes all integrality restrictions and solves the so-called LP relaxation of the original problem through the barrier method [216] (tuned parameter). Next, it will start the branch-and-bound procedure and will improve the bounds (relaxation) and/or incumbent solution using heuristics. Additional techniques proper to Gurobi are included to limit the size of the branch-and-bound tree that must be explored. The termination criterion is a parameter set to the time spent solving the model in this study.

4.2.4 Pareto frontier approximation

In order to solve the bi-objective optimization problem (4.1), it is customary in the radiotherapy field to compute a set of trade-off solutions, which are referred to as Pareto optimal points and from which the planner can conveniently pick a suitable solution. In the current article, this is carried out by blending (or scalarizing) the multiple objectives into a single compound objective, formulated as a linear combination of the initial objectives: $w_D f(\mathbf{d}) + w_T \text{BDT}^\approx(\mathbf{x})$. However, the adjustment of the weighting factors, to be used as constant parameters by the optimizer, is not trivial if one wants to obtain a good approximation of the Pareto front efficiently. We decided to implement Craft et al.'s [25] algorithm initially applied to optimizing multiple-objective intensity modulated radiation therapy (IMRT) convex problems. In our case, the feasibility space is a polyhedron resulting from a combination of linear constraints (aside from the integer constraint). Note that the smallest polyhedron that contains all the feasible integer points is called the "convex hull" (which results from a different, optimal, set of linear constraints: the perfect problem formulation - in fact, knowing these constraints amounts to solving the problem itself). The objective function is also linear, and therefore convex. The convexity of both the feasibility space and the objective function allows us to reuse the method to approximate the Pareto front and to best reduce its front position uncertainty. Note that due to the discrete nature of our problem (and henceforth the MIP integrality constraints), the Pareto front is here a discrete set of points.

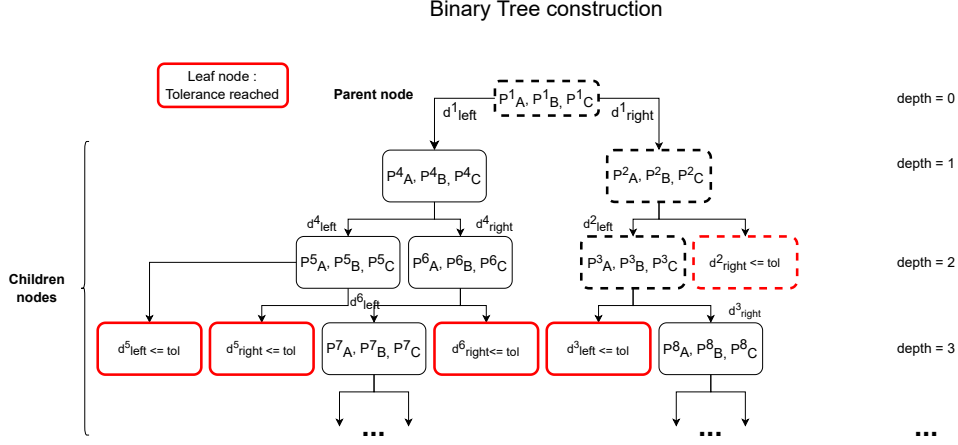


Figure 4.1: Pareto front construction in 2 dimensions as a binary search that produces a binary tree where each node corresponds to a point on the front. The figure contains technical details that are useful in the detailed explanation of the algorithm, in Appendix B.2.

For a bi-objective formulation, the BCO algorithm can be described as a search through a binary tree whose branches can be grown and explored in order to refine the Pareto front with new points wherever they are needed, such as illustrated in Figure 4.1. Our version of the BCO algorithm (Algorithm 3) differs from that of Craft et al.'s [25] by the use of a binary tree, hence specialized to bi-objective optimization.

In a nutshell, the construction of the binary tree goes as follows: we start from an initial node (the *parent* node) corresponding to three points P_A , P_B and P_C designated as the *anchor points*. P_A , and P_C relate to individual minima of each objective while P_B is an intermediate solution between P_A and P_C obtained with equal weights. Next, two new points (thus corresponding to Pareto optimal solutions) can be computed, respectively, between P_A and P_B and P_B and P_C , each corresponding to *left* and *right* branching in the binary tree depicted in Figure 4.1. In a 2D graph, maximal distances between segments lines $P_A - P_B$ and $P_B - P_C$ and the lower bounds of the Pareto front, *i.e.*, tangent lines through each vertex, are computed (d_{left} and d_{right} , respectively). Depending on the branching followed (the one with the minimal distance), points are renamed. For example, if one looks at left branching $P_A \rightarrow P_A$ but $P_B \rightarrow P_C$ and a new P_B is computed by solving the model using the weights average from P_A and P_C . In this way, P_B will always be the intermediate point between P_A and P_C that the optimizer needs to find. Then, new points can be subsequently and iteratively computed and

added to the Pareto frontier, each time between two existing points, and so on, until some tolerance threshold (or tree depth, or time limit) is reached. However, computing each new solution (*i.e.*, a point on the Pareto front) may potentially take a significant time. In order to cope with time limitation, one must obviously prioritize the order in which the new points are computed, or in other words the order in which the branches of the binary tree are constructed, so that we end up with a representative set of points, well-balanced, to approximate the Pareto front whenever the BCO algorithm is interrupted. The algorithm, and especially the prioritization method used, is explained in details in B.2, including the appropriate formulas and equations.

Algorithm 3 Binary search tree

```

while  $\mathcal{D}$  and  $i \leq I_{max}$  do
  input parent  $(\mathbf{P}_A^i, \mathbf{P}_C^i)$ 
  compute average weights  $(w_B^i)$  associated with parent  $(\mathbf{P}_A^i, \mathbf{P}_C^i)$ 
  optimize with weights  $w_B^i$  to find the single interior point  $\mathbf{P}_B^i$ 
  input the two presume descendants  $(\mathbf{P}_A^i, \mathbf{P}_B^i)$  and  $(\mathbf{P}_B^i, \mathbf{P}_C^i)$ 
  compute  $d_{left}^i$  and  $d_{right}^i$  and check tolerance of each
  if  $d_j^i < tol$  ( $j = \text{left, right}$ ) then
    remove  $d_j^i$  from  $\mathcal{D}$  (leaf node is reached)
  else
    update parent
  end if
  pick next node to run with largest  $d_j$  ( $j = \text{left, right}$ )
end while

```

4.2.5 Evaluation

As a proof of concept, we have tested the BCO algorithm on three clinical cases: a brain, a lung, and a liver. MCsquare [72] was selected as the dose calculation engine for the production of beamlets as well as the final dose computation. The sizes of each problem are reported in Table 4.1. The weighting factors in the dose fidelity objective were manually adjusted beforehand until the clinical goals were met in the situation where the weight of the dose fidelity term was much more important than the BDT weight. The aim was to get rid of additional dimensions in the problem and hence to keep a bi-objective formulation for the sake of simplicity.

For each case, we report a table with the metrics of interest, *i.e.*, the clinical goals, the irradiation time, the conformity index (CI_{RTOG} , [203]), and homogeneity index (HI, defined by the quotient of maximum and prescription dose to the

target volume) for three points lying on the computed approximation of the Pareto frontier. These points are carefully picked to demonstrate the trade-off machinery such as the impact of a large weight on the dose fidelity term (big D) combined with a small weight on the $BDT^{\approx}$ term (small t), denoted D_t and, conversely, d_T for low dose (small d) and high time weightings (big T). An intermediate point d_t is added to the comparison. It was chosen at a glance to show a more balanced solution although the plan quality might still be questionable in some cases. Other more specific points might offer a better compromise. Only a physician could make the final decision according to each patient's conditions. A series of plots are also provided and includes the computed Pareto fronts, the DVHs and dose distributions associated with the previously described three points, and finally the distributions of the absolute and relative weighting factors run by the BCO algorithm. In order to benchmark our results, we also evaluate an IMPT plan and a SPARC plan [17] for each disease site. They were produced in the same treatment planning system using the same solver (Gurobi) as the MIP plans, except that solely the dose fidelity term was minimized making it a linear program solved in polynomial time. Appendix B.3 provides additional details on the dose distribution and dose-volume histograms obtained with these modalities.

	Brain	Lung	Liver
CT voxel size (mm)	$1.17 \times 1.17 \times 1.00$	$1.23 \times 1.23 \times 3.00$	$1.17 \times 1.17 \times 2.00$
Beamlet voxel size (mm)	$2.5 \times 2.5 \times 2.5$	$2.5 \times 2.5 \times 2.5$	$3 \times 3 \times 3$
nSpots	5,816	10,220	17,866
nEnergyLayers	507	762	624
nBeams	37	48	41

Table 4.1: Size of each optimization problem

Partial arcs were defined for each clinical case to reduce the memory load and computation time. For the brain case, one partial arc was defined. With the couch angle set to 0° , it spans angles from 40° to 130° by steps of 2.5° , resulting in 37 beams. The IMPT plan is built with three beams whose gantry angles and couch angles are respectively 95° , 95° , 75° and 0° , 180° , 270° . The prescription to the tumor is set according to the IMPT clinical plan associated with the case, that is, 54 Gy to be delivered in 30 fractions where 95% of the tumor volume receives 100% of the dose. For this first case, beside the target coverage objectives, the sparing of the brain stem, the left optical nerve, and the optical chiasm were included in the fidelity objective given the close proximity of these OAR to the tumor.

The lung case was handled by creating two partial arcs with angles ranging from 2.5° to 197.5° and 322.5° to 357.5° , respectively, and spaced by 5° . Regarding

the IMPT plan, one posterior-anterior and one lateral beams were used. The patient was prescribed a hypo-fractionnated dose of 48 Gy in 12 Gy per fraction. For this patient, optimization objectives were only set on the target coverage and irradiation time given that the dose metrics for the surrounding OAR were particularly low and uncertainties were not considered.

The third case makes use of one partial arc covering angles from 170° to 250° with a control angular step of 2° . The IMPT plan for the liver case makes use of two oblique posterior-anterior beams (couch angle = 180°) at 155° and 186° . We followed the original prescription of 67.5 Gy to be delivered in 15 fractions in total. For this patient, we have included soft objectives to control the average dose received by the liver (target volume is subtracted) and the spinal canal maximum dose level.

All the SPARc plans were initialized and splitted so that the resulting arc would reach the same sampling angle frequency and arc span (start-stop) as the MIP plans.

4.3 Results

We apply the proposed BCO algorithm to three clinical examples as stated in 4.2.5. Their results are demonstrated case by case as the following:

4.3.1 Brain

From Table 4.2 and the DVHs in Figure 4.2b, one can already study the balance between the objectives. The three points included in the table and shown in Figure 4.2a, are Pareto-optimal plans. For example, the extreme anchor point D_{t} yields a plan that has the best target coverage ($D5\text{-}D98 = 3.45$ Gy) among all plans but also the worst irradiation time. If the user wants to improve the irradiation time, it translates automatically to a degradation of the dose quality (d_{t} solution: $D5\text{-}D98 = 4.39$ Gy) and vice-versa. Although the organs at risk are always better spared with higher $BDT^{\approx}$ weighting, it is only due to the fact that less beams and layers are activated in this case. For the d_{T} point, we did not pick the other extreme point (*i.e.* P_2 anchor point) because the dose distribution obtained has absolutely no clinical relevance due to too poor plan quality objective weighting versus the BDT one. The dose distributions (Figure 4.3) associated with each point are drastically different and emphasize the need to understand the compromises. Figure 4.2c shows the distribution of absolute weighting factors in decision space that were run during the BCO iterative algorithm. A color code is used to understand to which part of the Pareto front (Figure 4.2a) corresponds.

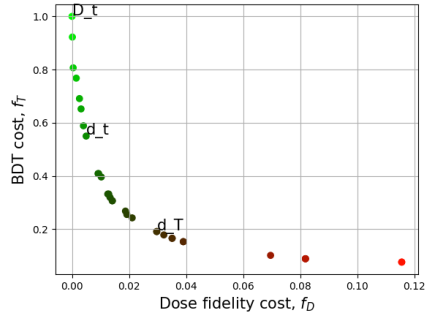
Their distribution appears semi-continuous in the sense that the fidelity objective minimization is continuous while the $BDT^{\approx}$ is discrete due to energy sequencing. Thus, this discrete behaviour impacts directly the aspect of the Pareto frontier, with few points at low dose weighting (high fidelity cost). When looking at the objective cost versus the relative weights this time in Figure 4.2d, the direct point-to-point correspondence is observed with almost a symmetrical behaviour supporting the hypothesis of conflicting objectives. Regarding the benchmark (last two columns of Table 4.2), the MIP demonstrates its superiority compared to both IMPT and SPARc plans in terms of OAR sparing thanks to the efficient use of the additional degrees of freedom available in the optimization. The SPARc plan is still faster but fails at satisfying the dose constraints.

Modality				MIP - PAT			IMPT	SPArc
				D.t	d.t	d.T		
	ROI	Metric	Goal (Gy)	Result (Gy)				
Fidelity	Target	D98	51.30	53.28	53.06	50.28	53.59	52.72
		D5	56.70	56.73	57.45	61.58	57.03	59.07
	Optical chiasm	D5	54.25	54.14	54.87	58.34	55.85	57.53
	Brain stem	Dmean	21.50	4.34	4.00	3.45	4.90	5.82
		D5	54.00	17.37	16.54	14.47	19.10	20.98
	LON	D5	54.50	54.04	55.53	54.87	54.48	56.21
	External	D5	56.65	2.31	2.12	1.10	4.18	3.27
CI				2.28	2.08	2.26	2.59	3.03
HI				1.06	1.07	1.19	1.06	1.12
Irradiation time (s)				126.2	69.4	24.1	72.2	51.40
# active spots				288	248	183	315	276
# active layers				37	26	13	40	27
# active beams				37	26	13	3	27
Optimization time (s)				18000	18000	18000	16	2155

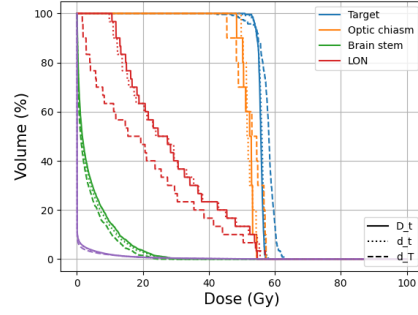
Table 4.2: Summary of optimization results for the brain case for three points standing on Pareto front illustrated on Figure 4.2a. Bold numbers correspond to best result for that metric.

4.3.2 Lung

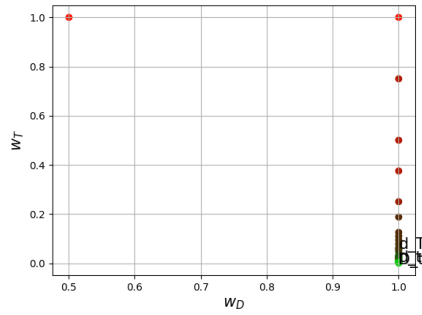
From the DVHs (Figure 4.4b) and metrics related (Table 4.3), sensitive trade-offs can be observed with the target losing rapidly its dose coverage if the time weighting gets too strong. Figure 4.5 illustrates the addition of beam angles to secure the dose fidelity objective and target conformity index. Compared to the brain case, this dataset is twice as large, making it harder for the MIP to find an optimal solution. This behaviour is reflected in Figure 4.4a where some points



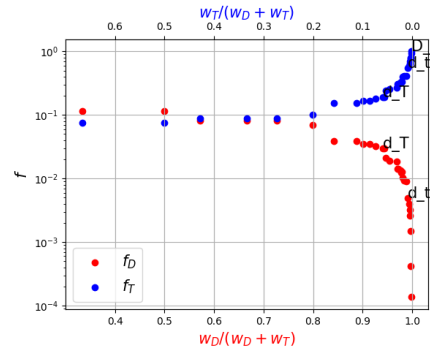
(a) Pareto front



(b) DVHs associated with 3 points on Pareto front



(c) Absolute weighting factors distribution



(d) Relative weighting factors distribution

Figure 4.2: BCO results obtained for the brain case. D_t is a solution where dose fidelity objective has large importance versus $BDT \approx$ objective while d_T just does the opposite and d_t is a more balanced combination of both objectives

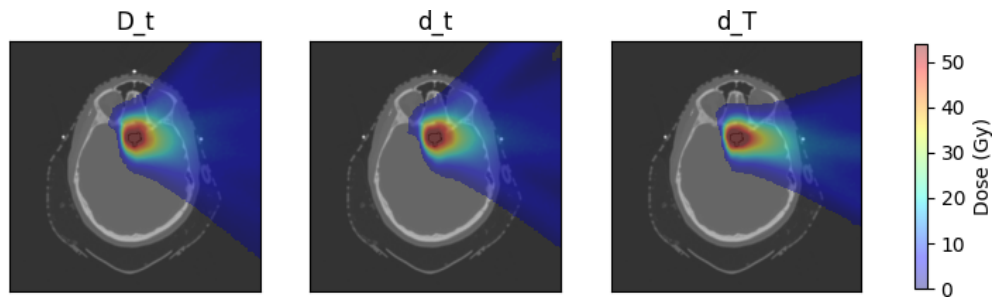


Figure 4.3: Brain dose distributions for three points standing on Pareto front

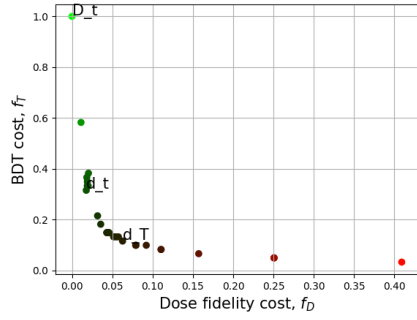
stack on levels, indicating the MIP was not able to push enough the optimality of the solution found in the allocated time. The IMPT and SPArc plans present significant improvement in the target coverage compared to the MIP plans. The underlying cause is that the plan parameters (spot spacing, layer spacing, and target margin) differed from those used in the MIP plans. For the lung case, MIP plan parameters create a spot placement with insufficient coverage of the target, which the SPArc method was unable to compensate for. Hence, we increased the target margin by 1 mm, which led to better dosimetric results for SPArc. An interesting consequence is that the intermediate MIP plan (d_t) is much faster than the SPArc plan to be delivered.

Modality				MIP-PAT			IMPT	SPArc
				D_t	d_t	d_T		
	ROI	Metric	Goal (Gy)	Result (Gy)				
Fidelity	Target	D98	45.60	47.13	46.79	44.74	46.89	47.25
		D5	50.40	52.21	54.89	61.78	52.04	51.89
	Heart	D5	15.00	0.62	2.15	0.03	0.00	0.70
	Lungs-GTV	Dmean	1.00	0.68	0.68	0.83	1.07	0.99
	Chest wall	D5	30.00	2.29	2.35	1.57	1.84	3.65
	External	D5	62.00	0.06	0.05	0.03	0.02	0.06
CI				1.29	1.31	1.60	1.82	1.74
HI				1.10	1.17	1.36	1.13	1.12
Irradiation time (s)				96.6	30.5	11.2	49.7	73.80
# active spots				199	141	114	186	195
# active layers				43	17	8	29	41
# active beams				43	17	8	2	41
Optimization time (s)				18000	18000	18000	45	950

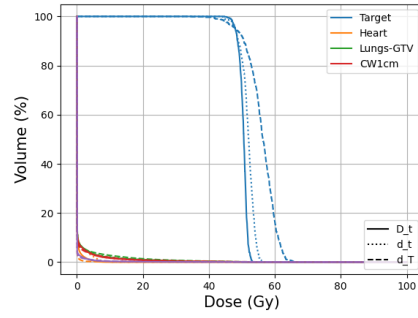
Table 4.3: Summary of optimization results for the lung case for three points lying on Pareto front illustrated in Figure 4.4a

4.3.3 Liver

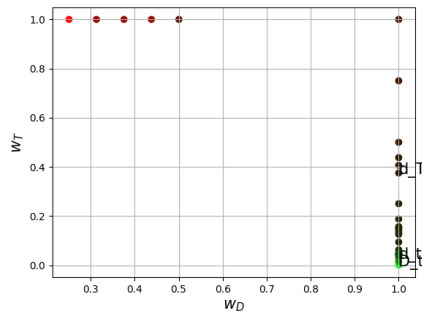
The extreme anchor point d_T point obviously leads to very poor dose quality (Figure 4.6b) and should not be considered at all in a real clinical study. Table 4.4 shows an interesting compromise where we would reduce by more than half the irradiation time and lose 2.8 Gy in the target coverage (D5-D98) when comparing points D_t and d_t. Considering the largest beamlet size, *i.e.*, lower dose resolution, the MIP optimizer was this time able to compute a very smooth approximation of the Pareto front (Figure 4.6a) within the time limit. Smoothness is referring in this case to the shape of the hypothetical curve connecting the solution points on the Pareto front. Figure 4.6d now shows clearly the inter-dependence of both objectives with a symmetric correspondence point by point. Finally, looking at



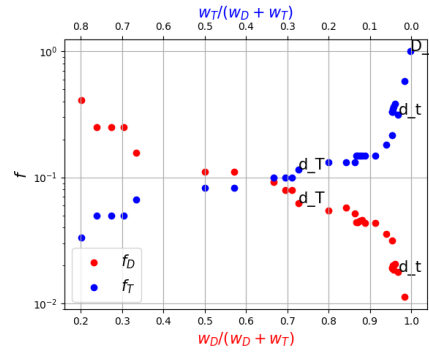
(a) Pareto front



(b) DVHs associated with 3 points on Pareto front



(c) Absolute weighting factors distribution



(d) Relative weighting factors distribution

Figure 4.4: BCO results obtained for the lung case. D_t is a solution where dose fidelity objective has large importance versus $BDT \approx$ objective while d_T just does the opposite and d_t is a more balanced combination of both objectives.

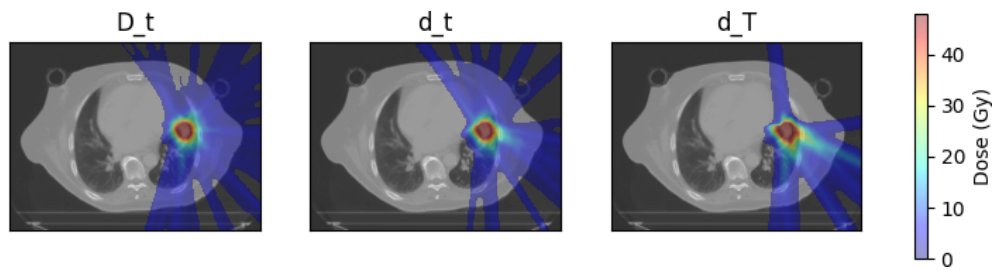


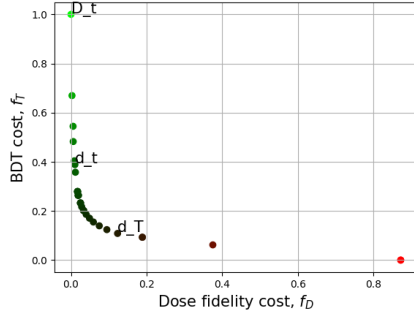
Figure 4.5: Lung dose distributions for three points lying on Pareto front.

the benchmark results, IMPT and SPArc plans show superior target coverage

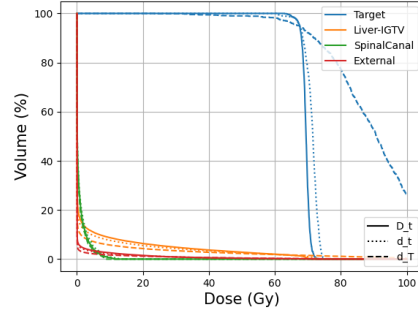
compared to the intermediate point d_T picked from the Pareto front. However, this can be explained by the fact that plan parameters had also to be tuned to obtain a sharp target DVH. The spot grid used by the MIP was too small and too sparse this time indeed. The layer spacing and spot spacing had to be both reduced, whereas the target margin had to be increased, in order to bring sufficient target coverage. B.3 discusses a few additional results to support this claim.

Modality				MIP - PAT			IMPT	SPArc
				D_t	d_t	d_T		
	ROI	Metric	Goal (Gy)	Result (Gy)				
Fidelity	Target	D98	64.13	66.37	66.27	60.12	66.20	66.17
		D5	70.88	71.13	73.82	99.96	72.01	71.88
	Liver-IGTV	Dmean	4.10	3.54	3.09	2.14	3.88	4.45
	Spinal Canal	D5	8.00	4.53	4.77	3.16	7.9	6.45
	External	D5	71.25	1.25	0.72	0.10	1.10	2.66
CI				1.14	1.00	0.57	1.52	1.59
HI				1.04	1.05	1.25	1.08	1.09
Irradiation time (s)				103.20	40.10	11.20	54.50	65.80
# active spots				640	557	354	680	714
# active layers				41	23	8	32	36
# active beams				41	23	8	2	36
Optimization time (s)				18000	18000	18000	40	1200

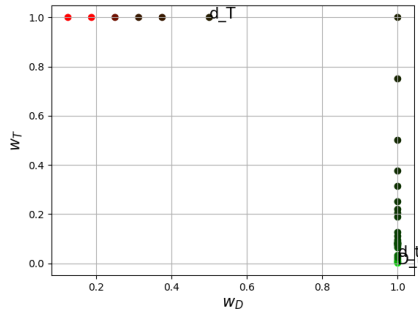
Table 4.4: Summary of optimization results for the liver case for three points standing on Pareto front illustrated on Figure 4.6a



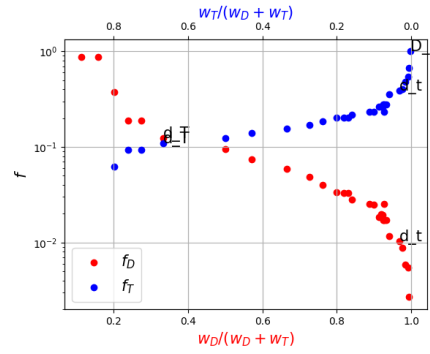
(a) Pareto front



(b) DVHs associated with 3 points on Pareto front



(c) Absolute weighting factors distribution



(d) Relative weighting factors distribution

Figure 4.6: BCO results obtained for the liver case. D_t is a solution where dose fidelity objective has nearly exclusive priority versus $BDT \approx$ objective while d_T just does the opposite and d_t is a more balanced combination of both objectives.

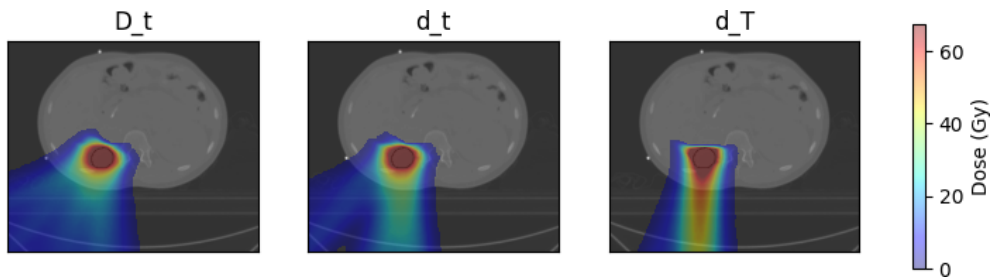


Figure 4.7: Liver dose distributions for three points lying on Pareto front.

4.4 Discussion

We have presented the first application of bi-criteria optimization to the PAT treatment planning problem, which requires innovative trade-offs between clinical and delivery time goals.

Mixed-Integer programming.

Finding an optimal point on Pareto front is converted into a traditional single objective optimization problem. We choose mixed-integer linear programming as most suited formalism to tackle it, and a well known optimizer to solve it. According to our previous study [212], the MIP problem statement is a proper formulation, allowing to model our full objective function and constraints in a convex form, which is a commonly admitted prerequisite to apply the BCO algorithm. Given enough computation time, solutions with optimality proofs can be obtained by using an appropriate solver (such as Gurobi) on our MIP model. Unfortunately, our problem is proven to be NP-hard [212], which means that such proofs, or even the corresponding solutions, are impossible to compute in reasonable time, for realistically sized problem instances. This is the case when the dimension of the problem increases, typically for higher dose resolution or larger tumor. This has been observed in some of the Pareto plots, where several points found by the optimizer were slightly “off the curve”. This indicates a clear limitation of our current optimizer to give an efficient approximation of the Pareto frontier. However, looking at the results, we can say that our optimizer is close to finding and proving the optimal solutions considering the smooth shape of the Pareto fronts with a slight experimental noise that also supports the claim of a convex approximation. Over 30 optimizations for each case, about 10 are useful to represent the Pareto front. Some runs were indeed not necessary because their final solution was already found in a previous run or eventually ended being not Pareto optimal. We suspect this behavior is induced by the discrete nature of the PAT problem. Because we are using a MIP model, the number of points in the Pareto front should be finite. The Pareto front is therefore discrete and disconnected, as observed in the results. A benefit of our MIP formulation is its ability to interchange soft objectives with hard constraints, although this has not been exploited in the current article. As shown in Wuyckens et al. [212] the ability to turn medical constraints, from soft to hard, may be beneficial in order to compute solutions with stronger treatment guarantees in some cases, depending on the complexity of the problem and clinical considerations¹. Finally, the benchmark produced for this study including a conventional IMPT plan and a SPARC plan

¹Slightly adapted from the published version.

allows to make several statements. Given that for each case, the plan parameters had to be adjusted to bring sufficient target coverage for both IMPT and SPArc plans, it only reflects the power of degrees of freedom that is efficiently used in the MIP plans. It also demonstrates that the energy sequence optimization matters and especially for more complex cases such as small brain tumors for which, in our case study, SPArc (through its greedy heuristic) was unable to satisfy the dose constraints even with tuned plan parameters.

Anchor points choice.

The weights of the anchor points we selected to start the BCO algorithm might not be very reasonable considering a large part of the Pareto optimal plans are too extreme for a physicist in real clinical practice. Indeed, many points lead to poor results in terms of dose distribution and would not be regarded at all. A less exhaustive and more economical exploration technique should therefore be devised to avoid some unnecessary optimization runs. For example, fixing hard constraints (made possible by MIP) to obtain a minimal acceptable target coverage in addition to the soft objectives might have given more clinically relevant results.

True delivery time.

In this work, we only consider an approximation of the *static* beam delivery time. However, the *true* delivery time is in reality *dynamic* and includes mechanical parameters on top of the irradiation parameters such as maximum gantry velocity, acceleration, and deceleration speeds. The modeling of such function and the search for a suitable optimizer are very complex and require further research.

Time considerations.

Solving each scalarized problem for a mixed-integer programming problem is very costly. For this study, each optimization has been run for 5 hours. Considering we have made 30 iterations to approximate a single Pareto front for each test case, it implies that 6 days in total were needed (and this, for small target volumes considered), which is certainly not realistically feasible for a clinical implementation at the present time. Nevertheless, this huge computation time is not exclusively attributable to the BCO layer. Underneath, PAT optimization is still recent and much progress is needed to improve the optimization efficiency. In our case, further effort could be dedicated to the MIP problem statement itself, as well as to the tuning of the optimizer parameters.

N dimension.

We have considered a bi-objective formulation although in reality the fidelity objective is itself a multi-objective problem. The planners state a new objective for each target/organ clinical goal specified in the prescription. A weighting factor is therefore attached to each term and requires a fine tuning to fulfill the combination of objectives. In this work, we indicated that these factors were adjusted beforehand through an iterative manual loop. However, one could ask about the implications for subsequent bi-objective optimization. The reason behind this choice is the following: the dose quality is the most important objective in any case, as we cannot deliver a plan that does not respect the dose limits. If we find a good set of weighting factors for the target and OAR that work to obtain a good plan quality, we assume the same set will also work for the subsequent bi-objective optimization including the BDT objective given that we only seek to minimize BDT as long as the dose quality remains sufficient. We understand that the last statement is weak and might deteriorate our algorithm efficiency to solve the PAT problem since degrees of freedom are cut out when adjusting the weights beforehand or that a set of weights might be more convenient than another for a specific BDT weight but this extends outside the scope of this work. The next logical step is therefore to use multi-criteria optimization. One could use the convex hulling method presented by Craft et al [25] to take all the dimensions of the problem into account. For example, in the liver case, the MCO would show the trade-offs between average liver dose, high dose volumes for the other OAR, or conformality. Advanced techniques for generating Pareto fronts of MIP-specific problems should also be considered though today this is still a very active field of research. Recently, Burachik et al. [217] presented a review of such algorithms where they ultimately proposed a new four-objective algorithm in a challenging mixed-integer programming problem.

Uncertainties.

In this study, uncertainties have necessarily been neglected in the treatment planning optimization and, consequently, no robustness evaluation was performed. However, robustness planning and evaluation are of utmost importance for proton therapy due to dose distributions sensitivity to the uncertainties of the position of the Bragg peaks [218, 219]. The integration of robustness in proton arc therapy treatment optimization is extremely difficult and also computationally too intensive at this time, when using optimizers with no restrictive pre-selection of the energy layers, *i.e.*, using the full available set of degrees of freedom. Actually, none of the works published in the literature on proton arc optimization with no pre-selection has presented results optimized with robust settings. In our case,

mixed-integer programming will require elaborated mathematical techniques in order to obtain robust treatment plans. Given a set of sampled outcomes, the L-shaped method [205] could be exploited to minimize expected deviations from the initial treatment. Monte Carlo sampling, such as the sample average approximation method [220], should also be considered, although the complexity of our problem will severely limit the size of the sample required to produce estimates. Few methods exist in order to reduce the size of a sample, while preserving its representativeness [221]. One could also look at the progress made with the well-known beam angle optimization (BAO) problem [67] and translate the methods to energy layer optimization (ELO) problem we have in proton arc therapy. Moreover, there is still limited data for robustness optimization or evaluation for PAT, and therefore no indication on what a good robustness evaluation strategy would possibly be used for this new modality. For this purpose, the difference of robustness between PAT and IMPT should be investigated and, based on that, tools to evaluate the robustness of proton arc plans can be developed in a clinically relevant way. Note that there is still no consensus for the robustness evaluation of a conventional IMPT plan itself, though the topic is widely recognized by the community [81].

MCO utility.

If planners made use of an efficient MCO algorithm, it would eventually speed up planning time by a great amount and ease the decision-making process to select the plan, which represents the most desirable compromise between target coverage, organ at risk sparing, and irradiation time, where the last is critical to make efficient use of the PAT modality. While interesting algorithmic challenges remain open for PAT treatment plan optimization (reliable approximation of beam delivery time, MIP modeling) as well as for Pareto front generation in multi-objective mixed-integer programming problems, we believe PAT potential could ultimately be unlocked when combined with MCO.

4.5 Conclusion

PAT is a new modality that delivers the treatment over multiple proton beam directions following an arc. The PAT optimization problem imposes minimizing the beam delivery time in addition to the traditional clinical and dosimetric goals. Trade-offs between these two objectives are therefore indispensable in order to deliver dose and time efficient PAT plans. In this work, we have demonstrated the feasibility of applying a bi-criteria algorithm to the PAT problem statement, which could further help the treatment planners understand the different options that

are actually available. The next steps as future work would be to further reduce the optimization time, e.g., by using an iterative grid size approach, implementing heuristics to shrink the searchable solution space, etc. Moreover, scaling to N dimension, with more than 2 concurrent objectives, and hence solving a clinically relevant MCO is another important milestone on the future roadmap of PAT.

Chapter 5

Streamlining the process

The content is based on a manuscript in preparation with the title *“Proton arc plan optimization based on organ at risk sparing and delivery time driven energy layer pre-selection”*.

Authors: Sophie Wuyckens¹, Guillaume Janssens^{2,3}, Macarena Chocan Vera¹, Dario Di Perri⁴, Edmond Sterpin^{1,5,6}, Kevin Souris^{1,3}, and John A. Lee¹

¹ UCLouvain, Institut de recherche expérimentale et clinique, Molecular Imaging and Radiation Oncology Laboratory, Brussels, Belgium. ² UCLouvain, Institute of Information and Communication Technologies, Louvain-La-Neuve, Belgium. ³ Ion Beam Applications SA, Louvain-La-Neuve, Belgium ⁴ Cliniques Universitaires Saint Luc, Brussels, Belgium. ⁵ KULeuven, Department of Oncology, Laboratory of experimental radiotherapy, Leuven, Belgium. ⁶ Particle Therapy Interuniversity Center Leuven - PARTICLE, Leuven, Belgium

Summary

Introduction: As proton arc therapy (PAT) approaches clinical implementation, optimizing treatment plans for this innovative delivery modality remains challenging, especially in addressing arc beam delivery time. Existing algorithms for minimizing delivery time are either optimal but computationally demanding or fast but at the expense of sacrificing degrees of freedom. In this study, we introduce a flexible method for pre-selecting energy layers (EL) in PAT treatment planning before the actual robust spot weight optimization.

Approach: Our EL pre-selection method employs metaheuristics to minimize a bi-objective function, considering an approximation of dynamic delivery time and tumor geometrical coverage penalized as a function of selected organs-at-risk (OAR) crossing. We evaluate the method using a young adult craniopharyngioma case, providing a comprehensive dosimetric analysis benchmarked against intensity modulated proton therapy (IMPT) and dynamic proton arc plans generated with early energy layer selection and spot assignment (ELSA) in RayStation.

Results: The algorithm efficiently generates Pareto-optimal EL pre-selections in approximately 5 minutes. Subsequent PAT treatment plans derived from these selections and optimized within the TPS, demonstrate high-quality target coverage, achieving a high conformity index, and effective organ-at-risk sparing. These plans substantially reduce the number of required energy layers to meet clinical goals while minimizing delivery time, resulting in a 40% reduction compared to the best ELSA arc plan. Additionally, the EL pre-selection improves hippocampal sparing by avoiding critical directions.

Significance: The proposed algorithm offers speed and efficiency, producing high-quality PAT plans by intelligently selecting energy layers prior to robust spot weight optimization. It efficiently reduces delivery time while maintaining good target coverage and sparing OAR.

5.1 Introduction

Proton arc therapy (PAT) is an advanced modality of radiation therapy that dynamically delivers proton beams during gantry rotation. Compared to conventional techniques, such as intensity modulated proton therapy (IMPT), multiple studies have shown that PAT may offer superior dose distribution [175], increase treatment robustness [176, 179], enhance radiobiological effects [161, 160], and simplify the treatment workflow [18]. However, the treatment planning optimization challenge associated with this novel modality demands sophisticated algorithms and innovative computational approaches to fully exploit its potential benefits while keeping realistic treatment times.

In addition to the usual dose-based objectives, the PAT optimization problem introduces a new aspect, which is the minimization of the delivery time. The arc delivery time is highly dependent on the number of irradiation directions and the number and sequence of energy layers (EL). On the other hand, the PAT optimization space contains a much larger number of potential beam angles and EL than conventional IMPT. Consequently, large efforts have been made to minimize delivery time by reducing the number of EL and finding an energy sequence such that energy switch-up (SU) occurrences are minimized, as these transitions are the most time-consuming.

There are currently two major approaches to solving the PAT problem. First, one can explicitly include both dose and delivery time surrogate in the optimization function. For instance, Gu et al. published a complex objective function, composed of several terms, each addressing a specific goal, such as EL sparsity and sequencing, during spot weight optimization [16]. Mathematically speaking, this integrated framework is the most rigorous, and the solution found is likely the most optimal one, utilizing the full set of degrees of freedom (DOF). Our group and Zhang et al. have also explored such methods [155, 212, 156]. However, these methods share common drawbacks, including the need for new algorithms to handle non-differentiable objectives, important tuning of objective weights, and heavy computational power.

Secondly, there are heuristic-based methods that decouple the problem by pre- or iteratively- selecting EL and then performing a regular spot weight optimization with the selected set of layers. Although these techniques eliminate part of the DOF, they yield encouraging results and do so usually in a highly efficient computing time, making them more suitable for clinical integration. For instance, Sanchez-Parcerisa et al. [153] proposed several approaches for selecting mono- and bi- energetic arcs based on the water-equivalent range relative to the

distal and proximal limits of the target. However, their method exhibits limitations when used for complex target geometries. The SPARC method [17] uses pre-defined beam angles, iteratively increases the sampling frequency of beam angles, and reallocates EL to newly formed angles. Its value was demonstrated for several treatment sites [174, 178, 175, 176, 177, 179], though computation resources suffer from the systematic re-computation of beamlets at each splitting and filtering steps. Based on geometrical target coverage, RaySearch Laboratories (Stockholm, Sweden) published an early energy layer selection and spot assignment (ELSA) method in RayStation [26]. Yet, to our knowledge, no organs-at-risk (OAR) information is taken into account in this pre-optimization step. In all of these methods, pre-defined gantry angles are used. Besides, the delivery time is never explicitly modeled in the optimization.

This work aims to present a new method for efficiently and flexibly pre-selecting EL before robust spot weight optimization, reducing the high computational load associated with full DOF optimization. The algorithm employs a bi-objective function that evaluates two competing objectives, a direct approximation of beam delivery time (BDT) and the geometric target coverage considering the location of OAR. The method also benefits from a parallelization paradigm, allowing for either faster execution or running multiple instances of the problem with different parameters. Generated solutions populate a Pareto front, providing the planner with a spectrum of trade-off solutions to choose from, offering flexibility that can be tailored both to planner preferences and specific patient needs. All the results are benchmarked to IMPT and ELSA plans. The EL pre-selection code is implemented in OpenTPS [19], an open-source research treatment planning system (TPS).

5.2 Materials and Method

In this paper, PAT planning is divided into two distinct steps, following a similar approach to [153, 26, 159]. The first step, which serves as the primary focus of this study, involves the pre-selection of EL in the arc plan. The second step entails the conventional robust spot weight optimization process, which is used to achieve dose-based objectives. In this section, we describe the EL pre-selection optimization problem and its solving algorithms. The patient selected is then presented as well as the comparison strategy established for the study.

5.2.1 Energy layer pre-selection problem

The primary objective of the pre-selection process is to determine the optimal sequence of gantry angles and associated water-equivalent thickness (WET) val-

ues, eventually converted into energy values, that simultaneously maximizes the geometrical target coverage while minimizing the BDT.

During the pre-selection step, we initiate by pre-computing the WET maps $\mathcal{M}_{\text{wet}}$ through a raytracing procedure from the target voxels in the planning CT for a sequence of gantry angles within the user-specified arc range. Figure 5.1 illustrates the WET ranges computed for a full arc rotation for a phantom case.

The arc can be divided into sectors, with each sector representing an arc irradiation window during which the energy can only decrease. At each sector transition, energy can increase and make the so-called energy switch-up (SU). The decision variable is determined by the angular position of each sector (i.e., the starting angle a_i and the angular span of each delivery window w_i). This approach substantially reduces the number of variables in the problem compared to selecting a decision variable for each possible gantry angle or EL in the arc. For M sectors, We consider the input solution $\mathbf{x} = \{a_i, w_i\}_{i=1}^M$, where $a_i \leq a_{i+1}$ for $i \in \mathbb{N}$, $a_i \in [\text{minArcRange}, \text{maxArcRange}]$ and $w_i \in [\text{minWindowSize}, \text{maxWindowSize}]$ for a clockwise rotation.

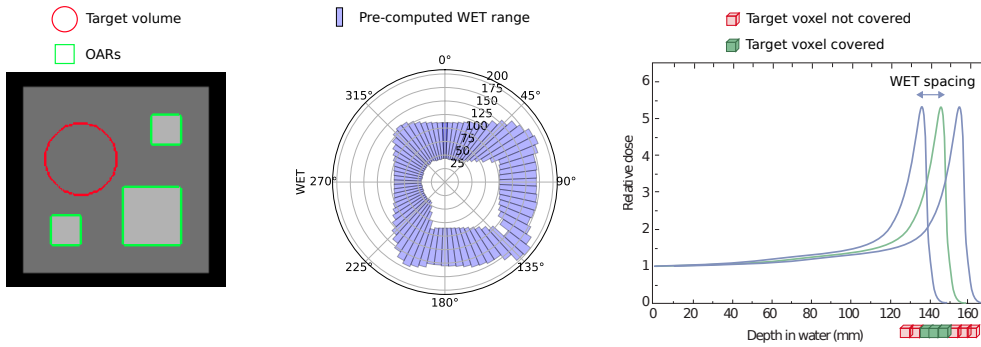


Figure 5.1: Left: Phantom CT example. Middle: Pre-computation of WET map with OAR considerations for a pre-defined sequence of gantry angles and WET sequence candidate found. Right: For each angle-layer combination selected, all target voxels that have a $\text{WET}_{\text{layer}} - \text{WET spacing} < \text{WET} < \text{WET}_{\text{layer}}$ are considered as covered voxels.

The objective function is then formulated as a multi-objective problem, divided into two parts, as shown in Eq. 5.1-5.2. The first part, f_{coverage} , is a weighted sum of several terms computed from the geometrical target coverage $\mathcal{C}$ over the N voxels in the 3D-CT image. To compute $\mathcal{C}$, we first need to know the sequence of angles in the arc. It can be easily inferred from utilizing the user-defined gantry spacing and the input solution $\mathbf{x}$. Next, we obtain the associated sequence of

WET by linearly arranging WET values from its maximum possible value from the starting angle to the minimum possible WET value from the last angle in the given sector with WET values obtained by interpolation from the look-up table $\mathcal{M}_{\text{wet}}$. Finally, the target coverage $\mathcal{C}$ is computed for each voxel by accumulating the number of times the voxel is covered by the angle-WET coupled sequence under consideration. In other words, for each angle, all voxels included in the WET band defined by the user-defined WET spacing and representing the Bragg Peak width have their coverage increased by 1, as represented on the right plot in Fig. 5.1.

From $\mathcal{C}$, we derive scalar metrics such as the 10^{th} percentile C_{low} and the minimal target coverage over the target voxels C_{min} . The exponential of minus the target coverage terms were chosen as objective functions to obtain positive objective values.

$$f_{\text{coverage}}(\mathbf{x}) = w_{\text{low}} \cdot \exp(-C_{\text{low}}(\mathbf{x})) + w_{\text{min}} \cdot \exp(-C_{\text{min}}(\mathbf{x})) \quad (5.1)$$

$$f_{\text{time}}(\mathbf{x}) = w_{\text{T}} \cdot T(\mathbf{x}) + \max(0, S(\mathbf{x}) - \text{maxSpan})^2 + \min(0, S(\mathbf{x}) - \text{minSpan})^2 \quad (5.2)$$

where

- $C_{\text{low}}(\mathbf{x}) = \text{percentile}(\mathcal{C}, 10)$
- $C_{\text{min}}(\mathbf{x}) = \min(\mathcal{C})$

and,

- $T(\mathbf{x})$: BDT approximation
- $S(\mathbf{x})$: Total arc spanning

To consider the location of some selected OAR in the pre-selection process, we intervene at two levels: at the pre-computation step of the WET maps $\mathcal{M}_{\text{wet}}$ and in the coverage objective function f_{coverage} . During the pre-processing step, target voxels are categorized depending if they are “hidden” by selected OAR or not. In this context, “hidden” implies that these target voxels are located distally to a selected OAR from the beam direction. Taking these information into account, we create a new map $\mathcal{M}_{\text{wet}}^{\text{OAR}}$ whose elements are less than or equal to zero only if they are not hidden by a selected OAR. As a consequence, the

WET range for a specific angle can be substantially shrunk compared to the situation where no OAR was considered. This limits the range of possible WET in the subsequent optimization. We also incorporate an OAR-crossing penalty weight when calculating the accumulated target voxel coverage for each sector of the input solution in the objective function evaluation. For this purpose, we first identify the hidden target voxels ($\mathcal{M}_{\text{wet}}^{\text{OAR}}(\mathcal{M}_{\text{wet}}^{\text{OAR}} > 0)$). The OAR-crossing penalty is then integrated into the objective function calculation by adjusting the current target coverage $\mathcal{C}$. Specifically, it penalizes the sectors that intersect with OAR using

$$\frac{1}{w_{\text{OAR}} \cdot \mathcal{M}_{\text{wet}}^{\text{OAR}}(\mathcal{M}_{\text{wet}}^{\text{OAR}} > 0)} \cdot \mathcal{C} ,$$

where w_{OAR} represents an additional penalty weight.

The BDT $T(\mathbf{x})$ is approximated by summing the irradiation time of each EL, the downward energy switching time between layers of a sector and the maximum between upward energy switching time and rotation time of the gantry between sectors. For our study, we considered a maximum gantry speed of 2 degrees per second, a fixed irradiation time of 1 second per layer, an upward energy layer switching time (ELST) of 5.5 seconds, and a downward ELST of 0.6 seconds. We included additional terms in the objective function to ensure that the minimal and maximal arc spans allowed are within specified bounds. They are formulated as semi-quadratics objective terms.

The weights integrated into the objective function were empirically determined. For this study, specific values were used, which are as follows: $w_{\text{low}} = 10^3$, $w_{\text{min}} = 10^4$, and $w_{\text{T}} = 10^1$. To maintain a balanced optimization process, we further increase the importance of the total coverage objective by multiplying it by a factor of 100. This adjustment ensures that the time-related objective function does not dominate the optimization, preventing the generation of unrealistic solutions with extremely short BDT but inadequate target coverage.

The objective function as stated in Eq. 5.1-5.2 allows for fast evaluation and drives the motivation of this study. Compared to a dose-based EL selection approach, in addition to reducing the number of optimization variables (spot weights $\rightarrow$ irradiation windows), it also substantially reduces the number of constraints in the objective formulation by replacing the heavy dose computation by fast geometrical target coverage and delivery time evaluation. To ensure clarity throughout the paper when discussing the EL pre-selection method developed in this paper, we will use the term “Flexi-Arc” algorithm.

5.2.2 Optimization algorithms

The two competing objectives could be combined in a single objective function using a weighting factor. The main difficulty is that the objective function is not convex at all, so a very large number of local minima exist. Furthermore, the space of solutions is large and multi-dimensional, even if it has bounds. To address these challenges, we selected global multiobjective methods based on evolutionary computation, utilizing the PyGMO (Python Parallel Global Multiobjective Optimizer) [222], a scientific Python library for massively parallel optimization.

For our study, we utilized the Generalized Island Model [223], a key feature of the PyGMO library, for solving complex optimization problems. This model enables parallel optimization by partitioning a population (i.e., potential candidate solutions to the given problem) into multiple “islands”, each operating independently with distinct optimization algorithms and/or parameters. Periodic migration of population individuals, between islands fosters diversity and helps in exploring the search space, ultimately leading to improved convergence properties. Our investigation focused on several multiobjective algorithms available in PyGMO: NSGA2 [113], MOEAD [224], NSPSO [225], and MACO [226]. The islands evolve asynchronously (i.e., running in the “background”), each within a separate process, thread or machine. The output of each island consists of evolved population where individuals are grouped on several non-dominated fronts, representing the set of optimal solutions that are not outperformed by any other solution in all objective dimensions, thereby indicating the best trade-offs among conflicting objectives. Several parameters, including the number of generations, individuals, and processes, can be adjusted during the optimization process.

Figure 5.2 depicts the workflow of the algorithm. It is executed in two main phases. In the first phase (A), we determine the number of sectors required to achieve proper geometrical target coverage. To achieve this, multiple processors run instances of the problem independently with varying numbers of sectors - e.g. ranging from 5 to 20 sectors for a full-arc rotation. Each island converges towards a non-dominated front corresponding to a specific number of sectors. By comparing the obtained Pareto fronts, we assess various trade-offs, enabling users to select the optimal number of sectors based on time and coverage efficiency.

In the second phase (B), we aim to refine the previously obtained Pareto front. For this purpose, we run the same problem instance (with the previously selected number of sectors) in parallel on different processors using a fully-connected topology. In this setup, islands are all interconnected and can exchange individuals between each other. After evolving their populations, the islands merge the solutions into several non-dominated fronts. The fast non-dominated sorting algo-

rithm available in PyGMO is then employed to retain only the solutions present on the highest ranked non-dominated front. Finally, we manually filter solutions with time objectives twice as the minimum time objective among the solutions, thus yielding a realistic set of optimal solutions. The final choice for the EL pre-selection is ultimately done based on the non-dominated front solutions displaying the BDT versus target coverage metric (e.g. the minimal target or 10th percentile coverage). The final output (C) is a list of the gantry angles and for each the associated EL found.

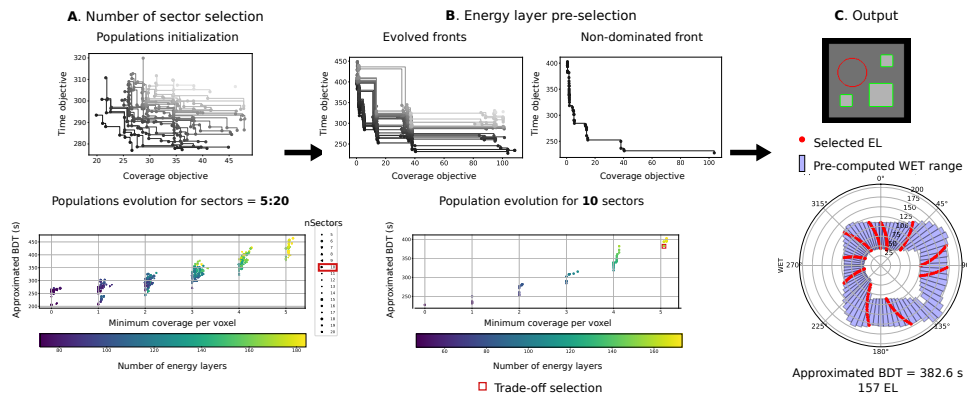


Figure 5.2: Energy layer pre-selection workflow implemented using PyGMO [222]. In Step A, we initialize a population for each possible number of sectors, allowing them to evolve towards convergence on their respective Pareto fronts. We select the 10-sector Pareto front as a suitable trade-off in this example. In Step B, we continue the evolution of the 10-sector population. After filtering and retaining the non-dominated front, we select a solution from it. Finally, in Step C, we present the solution, i.e. the selected EL for use in the dose optimization process.

The algorithm is fully integrated and openly available in OpenTPS [19], an open-source python-based TPS for research in proton therapy. Simulations were performed on a Intel(R) Xeon(R) Gold 6248 CPU @ 2.50GHz with a total of 80 CPU cores and 526 GB RAM. The computation time is highly dependent on the algorithm parameters. In this study, we use multiprocessing islands. Our approach for sector selection involved dedicating one island (i.e. one processor) per number of sectors (step A). Within each island, we initialized a population of 200 individuals that evolved over the course of 250 generations. This setup resulted in a total of 50,000 objective function evaluations per island and took approximately 5 minutes to construct all the corresponding Pareto fronts. When refining a specific Pareto front for a given number of sectors (step B), we adopted a different configuration. We employed 32 islands, evolving asynchronously over 40x10 generations, with each island having 52 individuals. The factor of 10 rep-

resents the number of evolutions during which individuals have the opportunity to migrate between the different islands. This configuration led to approximately 21,000 function evaluations per island, and the individual Pareto fronts generated using this method also took around 5 minutes to complete.

The resulting list of EL and corresponding gantry angles could now be imported in any TPS. Once the spot map corresponding to each of these layers is computed, one can proceed with the conventional robust spot weight optimization.

5.2.3 Patient data

We applied the previously described method on a young adult craniopharyngioma patient case, and benchmarked it to ELSA and IMPT plans. The anonymized patient data was retrieved from a database from Cliniques Universitaires Saint-Luc (Brussels, Belgium) for which ethical approval was granted. The prescription for this craniopharyngioma is 54 Gy in 30 fractions.

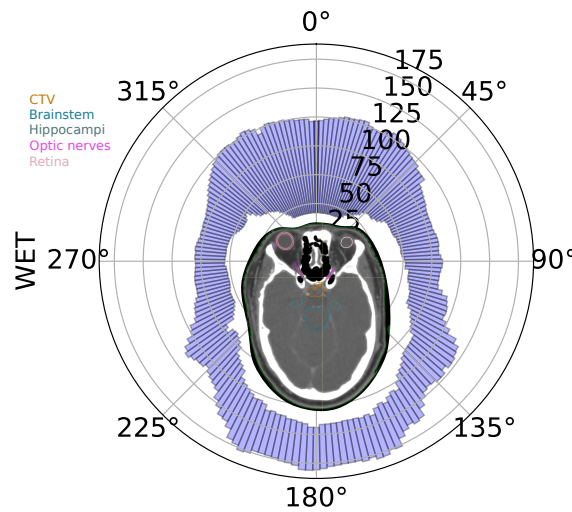


Figure 5.3: Patient geometry used for the study displayed with pre-computed WET ranges.

Treatment planning optimization was done in a research version of RayStation 2023B (Raysearch Laboratories, AB). An IMPT plan was generated with a 3-beam configuration including a right–anterior–superior beam (gantry $[G] = 70^\circ$, couch $[C] = 220^\circ$), a left–anterior–superior beam ($[G] = 70^\circ$, $[C] = 320^\circ$) and an oblique vertex beam ($[G] = 315^\circ$, $[C] = 90^\circ$). The proton arc plans were designed with a full 360° arc. The ELSA algorithm used in this study slightly differs from the original publication [26] in the sense that it produces an EL pre-selection with a

more predefined sector length, which explains the consistency in the number of sectors and associated energy SU in the plans. Several gantry angle spacing values were tested for the ELSA plans: we designed plans with 2° , 3° and 4° as gantry angle spacing, corresponding to 180, 120 and 90 EL. This matched with using 8, 5 and 5 sectors, resulting in 7, 4 and 4 SU respectively. The Flexi-Arc plans were generated with a 1-degree angle spacing to give more freedom to place sectors and their EL in the full arc rotation. The number of sectors and EL for Flexi-Arc is defined based on the trade-off solution picked on the Pareto fronts produced and is studied in Section 5.3. WET maps were pre-computed for each angle as well as the corresponding WET range taking into account two selected OAR: the optic nerves and the hippocampi. WET range was therefore slightly shrunk for angles that crosses these OAR to reach the target voxels. The brain anatomy and pre-computed WET ranges for the patient selected is displayed on Fig. 5.3.

After the EL pre-selection (either ELSA or Flexi-Arc), Monte Carlo dose calculation and robust worst-case optimization [80] were done considering 21 scenarios. These scenarios comprised three proton range error scenarios (-3% , 0% , 3%) combined with six setup isotropic error scenarios of 3 mm plus the nominal setup scenario, all considered for the CTV volume and brainstem. Same settings were used to generate the IMPT plan.

In summary, we look for an optimal EL selection using PyGMO algorithm, optimize the plan within RayStation and finally compare the clinical goals to the ones obtained with the ELSA and IMPT plans which will be described in the next subsection.

5.2.4 Plan evaluation

Table 5.1 lists the clinical goals in the field of neuro-oncology (adapted from [227]) we referred to for the planning. Doses were recalculated to equivalent dose in 1.8 Gy-fractions using the equivalent dose in 2 Gy-fractions (EQD2) formula $EQD2 = \frac{n \cdot d(d + \alpha/\beta)}{2 + \alpha/\beta}$ where $n = 30$ is the number of fractions used for the treatment and d is the unknown computed for each dose constraint.

Several other metrics of interest were also computed such as the homogeneity index (HI) as defined in ICRU 83, conformity index (CI) [228] and body integral dose (ID). Additionally, a robust evaluation of each plan was performed using the same parameters as for the robust optimization.

Finally, while we use a very rough approximation of the BDT of the EL set in the pre-selection stage, we would like to use a more reliable estimation of the BDT for the final plan comparison. We therefore computed the dynamic BDT with the

Region of interest (ROI)	α/β (Gy)	Clinical goal
CTV (nominal)	/	D98% > 52.92 Gy (98% Dp = 52.92 Gy)
CTV (worst-case)	/	D98% > 51.3 Gy (95% Dp = 51.3 Gy)
Brain	2	V60Gy $\leq$ 3 cc
Brainstem	2	Surface: D0.03 cc $\leq$ 60 Gy Interior: D0.03 cc $\leq$ 56 Gy
Chiasm & Optic nerve	2	D0.03 cc $\leq$ 56.5 Gy
Cochlea	3	Dmean $\leq$ 48.5 Gy
Hippocampus	2	D40% $\leq$ 12 Gy
Lacrimal gland	3	Dmean $\leq$ 31 Gy
Eyes	2	Dmean $\leq$ 20 Gy
Lens	1	D0.03 cc $\leq$ 18.5 Gy
Retina	3	D0.03 cc $\leq$ 48.7 Gy

Table 5.1: List of clinical goals in neuro-oncology adapted from [227] for the current fractionation scheme (30×1.8 Gy (RBE)). Dp = dose prescription, D0.03 cc = near maximum dose to 0.3 cc of structure/organ, Dmean = mean dose, Dx% = dose to x% of the considered volume.

ATOM¹ (Arc Trajectory Optimization Method) algorithm, an open-source tool published by RaySearch Laboratories [229]. ATOM is designed to determine a fast plan delivery while adhering to the mechanical constraints specified by the user. The machine parameters utilized in this study are detailed in Table 5.2. The selection of ATOM was driven by its utility, given that no proton arc machine has yet been established and commissioned in clinical settings; ATOM serves as a reliable estimator at present.

Maximum gantry velocity	2°/s
Maximum gantry acceleration	0.5°/s ²
Maximum gantry jerk	0.5°/s ³
Downward energy switching time	0.6 s
Upward energy switching time	5.5 s
Spot switching time	2 ms
Spot delivery time (per MU)	5 ms

Table 5.2: Machine constraints parameters used for BDT estimation from ATOM [229].

For clarity, when using ATOM to compute the BDT, we will refer to the *true* BDT while we will refer to the *approximated* BDT for its quick and rough estimation during the pre-selection stage.

¹<https://github.com/raysearchlabs/ArcTrajectoryOptimizationMethod>

5.3 Results

Figure 5.4 depicts the Pareto fronts, which illustrate the trade-offs between the approximated BDT and the minimum target voxel coverage. We conducted 21 parallel instances of the problem, varying the number of sectors from 5 to 20, hence the 21 Pareto fronts. The observed trend indicates that as we increase the number of sectors, the Pareto front shifts upward, meaning improved target coverage but longer BDT. Interestingly, it is the total number of EL that has a more substantial impact on BDT than the number of sectors. Moreover, the use of more sectors allows for a greater number of energy layers to be incorporated (as angle spacing is fixed), thereby enhancing target coverage. We also assessed the ELSA EL pre-selections regarding these two metrics employed for Flexi-Arc EL pre-selection (indicated by the red dots in Fig. 5.4). It is important to note that while ELSA pre-selections may not appear Pareto-optimal on this plot, the comparison is not entirely fair because ELSA has been optimized using a different definition of these criteria, and thus the 'worse' value observed for ELSA were somewhat expected. We chose to pick two Pareto fronts to continue the study; the 8-sector front to provide a direct comparison with 8-sector ELSA pre-selection and the 16-sector front that is more shifted to the right on the minimum coverage axis and present data points with low BDT approximation (not achievable with the ELSA method).

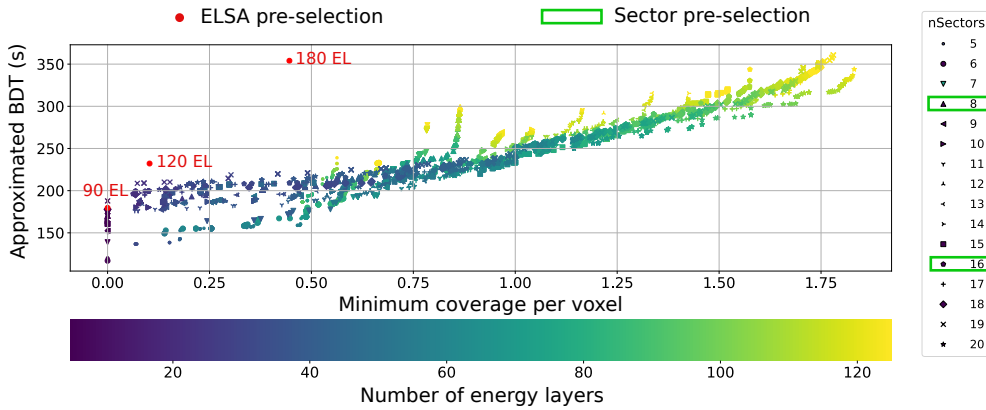


Figure 5.4: Non-dominated fronts for several numbers of sectors. Minimum coverage and BDT increases with the number of sectors. ELSA solution set, which included 90, 120, and 180 EL, was also evaluated in relation to these two objectives (though not optimized using the exact same definition).

Next, we conducted separate runs of the 8-sector and 16-sector problem instances. After filtering out non-realistic solutions and retaining only the highest-

ranked non-dominated solutions, we get the Pareto fronts, as shown in Fig. 5.5. As anticipated, the 16-sector solution achieves a higher minimum target coverage compared to the 8-sector solution. The 8-sector front completely dominates the ELSA pre-selections. The Pareto fronts in these plots clearly show that the BDT increases as the total number of EL increases. We carefully selected one data point from each of these Pareto fronts, taking into account not only achieving a reasonable BDT and a high minimum coverage but also conducting individual visual inspections of the EL layout on a polar plot to avoid configurations with unique EL within sectors, for instance. The retained EL pre-selection include 87 and 85 EL for 8- and 16- sector solution respectively (green squares on Fig. 5.5).

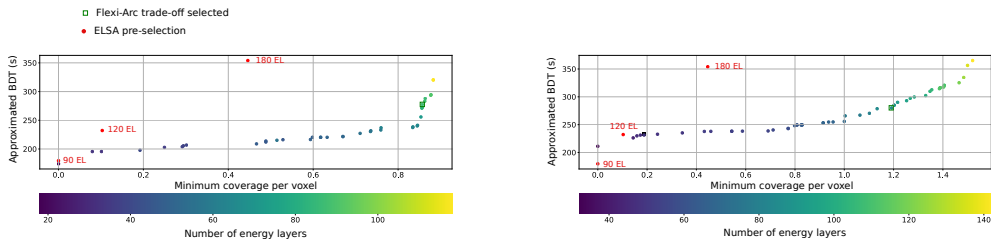


Figure 5.5: Non-dominated front refined for 8 sectors (left) and 16 sectors (right). Minimum coverage per voxel and BDT increases with the total number of EL. Final trade-off EL pre-selected shown in a green square.

Figure 5.6 presents two rows of information. In the first row, we display on a polar plot the EL pre-selection obtained using the two different methods: the Flexi-Arc method for both 8- and 16-sector configurations and the ELSA method for the three gantry angle spacing values under evaluation. The Flexi-Arc pre-selection naturally spaces the sectors, accommodating dead angles, useful when potentially crossing the selected sensitive structures (optic nerves and hippocampi). It also minimizes the gantry braking to accommodate the SU. In contrast, ELSA exhibits a more regular pattern with no consideration for OAR avoidance and no gaps between sectors. The appearance of some EL falling outside the WET range calculated for the ELSA pre-selection could be explained by the WET range computation already considering the selected OAR in our implementation but also from a difference of implementation for WET raytracing.

In the second row, we present the dose distributions obtained after optimizing the spot weights for each of these EL pre-selection methods within RayStation. The dose optimization was performed individually for each plan to maximize their quality. It seems from these dose distributions that the best target conformity was achieved with the 16-sector Flexi-Arc and the 2-degree spacing ELSA plans. The 16-sector Flexi-Arc plan effectively avoids anterior field irradiation to attempt

to spare the optic nerves system and employs oblique angles to protect the hippocampi as intended.

In Table 5.3, we present important plan quality metrics. The dose was normalized so that D98 corresponds to the CTV prescription, meeting the clinical goal outlined in Table 5.1. As anticipated, the highest CI is achieved with the 2° ELSA plan and the 16-sector Flexi-Arc plan. The CI deteriorates for the 3° and 4° ELSA plans, resulting in a higher integral dose to the body. The best HI (lowest) was obtained with the ELSA 2° plan, likely due to the increased number of EL. Additionally, the number of spots is much higher for the ELSA plans. When comparing the true BDT among the arc plans, the Flexi-Arc plans achieve the shortest time, aligning with the reduced number of EL needed for target coverage and the gap between sectors. For instance, the 16-sector Flexi-Arc plan achieves a reduction of 16%, 18% and 40% compared to the 4°, 3° and 2° ELSA plans respectively.

Now, in terms of a comparison with the IMPT plan, it is worth noting that the IMPT plan already exhibited very high quality, with an even lower integral dose compared to the arc plans. Regarding the BDT, it is important to consider that, for IMPT, it is calculated individually for each beam, resulting in what we refer to as “static delivery time”, in contrast to the dynamic delivery time of proton arc plans. The static BDT for this IMPT plan was estimated to be 36s. To make a fair comparison between static and dynamic delivery times, we need to account for manual interventions required between each fixed-angle beam delivery in the control room and/or treatment room. This involves tasks such as beam loading in the console, gantry and/or couch rotation, and potential re-imaging steps. After consulting with two separate proton centers, we introduce a simulated delay of 60 seconds between each beam, plus an additional time of 120 seconds required for each couch rotation. Considering the three IMPT beams with different couch angles, this results in an extra intervention time of 360 seconds, making the IMPT plan longer to deliver compared to the proton arc plans. The 16-sector Flexi-Arc plan was therefore 57% faster to deliver than the IMPT plan.

Finally, in our evaluation, we focused not only on target coverage and BDT but also on the dosimetry of OAR. Figure 5.7 provides a comparison of dose metrics related to the clinical OAR goals for the Flexi-Arc 16-sector plan and the ELSA 180 EL plan, relative to the corresponding metrics in the IMPT plan. Negative differences indicate that the arc plans outperformed the IMPT plan in that metric, meaning that dose was reduced to that OAR. Overall, the Flexi-Arc plans tend to better spare the OAR compared to ELSA, with the exception of the optic nerves, despite being selected as critical OAR during EL pre-selection. However, it was effective in sparing the hippocampi. Nevertheless, it is important to note that the

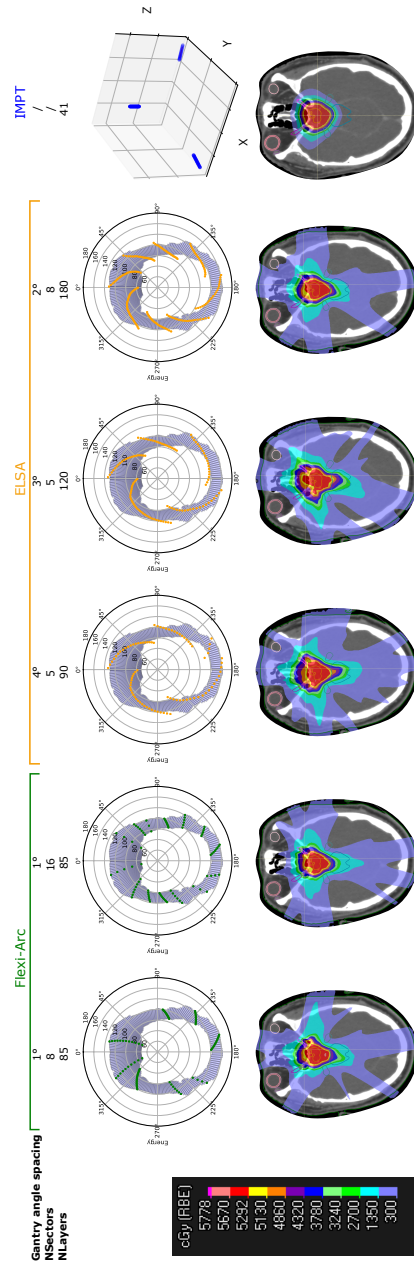


Figure 5.6: Energy layer pre-selection for each trade-off selected using the Flexi-Arc method and for each ELSA plan. Final dose distributions are displayed below after spot weight optimization in the RayStation TPS.

	Flexi-Arc		ELSA		IMPT	
Gantry angle spacing (°)	1	1	4	3	2	
NSectors	8	16	5	5	8	
D98(CTV)-nom (Gy [RBE])	54.0	54.0	54.0	54.0	54.0	54.0
D98(CTV)-wc (Gy [RBE])	53.2	53.2	53.4	53.5	53.5	53.8
D2(CTV)-nom (Gy [RBE])	56.3	56.0	56.6	56.3	56.0	56.5
D2(CTV)-wc (Gy [RBE])	57.0	56.5	57.0	56.8	56.4	56.6
CI	0.26	0.28	0.19	0.20	0.28	0.27
HI	0.041	0.045	0.047	0.041	0.037	0.045
Patient ID (Gy · L)	13.2	13.2	20.7	20.8	12.5	10.1
NSpots	2816	2019	5817	6206	4834	2620
NLayers	87	85	90	120	180	41
True BDT (s)	182	171	203	209	245	/

Table 5.3: Dosimetric evaluation and plan metrics. *Abbreviations : D98(CTV) = dose received at 98% of the CTV, nom = nominal, wc = worst-case, CI = conformity index, HI = homogeneity index, ID = body integral dose, BDT = beam delivery time obtained with ATOM [229].*

differences between ELSA and Flexi-Arc plans are generally quite small. Most data points in the figure are located to the right of the vertical central axis, indicating that IMPT better spares most of the OAR. IMPT especially stands out in the case of the ocular system, thanks to the use of non-coplanar beams. However, it is worth mentioning that doses to the brainstem, chiasm, and optic nerves could be effectively reduced using an arc beam. A full detailed dosimetric evaluation of the OAR in the nominal and worst-case scenario for all the plans generated is provided in Appendix C.

5.4 Discussion

In this study, we introduced a new method, Flexi-Arc, for pre-selecting energy layers in the context of proton arc therapy treatment planning optimization. To validate this method, we applied it to a test case involving a young adult cranio-pharyngioma and generated two robust PAT treatment plans using 8 and 16 sectors. We compared these plans with the state-of-the-art ELSA pre-selection method by RaySearch [26] and a conventional IMPT plan with three beams.

The results demonstrate that the Flexi-Arc method generated EL pre-selections with a good target coverage and a short BDT for the patient under study in a very short computation time. Approximately 85 energy layers were sufficient to meet clinical goals. Flexi-Arc plans outperformed ELSA plans with 90 and 120

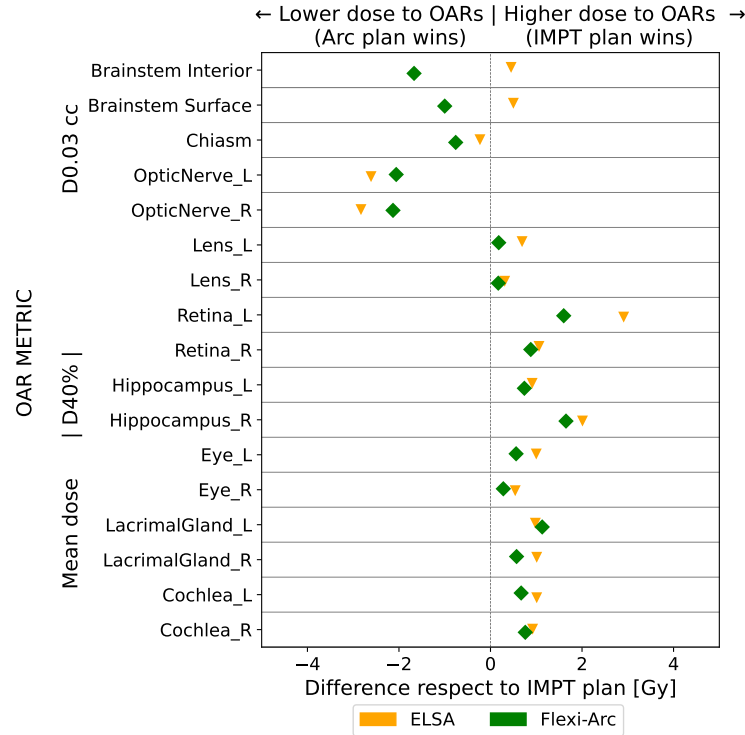


Figure 5.7: Differences in terms of clinical planning criteria for the OAR between the 16-sector Flexi-Arc plan, the 2° spacing ELSA plan relative to the IMPT plan. Negative differences imply that the arc plan was better than the IMPT plan for that metric (i.e., less dose to the OAR).

energy layers due to better sector and EL pre-selection. Their quality was similar to that of the ELSA plan with 180 energy layers, while achieving a 40% reduction in BDT.

From a purely dosimetric standpoint, the PAT and IMPT plans are likely clinically comparable for this patient. The IMPT plan showed high quality, particularly in sparing the ocular system, thanks to the use of non-coplanar beams. Nevertheless, this translated into a longer delivery time due to couch-kick and gantry rotations required for each beam to deliver. It may have been observed that, in this study, the integral dose is higher for all PAT plans compared to the IMPT plan, which contradicts claims and *in silico* planning evidence suggesting that PAT reduces integral dose [26, 179, 174, 178]. One potential explanation for this difference is the use of high WET arc directions. The major gain of the PAT plans generated for this study is therefore their faster delivery and simplified workflow

with a single arc to deliver compared to the 3-beams IMPT plan. Nevertheless, it may be beneficial to consider using multiple non-coplanar arcs in PAT if it results in significant dosimetric improvements, even if it increases the treatment time (to a reasonable extent). This was actually demonstrated to reduce radiation dose to the hippocampi in VMAT plans of craniopharyngioma patients [230]. A second potentially important advantage of PAT over IMPT, although not investigated here, is the increased incidence of beams, which could help to “dilute” the areas where there may be issues with high linear energy transfer at the end of the proton beam range and the associated potential toxicity [231, 159]. The choice of treatment modality ultimately rests with the physician, who can best assess the trade-offs.

A major advantage of the Flexi-Arc technique lies in the freedom it offers for sector placement within the arc. Through meta-heuristics, it optimizes the start and stop positions of each sector to minimize BDT, maximize target coverage, and improve OAR sparing. The sector optimizer will then take benefit of “dead angles” or passage above an OAR without irradiation to make the SU, reducing the need to brake and, hence ensuring smooth gantry rotation velocity. The method allows for a less sector-dependent approach with BDT being more dependent on the total number of energy layers used. It also enables planners to tailor treatment plans to each patient’s specific needs, in particular by prioritizing OAR to spare. In contrast, ELSA’s sector spacing is fixed by user-defined gantry angle spacing. As a result, it may require gantry slowdown or null velocity if spacing is too small to allow time to make the SU before the next sector begins, thus prolonging the BDT. ELSA’s BDT is therefore highly dependent on the pre-determined number of sectors, which, in turn, relies on user-defined arc beam range and angle spacing. In this study ELSA plans were initialized with a full rotation with 2, 3, and 4° gantry angle spacing. The plans could have been initialized differently based on planner’s intuition but the configuration choice is not straightforward. An experienced planner would have most probably tried to restrict the arc beam range such that it avoids some critical structures. For example, they could have used two partial arc beams with finer spacing. But, our method could be useful to compute the best arc directions and present that to the planner for consideration. This demonstrates that the planning process is not only dependent on the method but also on the planner’s experience and strategies in employing the available degrees of freedom.

One notable advantage, as well as a potential bottleneck, in the proposed EL pre-selection method is the utilization of evolutionary algorithms with parallel computation. These algorithms do not rely on derivatives, making them suitable for complex problems as the one we are dealing with in this article. This approach

offers the advantage of fast optimization and the exploration of the extensive parameter space for EL pre-selection. It enables us to assess EL pre-selection with varying numbers of sectors and refine the Pareto front for specific sector number by employing different starting points and algorithms. It then ensures a diverse initial population and a wide covering of objectives values. Yet, it is important to note that these metaheuristics do not guarantee optimal solutions and require careful parameter tuning. While we consider the non-dominated front as the Pareto front for our analysis, we cannot guarantee that it is indeed the true Pareto front. Points do converge to it, but the extent of convergence may vary depending on factors such as the number of generations, individuals, and other algorithm-specific parameters. Running the problem with different settings might lead to further convergence and refinement of the Pareto front.

Despite its advantages, the Flexi-Arc method introduces additional complexity due to numerous parameters in the EL pre-selection phase. These include parameters for evolutionary algorithms mentioned above but also weighting of sub-objective terms, OAR-crossing penalty weights, WET spacing, arc range, and gantry spacing. The choice of OAR for the pre-optimization step is also important. Incorporating too many OAR can become overly difficult for the optimizer as it does not have much freedom to place the sectors without being penalized. We therefore encourage to prioritize OAR that the planner wishes to absolutely avoid. While this complexity allows for diverse exploration and potential improvements, it also poses a challenge for planners. The decision-making process requires thoughtful parameter tuning to achieve desired outcomes.

Another point of discussion is considering whether the objective terms used in our study, as surrogates for dynamic BDT and plan quality, are suitable metrics. While there is a trend suggesting a correlation between geometrical target coverage and plan quality, the reflected dosimetric differences between two data points are much smaller than the large geometrical target coverage difference that could be observed in the Pareto plot. According to the results, dosimetric plan quality appears to be more related to the number of energy layers and sectors rather than geometrical target coverage directly, which may result from the use of a higher number of EL or sectors. Therefore, one should exercise caution when selecting the EL pre-selection with the highest minimal target coverage, as it may not necessarily translate to the highest plan quality. In the future, our estimation of BDT could be improved by considering factors like gantry acceleration. However, since we don't know the number of spots per layer in the pre-selection stage, we will still be limited in modeling the actual expected layer delivery time.

Additional improvement in the Flexi-Arc method can also be made. For instance,

refining the WET sequence generation within sectors could optimize coverage further instead of using the simple linear WET spacing between the sector start and end. Incorporating robustness parameters directly into the objective function during pre-selection optimization is another avenue for enhancement. Finally, applying this method to a broader range of patient cases and treatment sites will help demonstrate its potential, versatility and limitations.

5.5 Conclusion

In conclusion, the Flexi-Arc method introduces an efficient approach to EL pre-selection in PAT treatment planning. This pre-selection involves optimizing the angular position of decreasing energy layers within a specified arc range. The solutions are evaluated using a fast bi-objective function, which combines a metric for geometrical target coverage and a direct estimate of dynamic beam delivery time. This evaluation is performed through multi-objective evolutionary algorithms. The resulting solutions form a Pareto front, allowing the planner to choose the best trade-offs based on the clinical context. Flexi-Arc EL pre-selection achieves high plan quality after dose optimization and expedites delivery time compared to state-of-the-art EL pre-selection method. While it presents challenges related to parameter complexity, its advantages in terms of flexibility, parallel computation, and high potential for subsequent spot weight optimization make it a promising avenue for treatment planning in proton arc therapy.

Chapter 6

Bridging the gap

This chapter is titled 'Bridging the Gap' to reflect its aim: to join the more optimal dose-based energy layer selection with the fast geometrical-based energy layer pre-selection. This bond seeks to overcome the limitations inherent in both methods, thereby enhancing the optimization process for proton arc treatment planning. This work was produced in the context of a 3-months secondment at RaySearch Laboratories (Stockholm, Sweden) under the supervision of Dr. Erik Engwall. In what follows, we provide a brief overview of the main results.

Summary

Purpose: This study aimed to develop a dose-based energy selection method for proton arc therapy (PAT) treatment planning, combined with a fast energy layer pre-selection step, in a research version of RayStation 2023B (RaySearch Laboratories AB, Stockholm), addressing limitations in the existing energy layer (EL) selection process.

Methodology: The approach involved expanding the initial set of EL candidates and subsequently using a cardinality operator to penalize and filter these layers during spot weight optimization. The cardinality operator constrained the maximum number of EL per gantry angle or in the entire arc, leveraging dose-based knowledge in EL selection.

Results: Comparisons were made between the baseline proton arc plan and filtered plans with different EL initialization strategies. The study observed marginal improvements in organs-at-risk (OAR) sparing with filtered plans, with minimal changes in clinical target volume (CTV) coverage. Filtering strategies were particularly effective in cases with reduced baseline plan quality, highlighting the method's impact on plan improvement. Nevertheless, the followed strategies, inevitably prolongs the resulting filtered plan delivery time compared to the baseline plan.

Conclusion: This study introduced a dose-based EL selection method, to be used in combination with a fast EL pre-selection, for PAT treatment planning. Unfortunately, it could not demonstrate clear potential for enhancing plan quality, except in cases with reduced baseline plan quality. Further exploration of filtering strategies and associated parameters is warranted. Trade-offs between plan quality and beam delivery time must be carefully considered for clinical feasibility.

6.1 Motivations

The existing RaySearch dynamic proton arc therapy (PAT) treatment planning module relies on an early layer selection and spot assignment (ELSA) [26], which pre-selects energy layer (EL) before the actual spot weight optimization. While this method is fast, it does not include any dose-based knowledge in the EL selection process. Consequently, the EL chosen for the subsequent optimization may sometimes be suboptimal, as the pre-selection restricts many degrees of freedom. The objective of this work was to develop a dose-based energy selection method for PAT treatment planning in a research build of RayStation 2023B (RaySearch Laboratories, Stockholm, Sweden).

6.2 Methods

The approach we used is conceptually straightforward: we start by increasing the initial set of EL candidates and subsequently penalize and filter these energy layers during the spot weight optimization until we reach a plan with only one EL per gantry angle.

To penalize and reduce the set of energy layers, we introduce a cardinality operator formulated as follows:

$$\left| \forall b \in B : \sum_{e \in E_b} y_{be} > 0 \right| \leq N . \quad (6.1)$$

Here, B represents the set of beam angles, E_b denotes the set of energy layer candidates in angle b , y_{be} represents the sum of spot weight in energy layer e for angle b , and N is the upper limit on the total number of energy layers.

In practice, this formulation is a bit more complex to implement. Firstly, the RaySearch spot weight optimizer is gradient-based and performs continuous optimization, making it unable to handle such constraints. Secondly, the set of EL candidates still needs to be initialized carefully to ensure a reasonable number of energy switch-up (SU) in the arc.

To address these challenges, we replace Eq. 6.1 and approximate it with a sigmoid (smoothed step function):

$$\sum_{b \in B} \Theta \left(\sum_{e \in E_b} y_{be} \right) \leq N , \quad (6.2)$$

where $\Theta(x)$ is defined as:

$$\Theta(x) = \frac{1}{1 + e^{-x/\epsilon}} .$$

Here, ϵ controls the smoothness of the sigmoid function.

To initialize the set of EL candidates, we explored several approaches. We introduced the concepts of local and global cardinality constraints. A local constraint introduces a cardinality constraint for each angle, restricting it to one EL per angle ($N = 1$). A global constraint considers the maximum allowed number of EL in total (i.e. for all angles). Given that the PAT challenge involves minimizing both the number of EL and SU, increasing the set of EL must be done cautiously. If one wishes to add many EL candidates per angle, a new penalty function is required to minimize the number of SU in combination with local or global constraints.

To keep things relatively simple, we decided to retain the initial energy sequence obtained by ELSA but extended it to two revolutions or employed a finer gantry spacing. We combined this with a global constraint set to $N = \frac{\#EL}{2}$. This typically doubled the number of EL candidates ($2N$) and SU compared to the baseline plan.

The workflow is then:

1. Run spot weight optimization with penalty
2. Evaluate the EL candidates, remove the worst (lowest MU) and update steepness
3. (Repeat)

Therefore, our objective is to assess whether starting with $2N$ EL and subsequently filtering them down to N EL yields a superior plan compared to optimizing the baseline plan directly with only N EL. The EL filtering process was designed to activate the penalty function after the 5th iteration. After the 50th iteration, half of the EL were filtered at once, and spot weights were reset to their initial

values. Subsequently, we continued the optimization with an additional 250 iterations, totaling 300 iterations (with spot filtering occurring at the 250th iteration). In contrast, the baseline plan was optimized using a total of 250 iterations. This means that an additional 50 iterations were necessary to see some actual gain with the filtered plans. These numbers resulted from a manual parameter tuning study.

We conducted our evaluation primarily on patients with bilateral oropharyngeal cancer. Our assessment includes comparisons of the following aspects:

- CTV robust coverage (voxelwise minimum)
- Mean doses to OAR (voxelwise maximum)
- NTCP for dysphagia grade 2 and 3, as well as xerostomia grade 2 and 3.

For more detailed information regarding planning goals and robust settings, one can refer to Chapter 7, which pertains to the same group of patients. To keep it concise, we present results for a single patient.

6.3 Results

Figure 6.1 presents the DVH comparison between the baseline plan and two filtered plans using the 2-revolutions and half-gantry-spacing initialization strategies. We observe some differences in terms of OAR sparing, with the filtered plans showing slight improvements.

Figure 6.2 displays the final EL selection for each of the plans. It is now clear that the 2-revolution plans and half-gantry-spacing plans result in double the number of EL candidates compared to the baseline. The majority of filtered EL are situated on opposite lateral sides. Several explanations could account for this behavior. First, the optimizer may tend to avoid directions that go through the shoulders, which could require higher energy levels and deliver increased doses to the body. Additionally, the position of the primary tumor and the relative positions of the OAR in relation to the tumor may also contribute to this EL selection pattern.

The bottom row of Figure 6.2 presents DVH metrics and NTCP results. It appears that CTV coverage slightly worsens with the filtered plans, while mean doses to OAR show slight decreases. Consequently, NTCP values also decrease for all studied toxicities. However, it is important to note that the NTCP values presented were normalized for ease of comparison. The absolute values reflect the true NTCP value and should matter the most. As a result, while a 20% decrease

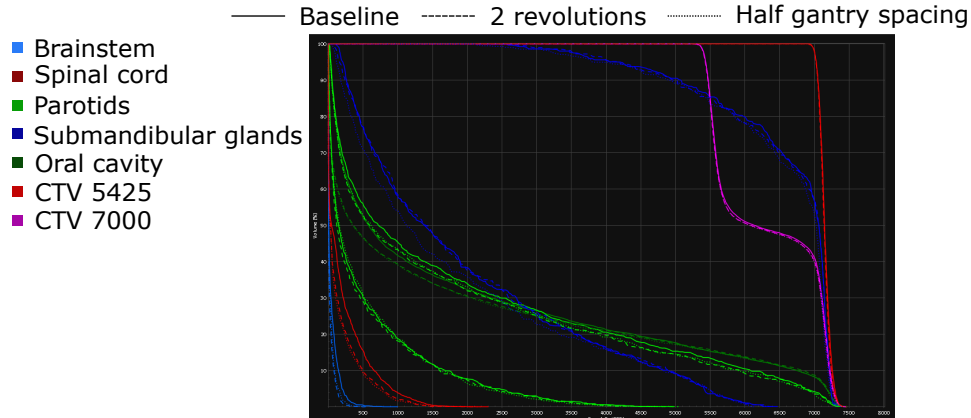


Figure 6.1: DVH comparison between baseline plan and filtered plans with initial set of EL candidates doubled either using two arc revolution or half gantry spacing strategy.

is observed for dysphagia grade 2, this corresponds to only a 0.2% reduction from an NTCP absolute value which is close to 1 for this toxicity. In this context, the filtered plans do improve NTCP values, but the improvements are still marginal compared to the baseline plan.

Figure 6.3 provides a comparison of DVH and NTCP, this time comparing the filtered plans to a discrete arc plan that was generated with 30 directions and 360 EL. As depicted in the polar plot, the EL in the discrete arc plan are stacked in the 30 directions, resembling to an IMPT plan but with many more irradiation angles. The DVH and NTCP plots show that the filtered plan (half-gantry-spacing) actually achieves a similar quality compared to the discrete arc plan. Consequently, discrete and filtered plans both achieve the same level of quality improvement when compared to the dynamic baseline plan.

Finally, in the next set of experiments, we repeated the process with larger gantry spacing, resulting in plans with a reduced number of EL, which is expected to yield lower baseline plan quality. Figure 6.4 compares DVH metrics for three baseline plans, each with 360, 180, and 120 EL, with three filtered plans. The filtered plans were initialized with double the number of EL but were then reduced to the same number as in the respective baseline plan. It is observed that the filtered plans achieve a similar quality compared to each other, with a slight decrease as the number of EL decreases. In contrast, the baseline plans suffer more in terms of plan quality as the number of EL decreases. This demonstrates the effectiveness of the filtering method in improving plan quality, with its impact becoming more

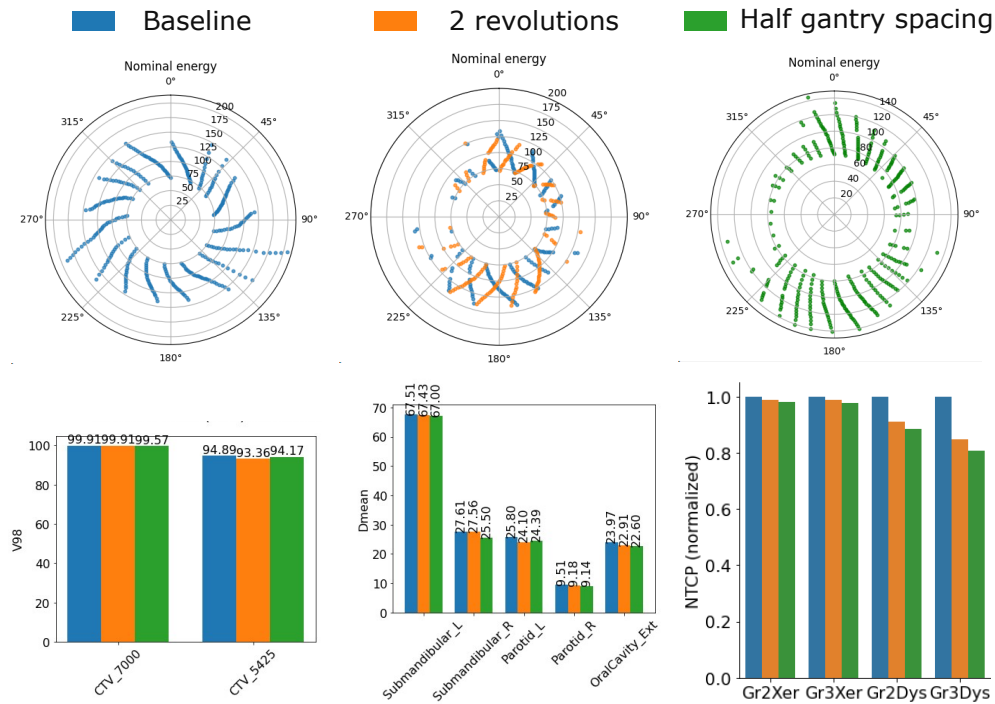


Figure 6.2: Upper: Final set of EL chosen after optimization and filtering. Lower: Robust DVH metrics comparison between the plans using voxelwise min (CTV) and voxelwise max (OAR) as well as NTCP comparison. Gr2(3)Xer = grade 2(3) xerostomia; Gr2(3)Dys = grade 2(3) dysphagia. NTCP magnitude order: Gr2Xer ~ 30 ; Gr3Xer ~ 8 ; Gr2Dys ~ 7 ; Gr3Dys ~ 1 .

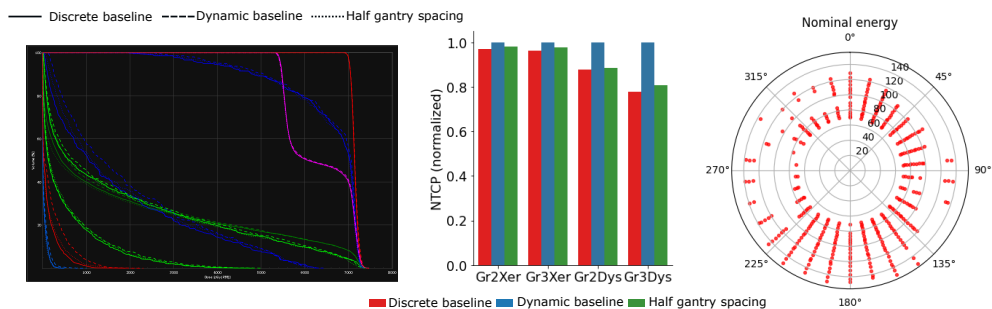


Figure 6.3: DVH and resulting NTCP comparison between discrete baseline plan (30 directions and 360 EL), dynamic baseline plan (360 directions, 360 EL) and filtered plan with half gantry spacing strategy (720→360 directions/EL).

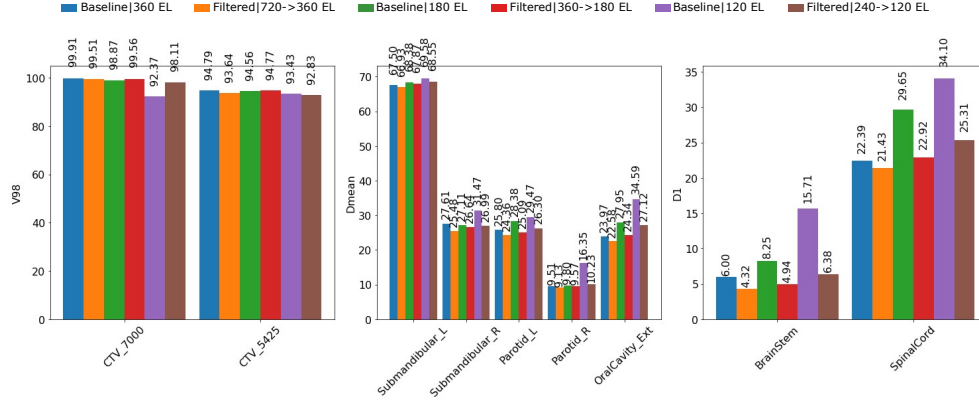


Figure 6.4: DVH metrics comparison for different gantry spacing (resulting in different number of EL) between baseline and filtered plans.

important as the baseline plan deteriorates.

6.4 Discussion and conclusion

In this study, we introduced a dose-based EL selection method for proton arc therapy treatment planning. This method was integrated directly into a research version of a commercial TPS, RayStation 2023B. This approach involves increasing the initial set of EL candidates and then reducing it using a cardinality operator that constrains the maximum number of EL, either in total or per direction.

The results suggest that there is minimal to no improvement when compared to a solution that does not filter energy layers, when the initial set of EL is simply doubled. Additionally, it is expected that the beam delivery time (BDT) will increase when using two revolutions or finer gantry spacing because it also effectively doubles the number of SU and consequently impacts the gantry rotation speed. While this might initially seem discouraging, we also observed that the baseline plans might already be of very high quality, making potential improvements marginal. This observation was further supported when we compared the plans to discrete arc plans, which are often considered the gold standard for proton arc therapy planning. The baseline dynamic plans were already quite close in quality to the discrete arc plans. When applying the dose-based EL selection, we observed that the filtered plans reduced the plan quality gap to the discrete arc plan. This suggests that the dose-based EL selection method is effective and correlates with initial baseline plan quality. If the quality of the discrete arc plan is already close to that of the dynamic plan, further improvements using a dose-

based EL selection may not be expected, as the dynamic baseline plan is already quite optimal. Another possible explanation could be related to our choice of EL initialization, which might still be too restrictive and does not provide enough freedom for the subsequent optimization process.

In addition to the choice of EL initialization strategy, there are other parameters related to the filtering process that had be considered. These parameters include whether to state the cardinality as a hard constraint or a soft objective. In the latter case, a penalty weight must be assigned to control the strength of this objective relative to the dose objectives. Another important parameter is how many times to filter EL, whether it should be done in a single step or at regular intervals. The steepness of the sigmoid function used in the filtering process is yet another parameter that can be adjusted. Several runs on different disease sites, although not presented here, have indicated that additional iterations compared to the baseline plan are required to see some improvements. Therefore, these numerous parameters must be fine-tuned, which can present a challenge in itself.

Our study has shown that even though we can achieve improvements in plan quality through dose-based EL selection, it often comes at the cost of increased BDT, particularly when using strategies involving two revolutions or finer gantry spacing. This highlights the need for careful consideration of important trade-offs. Additionally, there may be potential for further exploration of strategies, such as local constraints with a higher number of initial EL candidates per direction. However, implementing such strategies requires the formulation of new objective terms to penalize the number of SU in the arc, in order to avoid excessively large increases in BDT. This approach has been explored in previous studies [16, 155] as well as in Chapter 3 of thesis. We made efforts to include a similar objective term into our study; however, we faced challenges in determining the appropriate weight for this new term in relation to the sparsity and dose objective functions. Unfortunately, the results from these attempts were mixed and, as such, are not reported in this study.

In summary, while our study has shed light on the potential benefits and challenges of dose-based EL selection in PAT planning, there are still important considerations and trade-offs that must be thoroughly investigated to achieve a well-balanced and clinically feasible approach.

Chapter 7

Optimizing efficiency

The content presented in this chapter comes from a manuscript submitted to Medical Physics under the title: *“Faster proton arc optimization and delivery through energy layer pre-selection and post-filtering”*. This work is one of the main contributions resulting from a stay of three months at RaySearch Laboratories under the supervision of Erik Engwall.

Authors: Sophie Wuyckens¹, Viktor Wase², Otte Marthin², Johan Sundström², Guillaume Janssens^{3,4}, Elena Borderias-Villarroel¹, Kevin Souris^{1,4}, Edmond Sterpin^{1,5,6}, Erik Engwall², and John A. Lee¹

¹ UCLouvain, Institut de recherche expérimentale et clinique, Molecular Imaging and Radiation Oncology Laboratory, Brussels, Belgium. ² RaySearch Laboratories, Stockholm, Sweden. ³ UCLouvain, Institute of Information and Communication Technologies, Louvain-La-Neuve, Belgium. ⁴ Ion Beam Applications SA, Louvain-La-Neuve, Belgium ⁵ KULeuven, Department of Oncology, Laboratory of experimental radiotherapy, Leuven, Belgium. ⁶ Particle Therapy Interuniversity Center Leuven - PARTICLE, Leuven, Belgium

Summary

Purpose: Proton arc therapy has emerged as a promising approach for faster treatment delivery in comparison to conventional proton treatments. However, unexpected prolonged treatment times in several proton arc planning studies have raised concerns. In this study, we aim to address the challenge of minimizing delivery time through energy layer (EL) filtering combined with the fast EL pre-selection, comparing its performance to a baseline approach without filtering.

Methodology: We developed three methods of EL filtering: unrestricted, switch-up (SU), and switch-up gap (SU gap) filtering. The unrestricted method filters the lowest weighted EL while the SU gap filtering removes in priority the EL around a new SU to minimize the gantry rotation braking. The SU filtering removes the lowest weighted group of EL that includes a SU. These post-processing filters were used in conjunction with the RayStation dynamic proton arc optimization framework (ELSA). Four bilateral oropharyngeal cancer patients' data were used for evaluation. Objective function values, target coverage robustness, organ-at-risk doses and NTCP evaluations, as well as comparisons to IMPT plans, were used to assess plan quality.

Results: The SU gap filtering algorithm performed best in three out of the four cases, maintaining plan quality within tolerance while considerably reducing beam delivery time. It achieved up to approximately 22% reduction in delivery time. The unrestricted filtering algorithm followed closely. In contrast, the SU filtering showed limited improvement, suppressing hardly a few SU without substantial delivery time shortening. Robust target coverage was kept within 1% of variation compared to the ELSA baseline plan while OAR doses slightly decreased in all four patients. Both arc plans present large reductions in NTCP values for dysphagia and xerostomia compared to the reference IMPT plan.

Conclusion: This study provides insights to accelerate proton arc therapy delivery without compromising plan quality. These advancements could enhance treatment efficiency and patient throughput.

7.1 Introduction

Over the past decade, proton arc therapy has raised notable attention in the field of radiation oncology. This treatment technique involves continuous rotation of the gantry while delivering radiation, similar to Volumetric Modulated Arc Therapy (VMAT) in photon therapy [124]. The literature has witnessed a substantial increase in publications related to proton arc therapy between the first one in 1997 [127] and today, with a particular focus on clinical proofs of concept across various disease sites. Additionally, researchers have been developing treatment planning algorithms to effectively utilize the new degrees of freedom brought by the arc delivery.

One of the anticipated advantages of proton arc therapy is the potential for faster treatment delivery compared to conventional proton treatments as was seen with VMAT. Several studies have demonstrated that proton arc therapy can reduce overall treatment time [176, 179, 177, 174]. Importantly, though, reported delivery times are often estimated with rough approximations, relying solely on the total energy layer switching time (ELST), without accounting for the distinction between upward and downward energy switching, likely due to the limitations of earlier delivery systems. The arc timing advantage was actually only evident when ELST fell between 0.5 and 2 seconds, but this assumption for upward energy switching is not realistic in most of the currently available hardware. This makes it challenging to draw definitive conclusions based on these estimates only. Furthermore, contradictory findings have also been reported, where proton arc therapy exhibited prolonged treatment delivery durations compared to intensity modulated proton therapy (IMPT) [173, 183], especially in the discrete arc mode where the delivery entails a large number of fixed-angle fields for which the gantry needs to come to a full stop at each discrete direction to deliver a number of stacked energy layers [157]. In the dynamic mode, the prolonged delivery times stem from a complex combination of several factors, including energy layer switching time, spot irradiation, gantry rotation, a large number of layers and spots, and the requirement for accurate delivery from the right directions. Consequently, the gantry needs to undergo frequent deceleration and acceleration, contributing to the overall prolonged delivery times.

The scientific community has therefore dedicated efforts to devise approaches to minimize the delivery time of proton arc plans. One such approach is to formulate additional objectives to be minimized, which are surrogate measures of the delivery time, while simultaneously optimizing the spot weights within an integrated framework [16, 155, 212, 232]. While the concept is elegant in theory, this approach encounters practical challenges. These include the requirement for sim-

plified mathematical models to estimate beam delivery time, the need for massive computational resources and suitable solvers, as well as the necessity for careful tuning of objective weights. These limitations hinder its practical use and effectiveness in achieving substantial time savings in proton arc therapy planning and delivery. Another approach involves the pre-selection of a user-defined number of energy layers using heuristics based on dosimetry [17, 18] or on geometrical considerations [26, 159], with the intention of controlling the maximum number of upward energy switches. The spot-scanning proton arc (SPArc) therapy algorithm has demonstrated efficacy across various disease sites, primarily due to its dose-based approach. However, its main drawbacks are long optimization times and limited flexibility to select energy switch locations. On the other hand, the generation of dynamic arc plans with early energy layer selection and spot assignment (ELSA [26]) prior to robust spot weight optimization has been noted for its fast execution time. This is due to two main factors: firstly, it reduces the spot weight optimization problem size, and secondly, it reduces the need to compute dose for a vast number of potential energy layers and spots. However, using these rapid methods comes at a cost, as it constrains the optimizer's freedom for the subsequent dose optimization process.

To reap the benefits of the dose-based and geometry-based approaches, we propose a hybrid method. Our hybrid approach extends ELSA that rapidly creates an initial set of energy layers, which are filtered after robust spot weight optimization to a final set using information from the optimized dose distribution. As a result, the filtered plan accelerates delivery. This approach is inspired by iterative filtering step in the SPArc algorithm [17] and the Reverse Greedy algorithm [144] originally proposed for discrete proton arcs. These algorithms progressively eliminate the energy layers contributing the least to the plan quality. We have developed and compared three distinct methods capable of energy layer filtering. These methods act therefore as post-processing tools that filter after the spot weight optimization process. Through the implementation of these filtering techniques, our aim is to enhance the overall efficiency of proton arc therapy delivery.

7.2 Material and methods

7.2.1 Energy layer filtering algorithms

Unlike IMPT plans, where each beam angle is optimized to cover the target in depth using a decreasing energy sequence, proton arc plans typically adopt a sector-based approach. Arc beams are divided into sectors, with energy decreasing within each sector and increasing only between sectors. This design stems from the anticipated increased time required for upward energy switches. Delivery occurs

within fixed intervals (centered around a discrete direction) known as control points across each sector. Although more than one energy layer per discrete direction could be delivered in practice, doing so would slow down gantry rotation and compromise treatment efficiency.

The number of sectors in a proton arc plan, which defines the number of switch-ups (SU) between energy levels, is a carefully studied topic as it directly correlates with the delivery time for most machines [16, 212, 155, 156]. Researchers have sought to develop methods of minimizing the number of SU in these plans. The basic assumption of this conceptual model is that delivery time is proportional to the irradiation time of each energy layer and the number of switch-ups in the plan. Therefore, minimizing the delivery time can be achieved by filtering energy layers and/or reducing the number of SU.

In this study, we developed methods and optimized proton arc plans in a modified research build of the RayStation dynamic proton arc optimizer. While not exactly the same as the one showcased in the original publication [26], our approach still follows the ELSA framework for selecting energy layers based on geometry followed by the classical spot weight optimization.

Considering an existing arc plan as a starting point (Fig. 7.1(a)), we have developed three distinct methods that can filter energy layers (EL) and/or reduce the number of SU as a post-processing step. Assuming N initial energy layers and M initial SU in the plan, we aim at keeping at most X energy layers or Y SU. The three methods are summarized below:

1. **Unrestricted filtering:** This algorithm filters the $N - X$ energy layers with the lowest corresponding monitor unit (MU) from the plan. By removing the lowest MU energy layers, we aim to reduce the overall number of energy layers in the plan while preserving the important dose contributions. In Fig. 7.1, from (a) to (b) individual holes appear where the lowest EL have been suppressed.
2. **SU filtering:** This algorithm filters $M - Y$ SU by deactivating the partial sectors with the lowest corresponding MU. It iterates through two nested loops over the energy layers in the plan. It checks if at least one SU occurs between the two indexed energy layers and verifies that the energy of the second indexed angle is lower than the energy of the first indexed angle. When these conditions are met, the algorithm accumulates the MUs of the energy layers between the two indexes, forming a partial sector. These partial sectors, containing SU, are then ordered based on increasing MU values. The

energy layers included in each partial sector are subsequently removed from the plan. By eliminating the energy layers associated with partial sectors containing SU and having the lowest MUs, the SU filter effectively reduces the overall number of SU in the plan. Figure 7.1(c) displays one large hole between gantry angles $\sim 100^\circ$ and 125° , effectively reducing from 5 to 4 SU in the arc beam.

3. **SU gap filtering:** This algorithm removes the energy layers with the lowest MU, similarly to the unrestricted filtering algorithm. The difference is that this algorithm is restricted to only choose from the energy layers that are located just before or just after a SU. Note that after an energy layer is removed to create a gap, the adjacent energy layers on either side of the gap become the ones positioned just before and just after a SU. The rationale behind this approach is that the gantry is often forced to decelerate to a rather low velocity in conjunction with a SU and increasing the corresponding angular span might mitigate this effect. The algorithm contains a maximum size of this gap, since at a certain point the gantry will be able to keep a constant speed and still perform a SU. We establish this maximum gap size, which is measured in terms of the number of removed layers, as follows:

$$\left\lfloor \frac{t_{\text{SU}} + t_{\text{EL}}^{\text{mean}}}{t_{\text{SD}} + t_{\text{EL}}^{\text{mean}}} \right\rfloor, \quad (7.1)$$

where $\lfloor \cdot \rfloor$ is the floor function, t_{SU} and t_{SD} are the times to switch the energy upward and downward respectively and $t_{\text{EL}}^{\text{mean}}$ is the average irradiation time per energy layer. Which is to say, this is the ratio between the delivery time of a layer with an SU and a layer with a SD. Note that this means that different plans may have different maximum gap sizes.

Each algorithm, depicted in Fig. 7.1, is run multiple times, decreasing monotonically the allotted budget of energy layers (for unrestricted and SU gap filters) or switch-ups (for SU filtering) in the plan, hereinafter denoted as the MaxBound-Sequence s . Each element s_i is the maximum number of EL or SU we want to achieve after the filtering in the run i . Following each filtering operation indexed by i , we re-run the spot weight optimization with a reduced number of iterations to compensate for the loss of monitor units (MU) and to maintain plan quality. Each filtering run is independent from the others and start from the initial optimized spot weights of the optimized arc plan.

This strategy allows us to populate a Pareto front for each algorithm tested. A

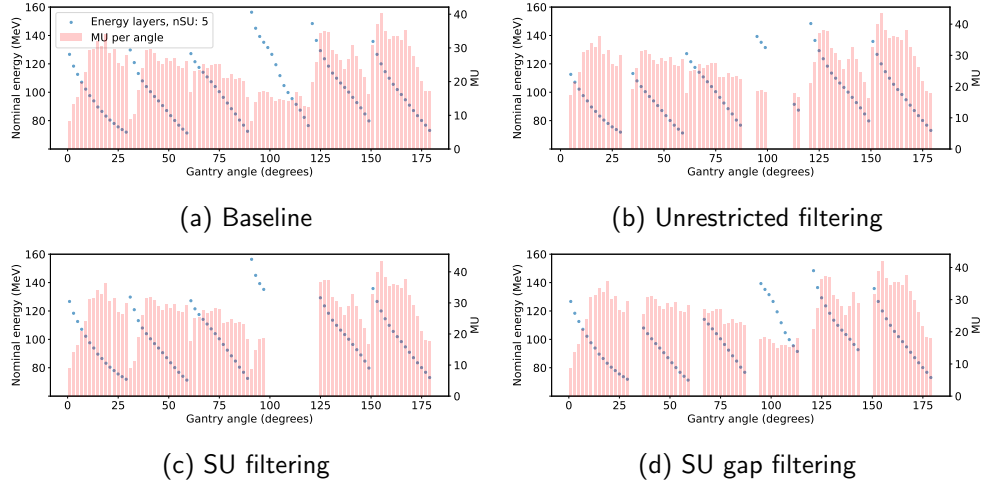


Figure 7.1: Strategies to filter energy layers (EL) for a random generated case. (a) Nominal arc plan used as starting point. (b) Unrestricted filtering removes the 15 lowest MU (EL). (c) SU gap filters the 15 EL that reduce the gantry time to break before each SU. (d) SU filters the lowest MU partial sector that includes at least one SU.

Pareto front represents the set of optimal solutions in a multi-objective optimization problem, where improving one objective comes at the expense of worsening another [21, 233, 234]. In this context, the Pareto front generated relates the attainable dosimetric objective values to the necessary dynamic delivery times. By examining the trade-off between these two sides of the problem, one can select the most suitable trade-off plan from the Pareto front. Algorithm 4 summarizes the workflow of the methodology.

7.2.2 Patient data and treatment planning

To demonstrate the efficiency of the algorithms, we selected four patients with oropharyngeal cancer from an anonymized database obtained from Cliniques Universitaires Saint-Luc in Brussels, Belgium. The local ethics committee approved the retrospective use of this database. The choice of oropharyngeal cancer as the disease site follows from the expectation that proton arc therapy would benefit to those patients, as compared to conventional treatments, with a lower integral dose, reduced doses to OAR, and overall less toxicity [179, 157, 187]. Three out of the four patients were selected from the database based on their higher NTCP in the IMPT plans for the side effects under consideration. Patient 1 showed the highest probability for dysphagia (swallowing difficulty) [235] grade 2, patient 2

Algorithm 4 Filtering algorithm workflow

Input data: Optimized spot weight x^* , N EL, M SU, MaxBoundSequence s , R runs, filtering algorithm $\mathcal{A}$

Require: $\{s_i\}_{i=1}^R : s_i \geq s_{i+1} \forall i \in \mathbb{N}$ ▷ Monotonic decreasing sequence

Ensure: $x = x^*$

```

1: for  $i$  in range( $R$ ) do
2:   if  $\mathcal{A}$  is unrestricted or SU gap then
3:     Filter  $N - s_i$  EL
4:   else if  $\mathcal{A}$  is SU then
5:     Filter  $M - s_i$  SU
6:   end if
7:   for  $i$  in range( $n\_iter$ ) do
8:      $x \leftarrow \text{spot\_weights\_optimizer}(x)$  ▷ Continue optimization
9:   end for
10:  save_spot_weights( $x$ ) ▷ Save plan properties
11:   $x \leftarrow x^*$  ▷ Reset to initial optimized spot weights
12: end for

```

had the highest probability for dysphagia grade 3, and patient 3 had the worst NTCP for xerostomia (dry mouth) [236] grade 2 and 3. The fourth patient was selected randomly from the database.

The patient database comprises IMPT plans manually generated by the same experienced dosimetrist using a treatment planning protocol to ensure consistency (see Table SM1 from [237]). These treatments were planned retrospectively in the treatment planning system (TPS) RayStation 11B (Raysearch Laboratories, AB). With the four selected patients being bilateral cases, the proton plans were designed using a simultaneous integrated boost IMPT technique which treats different target volumes to different dose levels within a single treatment session without increasing the toxicity [238]. For all plans, we use a constant relative biological effectiveness (RBE) factor of 1.1. The prescribed dose was 70 Gy(RBE) to the high-risk clinical target volume (CTV) and 54.25 Gy(RBE) to the low-risk CTV nodal regions, delivered in 35 fractions. The plan was designed with four beams at couch and gantry angles of (10°, 60°), (10°, 120°), (350°, 240°), and (350°, 300°), respectively. A rangeshifter was systematically used for each beam. Monte Carlo dose calculation and robust worst-case optimization [80] were employed for the proton plans, considering 21 scenarios (2.6% proton range error and 4 mm systematic error in all three spatial directions) for the CTV volume, the spinal cord and the brainstem.

The proton arc plans, referred to as ELSA plans according to [26], were designed with a single full arc beam with one-degree spacing between each control point, resulting in 360 initial energy layers available for optimization. It is worth noting that the version of ELSA used here is adapted with a more predefined sector length, which also explains the consistency in the number of SU across the four patients. They were planned for the study in a research version of RayStation 2023B using a total of 250 iterations and spot filtration iteration set at 150. Filtered ELSA plans, employ the same set of objectives and robust settings as the original ELSA plans which were initially obtained from the IMPT plans. However, in order to expedite the planning process, the total number of iterations for optimization is reduced to 50. All plans are normalized to the median dose (D50). The Monte Carlo (MC) dose calculation was carried out until it reached an uncertainty of 0.5%.

Going back to Algorithm 4, given $R = 10$ runs, the 360 initial EL and the 16 initial SU in each ELSA plan, we chose the MaxBoundSequence to be 10 evenly spaced numbers over the interval $[350, 200]$ for the maximum number of allowed energy layers (Method 1 & 3) and the sequence of numbers decremented by 1 over the interval $[15, 5]$ for the maximum number of allowed SU (Method 2). These specific numbers were chosen intentionally to ensure a smooth transition towards the number of energy layers present in the IMPT plans, which is approximately 160 energy layers.

7.2.3 Treatment plan evaluation

Delivery time

Since the objective of this study is to achieve faster delivery rather than improving plan quality, it is noteworthy that suppressing energy layers from the initial set of energy layers is not expected to enhance plan quality. The aim is merely to minimize the delivery time with precautions to ensure that plan quality remains comparable to the ELSA baseline plan.

Dynamic beam delivery time (BDT) was calculated using the ATOM¹ (Arc Trajectory Optimization Method) algorithm, an open-source tool [229] published by RaySearch Laboratories. ATOM is designed to determine a fast plan delivery while adhering to the mechanical constraints specified by the user. The machine delivery assumptions utilized in this study are detailed in Table 7.1. We made these selections based on realistic assumptions rather than referring to a particular proton machine's specifications. For example, it is acknowledged that, at present,

¹<https://github.com/raysearchlabs/ArcTrajectoryOptimizationMethod>

Maximum gantry velocity	5°/s
Maximum gantry acceleration	0.5 °/s ²
Maximum gantry jerk	0.5 °/s ³
Downward energy switching time	0.5 s
Upward energy switching time	5 s
Spot switching time	2 ms
Spot delivery time (per MU)	5 ms
Maximum window size	0.99°

Table 7.1: Machine constraints parameters used for ATOM BDT estimation [229].

switch-up times are typically an order of magnitude higher than switch-down times in most proton centers due to the magnetic hysteresis effect [18]. Regarding the maximum gantry velocity, we opted for a slightly more conservative value, considering that current proton gantries typically rotate at one revolution per minute. The maximum window size specifies the angular safety tolerance around each irradiation direction. The selection of ATOM was driven by its utility, given that no proton arc machine has yet been established and commissioned in clinical settings; ATOM serves as a reliable estimator at present. Nevertheless, in the final evaluation, we also provide BDT estimations obtained with ATOM but using the layer irradiation times and layer switching times based on a simulation of an IBA ProteusPlus machine using the IBA ScanAlgo system.

Dosimetry

In addition to our primary objective of achieving faster delivery, we have included a comprehensive dosimetric evaluation to consider the selected trade-off from the generated Pareto front. Target coverage is evaluated in the IMPT, ELSA, and filtered ELSA plans employing the worst-case scenario robustness evaluation approach. This approach considers 14 different isocenter shifts, comprising 6 points along the main axes and 8 points along the diagonals, which are defined by the setup uncertainty. Additionally, three density shifts were taken into account, resulting in a total of 42 scenarios. The assessment of robust target coverage was based on the criterion of achieving a minimum dose of at least 95% of the prescribed dose at 98% of the target volumes in a worst-case scenario dose distribution ($D_{98\%} \geq 95\% D_p$). This criterion was applied to both the high-risk and prophylactic lymph nodal regions, ensuring sufficient dose coverage for these target areas. Both arc plans (ELSA and filtered ELSA) are compared by evaluating the deviations in target coverage from the values achieved by the IMPT plan in the nominal and worst-case scenarios. We also assess the nominal average doses

received by organs-at-risk in the head-and-neck region using the same comparative approach. Finally, we have incorporated the evaluation of NTCP for xerostomia and dysphagia using the models from the Dutch model-based approach [239]. The Δ NTCP is calculated by taking the difference between the arc plans and the IMPT plan in the nominal scenario for each patient.

7.3 Results

7.3.1 Energy layer filtering algorithms comparison

The plan quality of the three algorithms is compared for each of the four patients, using the objective function value as a global scalar surrogate of the plan quality. Figure 7.2 illustrates the evolution of the objective value with respect to the remaining number of energy layers in the plan, where the maximum number of allowed EL or SU decreases linearly from right to left. The red star represents the optimized plan from which EL or SU are filtered. The filtering process is performed independently for each run, starting from the red star (i.e. the baseline ELSA plan). The multiple runs for each algorithm form a front, with a pink band denoting the tolerance range around the initial objective value that should not be exceeded (above). A tolerance value of 10% is employed in this study.

From the plotted data, it looks like that the unrestricted filtering algorithm performs better compared to others, with a majority of data points falling within the tolerance band compared to the other two methods. The SU gap filtering also shows favorable results, closely following the unrestricted filtering, for patient 1 & 2, with a behavior indicating the importance of preserving energy layers for optimal plan quality. SU filtering struggles more, as it experiences a rapid increase in the objective function value. Notably, SU filtering suppresses many more layers than the other two filters.

Interestingly, the assumption that suppressing more energy layers would lead to a stronger decrease in BDT is proven incorrect in Fig. 7.3, that illustrates the true Pareto front. The SU filtering, despite suppressing more energy layers simultaneously, actually exhibits the smallest reduction in delivery time. On the contrary, the SU gap filtering effectively reduces the beam delivery time with just a small number of suppressed energy layers. For instance, in patient 1, implementing the SU gap filtering and suppressing 44 energy layers in the third run resulted in an important reduction of the BDT from 422 seconds to 330 seconds. In contrast, the unrestricted filter required 77 energy layers to be suppressed in the fifth run to achieve a comparable decrease in BDT (336 seconds). We also observe a decrease of approximately 30 seconds by only filtering 10 energy layers with the SU gap

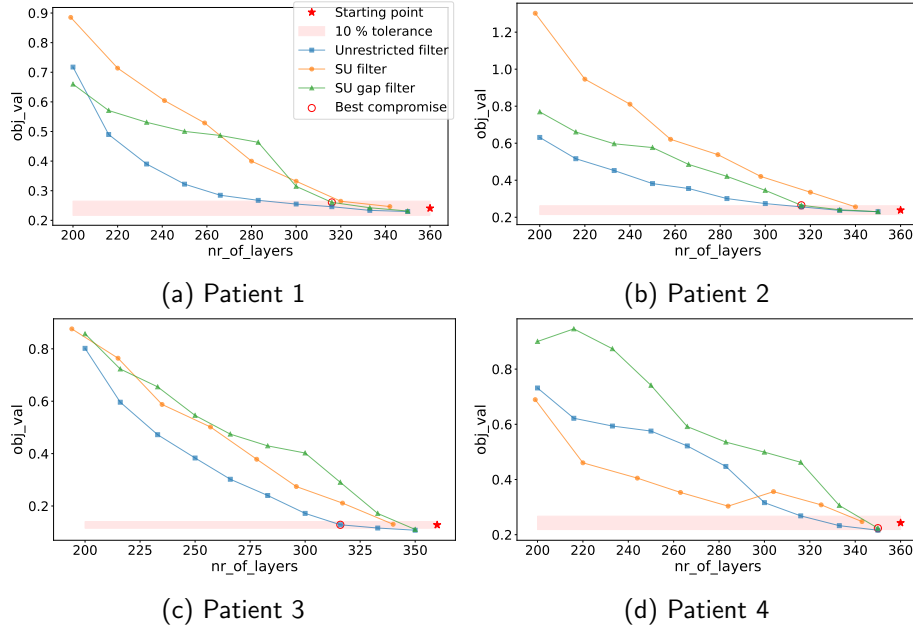


Figure 7.2: Objective function value versus **number of energy layers** for three energy layers filtering methods for the four oropharyngeal patients. The unrestricted method filters the lowest weighted EL while the SU gap filtering removes in priority the EL around a SU. The SU filtering removes the lowest weighted group of EL that includes a SU. The red star represents the baseline plan from which we filter EL. The tolerance band encompasses data points corresponding to plans that deteriorates by maximum 10% the objective value from the baseline plan. Plan quality is deteriorating as more energy layers are filtered from right to left.

method in patient 4.

Upon closer look at the Pareto front data points within the tolerance band in Fig. 7.3, the optimal trade-off between plan quality and delivery time can be determined for each of the four patients. The SU gap filter proves to be the most efficient for patients 1 and 2, achieving an effective reduction of approximately 22% in beam delivery time by suppressing 44 energy layers in the third run, while still maintaining plan quality within the tolerance band. It is closely followed by the simplistic unrestricted filter, which reduces beam delivery time by up to 18% and 15% for patients 1 and 2, respectively. SU gap filtering was also the best compromise found for patient 4 although it could only decrease the BDT by 6%. Conversely, the SU filter demonstrates limited improvement, as it could only

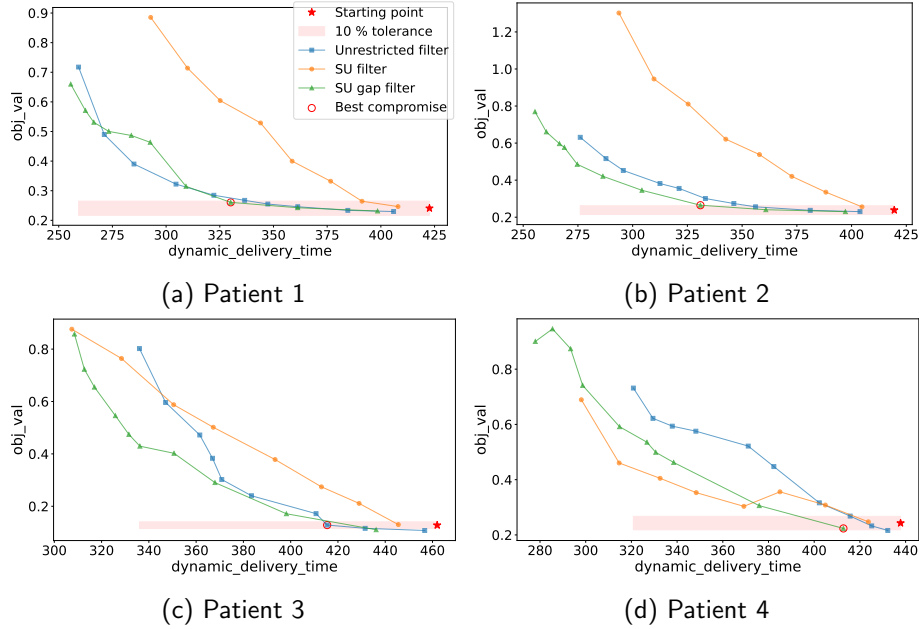


Figure 7.3: Pareto front comparing the objective function value versus **dynamic beam delivery** for three energy layers filtering methods for the four oropharyngeal patients. The unrestricted method filters the lowest weighted EL while the SU gap filtering removes in priority the EL around a SU. The SU filtering removes the lowest weighted group of EL that includes a SU. The red star represents the baseline plan from which we filter EL. The tolerance band encompasses data points corresponding to plans that deteriorates by maximum 10% the objective value from the baseline plan. Best compromise is circled in red showing the shortest delivery time with reasonable objective value.

suppress 1, 2, 1 and 1 SU for patients 1, 2, 3 and 4, respectively, without exceeding the tolerance zone for plan quality. Consequently, the SU filter shows the smallest gain in beam delivery time. Patient 3 presented a bigger overall CTV, compared to the three other cases and raised tougher planning challenges. The filtering for this patient exhibited better results with the unrestricted filtering method. It achieved a 10% reduction in beam delivery time in the third run. This smaller reduction in beam delivery time, as the one observed for patient 4, compared to the two first patients may indicate that more energy layers are required to meet the clinical goals effectively for these particular patients. Nevertheless, SU gap Pareto front in patient 4 looks like we could have had filtered a few more energy layers while staying in the tolerance band. The final best compromises for each patient were circled in red in Fig. 7.2-7.3.

Figure 7.4 provides an illustration of the energy layers retained after determining the best trade-offs for each filtering method, compared to the ELSA baseline plan. Additionally, relevant metrics are presented to assess the timing results. Notably, the SU filtering demonstrates a consistent pattern of energy layers suppression around 270° for all four patients. Furthermore, the removal of the second SU for patient 2 appears to be symmetrically suppressed relative to the first SU. One possible explanation for this behavior is that the optimizer tends to avoid shooting through the shoulders, which may require higher energy levels and deliver higher dose to the body. Another explanation could be related to the position of the primary tumor and also the relative position of the OAR to the tumor. The primary tumor is centrally located for the first three patients while it slightly shifted to the right for patient 4. In contrast, the unrestricted filter sporadically suppresses energy layers, resulting in holes across each sector. Moreover, this filter tends to remove high energy layers. On the other hand, the SU gap filtering displays a cleaner pattern by sequentially removing energy layers around the SU points. This method generates gaps primarily in between sector transitions.

After optimizing filtered proton arc plans (filtered ELSA plans) for each patient and keeping the best compromise found, the next question that arises is whether these plans can be delivered more rapidly than IMPT plans. In the case of the IMPT plans, the delivery time is computed for each beam individually, resulting in what we refer to as static delivery time, as opposed to the dynamic delivery time of proton arc plans. To enable a fair comparison between static and dynamic delivery timings, one should consider the manual interventions required between each beam delivery in the control room and/or treatment room. This includes beam loading in the console, rotation of the gantry and/or the couch with potential re-imaging step. Moreover, given that rangeshifters are used for each IMPT beam, extra caution is needed for the sake of collision. After consulting with two separate proton centers, we introduce therefore a simulated delay of 60 seconds between each beam plus a unique additional time of 120 seconds required for the couch-kick for each of the IMPT plans. Figure 7.5 illustrates this comparison.

Including the simulated delay, all arc plans are faster to be delivered than the IMPT plans when using the simple static approximation from ATOM. The filtered ELSA plans lead to considerable improvement compared to the IMPT delivery time. The biggest improvement is observed for patient 1 and 2 where the filtered ELSA plans speed up the delivery by 25% and 31% respectively compared to IMPT. Regarding ELSA baseline plans, they were only able to speed up the delivery by 5% and 9% for these patients. Patient 3 and 4 show more resistance for the delivery speed-up but filtered ELSA plans still show an improvement of 15% and 6% compared to IMPT. However, the estimation obtained with a realistic model

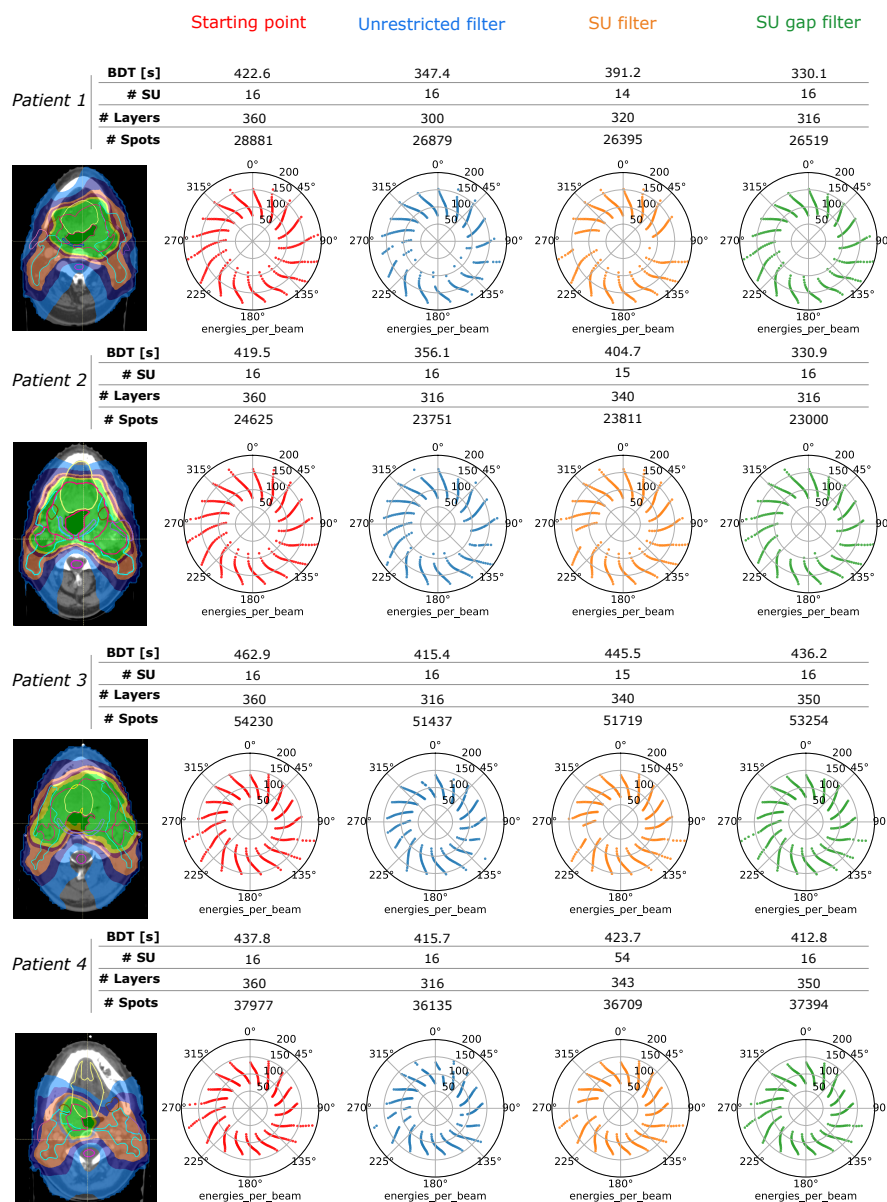


Figure 7.4: Comparison of energy layers kept in the plan after best filtering compromise found for the different algorithms. Rows display each patient whereas columns showcase the filtering algorithm.

from an IBA PPlus system shows the inverse behavior with IMPT plans still faster to deliver compared to arc plans. The filtering could help to reduce the gap

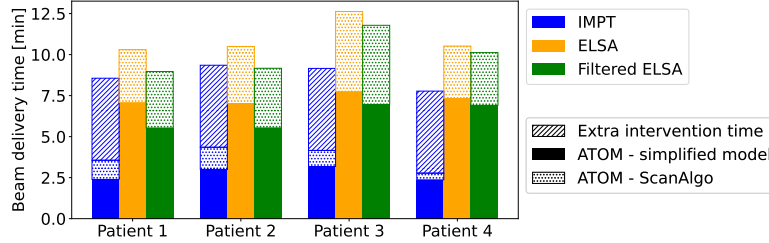


Figure 7.5: Comparison of beam delivery times between IMPT, ELSA and filtered ELSA plans for the best compromises found for each patient. Filtered ELSA employed the SU gap filtering algorithm for patients 1, 2 and 4 whereas unrestricted filtering was applied for patient 3. Plain bars represent a simplified approximation of the static BDT (energy layer irradiation time + switching time) used in ATOM plus (but only for arc plans) the dynamic BDT as computed by the ATOM simulator. Dotted bars represent the static BDT as obtained for an IBA PPlus system with ScanAlgo combined with (but only for arc plans) ATOM dynamic BDT for in-between sector transition times.

with the IMPT BDT but there is still some margin left for further improvement. Using ScanAlgo BDT, filtering the ELSA plan achieved smaller gain in terms of delivery time (compared with the ATOM estimates) with 13%, 13%, 7% and 4%, respectively for patient 1, 2, 3 and 4.

7.3.2 Treatment plan evaluation

The previous results can be complemented with a detailed dosimetric evaluation for each of the best compromises found for each patient and a comparison with the ELSA baseline plan and the IMPT plan.

Figure 7.6a reports the evaluation results for the CTV targets in terms of D98 in the nominal and worst-case scenario for each patient. The D98 is slightly reduced in the filtered ELSA plans compared to the ELSA baseline, although the deviation is minimal, not exceeding 1% for all targets in nominal and worst-case (corresponding to at most 0.6 Gy). These small reductions are therefore considered negligible. Both ELSA plans yield comparable or better results compared to the IMPT plans with most of the data points displayed on the right-hand side of the plot, bringing higher CTV coverage for arc plans. The clinical goals for target coverage (i.e., $D98\% \geq 95\%$ Dp) in both high-dose and low-dose CTVs were met for all the patients in the nominal scenario and the worst-case scenario.

Figure 7.6c illustrates the DVH metrics differences with respect to the IMPT

plan for the OAR included in the clinical goals. Comparing ELSA and its faster delivered version, it is good that we can maintain OAR sparing and similar target coverage while reducing the number of energy layers. In the transition from IMPT to proton arc modality, only patient 3 shows a slight increase in doses to the spinal cord and brain stem but still under the limit. However, average doses in organs generally decrease (85% of all data points lie on the left-side of the plot), which is highly favorable for arc modality in oropharyngeal patients.

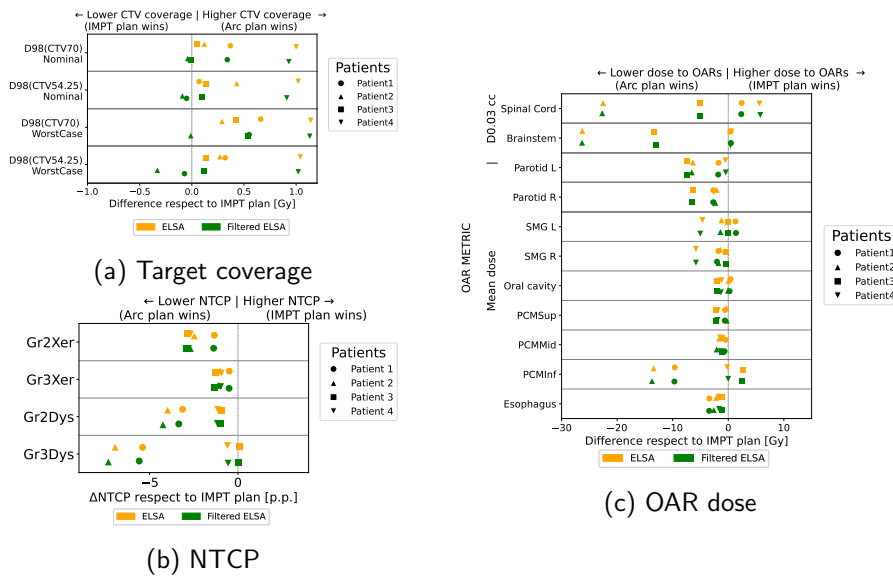


Figure 7.6: Differences in (a) target coverage (D98 - dose received by 98% of the CTV) (b) clinical planning criteria for the OAR (c) NTCP, between arc plans and the IMPT plan. (a) Negative differences imply that the IMPT plan was better than the arc plan in terms of target coverage. (b,c) Negative differences imply that the arc plan was better than the IMPT plan in terms of dose to OAR and probability to develop the side-effect (i.e., less dose to the OAR and lower NTCP). Abbreviations: SMG, submandibular gland; PCM, pharyngeal constrictor muscle. IMPT NTCP values for the ordered four patients: Grade 2 xerostomia = [43.5, 48.7, 53.1, 2.1], grade 3 xerostomia = [12.1, 14.1, 16.1, 11.6], grade 2 dysphagia = [44.2, 43.8, 19.3, 17.2], grade 2 dysphagia = [22.6, 24.4, 3.2, 7.1].

Figure 7.6b displays the difference in NTCP relative to the IMPT plan ($\Delta\text{NTCP} = \text{NTCP}_{\text{arc}} - \text{NTCP}_{\text{IMPT}}$). The arc modality substantially decreases the NTCP for all reported side effects, including grade 2 and 3 xerostomia and dysphagia. This finding further supports the adoption of proton arc therapy for the treatment of this

particular cancer site. Moreover, patients 1 and 2, selected for their high NTCP in the IMPT plan, show a reduction of approximately 5 and 7.5 percentage points (p.p.) in NTCP for dysphagia grade 2 and 3, respectively. Similarly, patient 3, chosen for its elevated xerostomia NTCP value, demonstrates a decrease of around 2.5 p.p. in NTCP for the proton arc plans, further highlighting the advantages of proton arc therapy.

7.4 Discussion

In this study, we presented three post-processing energy layer filters to accelerate the delivery of proton arc treatment plans for oropharyngeal cancers. These filters were designed to suppress the low-weighted energy layers or partial sectors in the arc plans generated with ELSA. Among the three methods, the SU gap filter was proved to be highly efficient in reducing the dynamic delivery time by suppressing a minimal number of energy layers. This primarily stems from the inherent design of ELSA plans, built with equi-spaced sectors and energy layers. Filtering energy layers around sector transition is intuitively efficient as it gives the system some additional time to make the energy switch with minimal gantry braking. At first glance, this method therefore only works for specific proton arc design which limits the angular spacing to make an energy layer switch. In fact, this is mostly the case as proton arc plans are generated with a fixed angle spacing. However, it is not parameter-free as it requires the knowledge of machine-specific parameters (Eq. 7.1) to know maximum how many energy layers it can remove around each sector transition. The unrestricted filter achieved similar results to the SU gap, although it required more energy layers to be suppressed for the same reduction in BDT. This technique is the simplest to set up and was already identified as an efficient mean to filter proton arc plans [17, 144] but rather focused to have exactly one energy layer per angle to enable efficient dynamic delivery. Here we look at filtering technique in the context of plans that already have a single energy layer, and we reduce the total number of energy layers instead of energy layer per angle. Regarding the SU filter, despite removing a large number of energy layers (partial sectors), was the least effective in reducing the BDT. It therefore brings little value. Although we applied these techniques in combination with ELSA for this particular study, they could likely be applied to any proton arc plan, offering particular advantages for plans generated using an energy layer pre-selection method that lacks dose information. This approach makes a good compromise between rapid pre-selection algorithms and comprehensive dose-based optimization [16, 155, 212], as the post-filtering process incorporates dose information.

Pareto plots, showcasing the objective value against the BDT, revealed that

the SU gap filtering algorithm, which suppressed 44 energy layers out of 360 for patients 1 and 2, provided the best trade-off between plan quality and delivery time, accelerating the delivery by ~ 100 s. Further removal of energy layers resulted in a serious increase in the objective value, correlating with a degradation in plan quality. ELSA plans for patient 3 and 4 were harder to accelerate where the best speed-up gain was about 50 and 30 seconds but removing as few as 44 and 10 energy layers respectively. Both patients had much higher initial number of spots, attesting for their more complex tumor geometry. Energy layers could therefore hardly be filtered. Although the choice of a 10% tolerance criterion for the objective values was arbitrary, further investigation might determine the actual threshold passed which plan degradation becomes unacceptable. Moreover, additional iterations may be necessary after the post-processing filter, depending on the number of energy layers removed from the plan. Ultimately, the Pareto graphs could be used for MCO-like optimization and thus help the treating physician to find the best trade-off for each patient.

For the patients treated with IMPT plans, four beams were used with about 40 energy layers each, resulting in a total of 160 energy layers. In comparison, the ELSA baseline proton arc plans needed 360 energy layers, one for each beam direction, to achieve a comparable plan quality. This discrepancy might arise because it is more challenging to combine beamlets from each energy layer in the arc, ensuring an adequately spread-out Bragg Peak to cover the targets. It is also important to understand that both the plan quality and BDT heavily rely on the preliminary energy selection. The angular sampling and sectors count may not be optimal and could potentially benefit from adjustments that are not purely geometry-driven. Yet, incorporating dose-based objectives into the initial energy layers choice can be complex due to the size of the proton arc problem [212]. Hence, dose-based filtering methods paired with rapid energy layer selection become pivotal to address the shortcomings of purely geometry-based energy layer choices.

Our study presents BDT results obtained using the ATOM simulator [229], but it is important to approach these results with caution, as ATOM uses a simplified layer delivery and layer switching model that might not correspond to the exact behavior of most systems. To address this uncertainty, we conducted a benchmark comparison by incorporating BDT data from an IBA PPLus system. While there were some differences in the specific numerical values related to BDT reduction when filtering treatment plans between the two models, we observed that the overall trends remained consistent. The disparity can be attributed to the fact that the IBA ScanAlgo system included more static time in its estimates, a factor directly influenced by the choice of model and parameters. Additionally, the

dynamic BDT component, associated with rotation movement, was still simulated using ATOM. This is because, currently, we are unable to provide a validation of any dynamic BDT model as we still await for the first dynamic arc product integration in clinical settings. Despite these considerations, we do not anticipate a substantial impact on the overall reliability of our comparative analysis if a more refined dynamic model were to become available. Furthermore, it is worth noting that the current body of literature in PAT planning often employs even simpler approximations of BDT with no consideration for acceleration or jerk using in-house models [18]. In this context, ATOM may represent one of the most advanced algorithms for computing dynamic BDT and most importantly available open-source. In any case, BDT will inevitably vary for each system in relation to the delivery assumptions, and yet the conclusions of this study remain valid.

Another key assumption made for the BDT estimation relates to the IMPT plan. We factored in extra time between each static beam delivery to ensure a more fair comparison with the arc plans. After consulting with two separate proton centers about this additional time, we chose the lowest values (i.e. faster IMPT delivery). This suggests that the difference in BDT between IMPT and arc plans could be even more pronounced. Moreover, the filtered arc plans further reduced the BDT, attaining up to a 30% speed-up in comparison to IMPT (using ATOM estimates), emphasizing the utility of the proposed method. Nevertheless, IBA ScanAlgo BDT estimates are more pessimistic, hence the proton arc plans take longer to be delivered compared to ATOM estimates. While filtering narrowed the gap towards IMPT BDT, there is still room for further optimization in reducing the delivery time of proton arc plans.

The plan evaluation and robustness analysis demonstrated that the filtered ELSA plans maintained target coverage with negligible deviations from the ELSA baseline plans in both nominal and worst-case scenarios. Comparisons with the clinical IMPT plans, generated by an experienced dosimetrist, showed that both the non-filtered and filtered ELSA plans outperformed the IMPT plans in most clinical goals, as observed across all four patients. Additionally, the study of NTCP for dysphagia and xerostomia further supported the superiority of proton arc therapy for oropharyngeal cancer. These findings concur with existing publications on proton arc therapy in this disease site [179, 187].

Methods of layer filtering have already been extensively studied in the context of IMPT [240, 241, 242, 243]. However, recent advancements in beam energy layer selection hardware have considerably reduced the impact of suppressing energy layers in IMPT plans. This raises the question of whether this would also apply to proton arc plans. Proton arc plans actually involve more energy jumps

and they must be dynamically delivered. Most machine capabilities face challenges with achieving the same efficiency for upward energy switching as they do for downward energy switching. This is due to the physical limitations of the magnets, known as the magnetic hysteresis effect, which requires adjustments to the magnetic field strength for each energy level to guide protons effectively to the treatment room [149]. In this context, changing energy upwards takes more time, as it is preferable to complete a full cycle, starting from the highest energy level and then descending to the desired level. Vendors are aware of this limitation in dynamic proton arc treatments and are exploring improvements in energy layer selection systems. Potential solutions include the use of achromatic magnets, which eliminate the need for retuning the entire beamline when switching energy levels, thereby achieving equal speed in both directions [244, 245, 246, 247]. Another solution involves dynamic ridge filters that broaden the width of Bragg Peaks and reduce the number of energy layers required for adequate target coverage [248]. Such advancements would simplify proton arc optimization by reducing the burden of dealing with costly SU in the plan, which is currently a major bottleneck in proton arc treatment planning.

Nevertheless, despite the complexity of proton arc plans, the implementation of a post-processing algorithm that filters out the lowest MU-weighted energy layers can still have a positive impact on treatment delivery. Even a small reduction in BDT could potentially enhance patient throughput in clinical settings. Furthermore, the integration of delivery-accelerating strategies could be seamlessly implemented into existing treatment planning systems. It is noteworthy that post-processing filter, including the subsequent short re-optimization, only takes around 10 minutes for the patients evaluated in this study (with 5 minutes dedicated to pre-dose calculation which could be cached). This relatively short running time might fall within an acceptable range of delay for proton treatment planning.

This study selected oropharyngeal cancer patients to showcase the filtering algorithms, as these patients appear to highly benefit in terms of plan quality, even though the delivery time remains long. However, the proposed post-processing filters could potentially be applied to other treatment sites with long treatment times, such as esophageal cancer that has shown very large difference compared to its IMPT plan counterpart in a recent institutional study [183]. Finally, it is important to acknowledge the limitations of this study, including the small sample size of four patients and the absence of statistical analysis. Nevertheless, this study provides a proof of concept and lays the groundwork for future investigations.

7.5 Conclusion

In conclusion, this study introduces energy layer filtering as a post-processing step of the dose optimization to accelerate the delivery of proton arc treatment plans. The results demonstrate that these filters effectively reduce beam delivery times while maintaining plan quality. These methods hold particular promise when combined with fast energy layer pre-selection techniques for designing arc plans that lack dose information for their selection. The findings support the use of proton arc therapy in reducing organ-at-risk toxicity and improving local tumor control in oropharyngeal cancers. Further research is needed to optimize proton arc planning and understand the underlying requirements for energy layer selection. The proposed filters have the potential to enhance clinic efficiency and patient throughput without compromising treatment quality.

Chapter 8

Unveiling complexity

The content presented in this chapter originates from a manuscript titled “The proton arc therapy treatment planning problem is NP-hard”, which has been submitted to the Computers in Biology and Medicine journal and is currently under revision.

Authors: Viktor Wase¹, Sophie Wuyckens², John A. Lee², Michael Saint-Guillain³

¹ *RaySearch Laboratories, Stockholm, Sweden.* ² *UCLouvain, Institut de recherche expérimentale et clinique, Molecular Imaging and Radiation Oncology Laboratory, Brussels, Belgium.* ³ *UCLouvain, ICTEAM, Louvain-La-Neuve, Belgium.*

Summary

Proton arc therapy (PAT) is an advanced radiotherapy technique using charged particles in which the radiation device rotates continuously around the patient while irradiating the tumor. Compared to conventional, fixed-angle beam delivery mode, proton arc therapy has the potential to further improve the quality of cancer treatment by delivering accurate radiation dose to tumors while minimizing damage to surrounding healthy tissues. However, the computational complexity of treatment planning in PAT raises challenges as to its effective implementation. In this paper, we demonstrate that designing a PAT plan through algorithmic methods is a NP-hard problem (in fact, NP-complete), where the problem size is determined by the number of discrete irradiation angles from which the radiation can be delivered. This finding highlights the inherent complexity of PAT treatment planning and emphasizes the need for efficient algorithms and heuristics to address the challenges associated with optimizing the delivery of radiation doses in this context.

8.1 Introduction

Proton arc therapy (PAT) is a novel modality in the field of radiation therapy for cancer treatment. Unlike the conventional intensity modulated proton therapy (IMPT) that delivers irradiation from a few discrete angles, PAT administers treatments dynamically, which means that the irradiation device (the gantry) rotates continuously around the patient. This dynamic method allows for irradiation from dozens to hundreds of discrete directions following a sequence of energy layers organized in sectors, which define the proton beam range in the patient to effectively reach and cover the tumor in depth, as illustrated in Fig. 8.1. The anticipated benefit of such a diverse range of irradiation angles is an increased adaptability to meet the clinical objectives set by radiation oncologists. The continuous movement also simplifies the treatment workflow as a single field is delivered without pauses at discrete intervals along the arc.

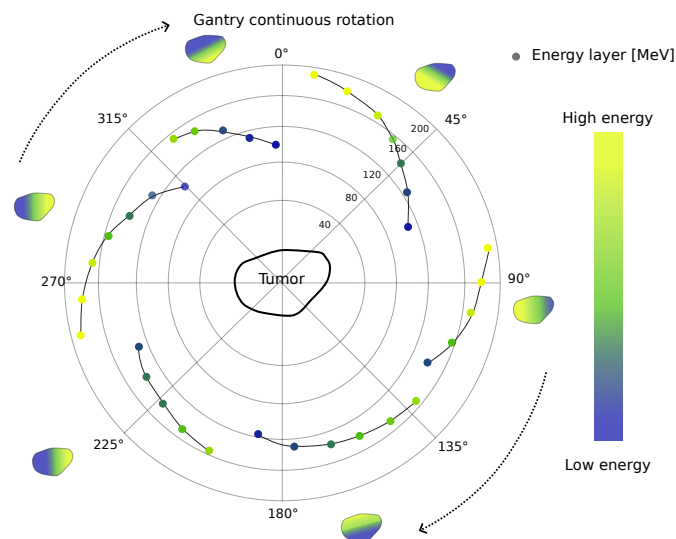


Figure 8.1: Dynamic proton arc delivery concept. An arc beam is subdivided into angular sectors. Each irradiation direction in each sector delivers a unique energy layer corresponding to a tumor layer. Energy decreases within a sector but increases in-between.

However, this increase in degrees of freedom introduces significant complexities into the treatment planning optimization process. While the extended pool of angles potentially offers greater flexibility, it concurrently demands a meticulous balancing act. It goes beyond satisfying treatment goals; it also entails optimizing the total delivery time due to the extensive number of irradiation angles. The quest for optimal proton arc plans has been a hotbed of research, with various strate-

gies under scrutiny. For instance, convex optimization techniques predominantly focus on minimizing a surrogate of the delivery time [16, 155, 156]. Approaches such as simulated annealing with dedicated neighborhood operators have been explored, and global optimization strategies have been applied using mixed-integer programming formulations [249]. Recognizing the sheer scale and intricacy of the problem, some researchers have proposed heuristics, albeit at the cost of reducing the overall flexibility [17, 26, 161].

In light of these challenges, the following question emerges: Is the optimization problem associated with PAT inherently difficult, more formally, is it NP-hard? This question was initially posed by Wuyckens et al. [212], providing complexity results and insights in some particular configurations. So far, no complexity proof for the general PAT problem has been given yet. In this paper, we demonstrate that the proton arc therapy treatment plan optimization problem is NP-hard indeed. We use a reduction from a well know NP-hard problem, the vertex cover problem, hence proving PAT is at least as difficult.

The paper is organized as follows. Section 2 introduces the proton arc therapy treatment planning problem and a formal mathematical description is provided. Section 3 proves that this problem is NP-Hard, whereas Section 4 briefly discusses the implications of this finding and provides some broader perspectives.

8.2 Background

The construction of a proton therapy treatment plan delivered with the most advanced delivery technique, namely, pencil beam scanning (PBS) [39], is commonly formulated as an optimization problem in which a set of spots, which are narrow proton pencil beams, are defined for each specified discrete gantry angle to cover the target volume. These spots within each angle are organized into energy layers, as illustrated in fig. 8.2. The energy layer controls the energy level at which the proton beam irradiates, corresponding to a specific depth within the patient (i.e., the higher the energy, the deeper we can reach) whereas each spot represents a (x,y) location in the tumor targeted by the proton beam. A weight, or beam modulation factor, is assigned to each spot and directly correlates to the amount of dose to be delivered for that plan configuration. Beam modulation can be expressed as $\mathbf{d} = \mathbf{A}\mathbf{x}$ where each entry in vector $\mathbf{d}$ is the dose in a voxel (i.e., a 3D pixel in the planning image), each entry of vector $\mathbf{x}$ represents the weight of a spot, and $\mathbf{A}$ is the linear operator that relates the weights to the total dose in all voxels. Each column of $\mathbf{A}$, thus represents the elementary dose contribution of a spot to all voxels. Given a plan with S spots and V voxels, $\mathbf{A}$ is therefore a matrix of size $V \times S$.

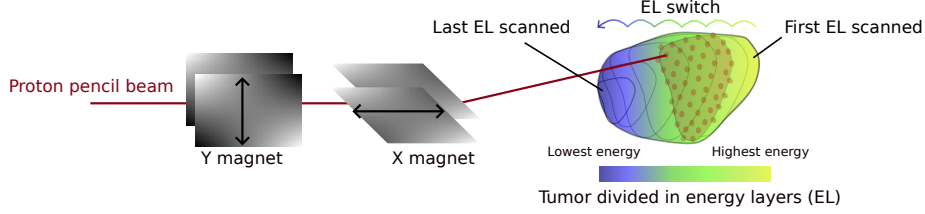


Figure 8.2: Proton pencil beam scanning technique. By means of steering magnets and adequate energy selection system, the proton beam irradiates the tumor spot by spot and layer by layer.

The details of how the design of a plan is modelled as an optimization problem varies a lot depending on the problem specific statement and implementation design choices. We adopt a simplistic, yet realistic, model in this paper. It integrates a minimum dose constraint on the target volume (TV), ensuring each voxel in the TV receives at least the prescribed dose p set by the user. Another common restriction limits the average dose c an organ at risk (OAR) might receive. Mathematically, the classical proton therapy (IMPT) optimization problem is formulated as

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad (8.1)$$

$$\text{s.t. } \mathbf{d} = \mathbf{A}\mathbf{x} \quad (8.2)$$

$$\forall i \in \text{TV}, \quad d_i \geq p \quad (8.3)$$

$$\forall \Omega \in \text{OAR}, \quad \frac{1}{|\Omega|} \sum_{i \in \Omega} d_i \leq c_{\Omega} \quad (8.4)$$

$$\forall s : 1..S, \quad x_s \in \{0\} \cup [\epsilon, +\infty) . \quad (8.5)$$

Here ϵ is the minimal intensity a spot can deliver (machine-specific parameter) and $f(\mathbf{x})$ is a linear function on the spot weights $\mathbf{x}$. For example, if one seeks the combination of spot weights that meets the prescription p on the TV as well as the mean dose constraint on the OAR, while minimizing at the same time the irradiation overhead on normal tissue voxels, then one could define $f(x) = (\mathbf{A}\mathbf{x})^T \mathbf{I}^n$, that is, the total irradiation over all the voxels, with $\mathbf{I}^n$ being the indicator vector such that I_v^n equals one if v is a normal tissue voxel, and zero otherwise. This problem boils down to a linear program, which can be solved in polynomial time.

As already identified in Wuyckens et al. [212], the problem becomes less trivial

when taking time considerations into account. One aspect to consider is constraining the BDT to compete with simpler treatment delivery methods that involve fewer spot candidates. Another important reason for limiting the BDT is to enhance patient comfort and reduce potential time-related treatment uncertainties, such as patient motion. In practice, an optimal arc plan x found for (8.1)-(8.5) may typically take too long to be delivered by the gantry because many energy layers and spots must be delivered. The actual BDT is a complex function influenced by a multitude of parameters. It takes into account irradiation parameters such as the spot scanning time, that is, the time it takes to irradiate a specific spot. It also depends on electro-mechanical parameters, notably the time it takes to switch between spots and energy layers (spot and energy layer switching time) and quantities such as the maximum velocity and acceleration of the gantry to reach the next irradiation window. This multifaceted nature of BDT may result in a highly non-convex objective function, making it difficult to evaluate accurately and hardly tractable for many optimization methods.

While seeking for possible BDT approximation functions, a dominant cause of time consumption in delivery happens to be the energy layer switching time (ELST). In order to mitigate it, the community has shown a growing interest in the energy layer structure of proton arc plans. A common strategy is to impose that each angle of irradiation be restricted to a single energy layer. Additionally, there is a deliberate effort to minimize the number of times the energy of the proton beam increases during the plan delivery. Any change of energy level is slow indeed, but going upwards turns out to be an order of magnitude slower than going downwards. For consistency in our discourse, upward energy switching will be termed “SU” for energy switch-up, thereby matching the nomenclature from [155]. Given the time concerns of SU, proton arc plans often use a sector-based representation for the sequence of irradiation angles. Here, each angular sector follows a decreasing energy pattern, with energy only allowed to increase between sectors. In this setup, each sector acts as an IMPT beam, aiming to cover the target in depth as depicted in Fig. 8.1. Conventional IMPT plans usually have a limited number of beams, typically 2 to 4, and a consistent downward energy sequence at each angle, and thus they do not face the same challenges with ELST, making the planning process more straightforward.

The optimization problem stated in Eq. 8.1-8.5 must therefore be refined in the context of PAT and introduce a new constraint such that the produced PAT plan will be delivered in a reasonable time, by limiting the number of SU to an upper

bound k . This leads to the following generalization:

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad (8.6)$$

$$\text{s.t. } \mathbf{d} = \mathbf{A}\mathbf{x} \quad (8.7)$$

$$\forall i \in \text{TV}, \quad d_i \geq p \quad (8.8)$$

$$\forall \Omega \in \text{OAR}, \quad \frac{1}{|\Omega|} \sum_{i \in \Omega} d_i \leq c_\Omega \quad (8.9)$$

$$\theta(\mathbf{y}) \leq k \quad (8.10)$$

$$\forall s : 1..S, \quad \frac{1}{M} x_s \leq y_s \leq M x_s \quad (8.11)$$

$$\forall s : 1..S, \quad x_s \in \{0\} \cup [\epsilon, +\infty), \quad y_s \in \{0, 1\}, \quad (8.12)$$

where M is a sufficiently large scalar so that $y_s = 1$ if $x_s \geq \epsilon$, zero otherwise (e.g., $M = 1/\epsilon$). Here $\mathbf{y}$ is therefore a binary vector (of redundant variables) that indicates which spots s are activated ($x_s > 0$) or not. Note that if all spots of an energy layer are zero, then that energy layer is not used. $\theta(\mathbf{y})$ is therefore the BDT constraint. Because, most research in the field has been focused on the reduction of SU to approximate and indirectly minimize the BDT, we will define $\theta(\mathbf{y})$ as the function that returns the number of energy SU according to active spots $\mathbf{y}$.

In 2022, Wuyckens et al. showed that an extension of the PAT problem, in which there are a lower limit as well as an upper limit on the number of SU is NP-hard in the general case [212]. In this study we will show that regular PAT plans are also NP-hard.

8.3 Main results

8.3.1 Hardness of PAT

Given the inherent computational challenges associated with PAT treatment planning problems, investigating and qualifying its computational complexity becomes valuable. In this section, we prove Theorem 8.3.1 by reducing the PAT problem to the well-known vertex cover problem (VCP).

Theorem 8.3.1. *The proton arc therapy treatment planning problem is NP-hard.*

Proof.

The Vertex Cover Problem. The VCP is a classic NP-complete problem [250]. In its decision version, and given a graph G , the objective is to determine whether there exists a set of *at most* K_1 vertices such that every edge in the graph is connected to at least one vertex in the set. The optimization version hence simply consists in finding a minimal set of such vertices. Formally, the decision VCP can be defined using binary integer programming as:

$$\sum_i v'_i \leq K_1 \quad (8.13)$$

$$v'_i + v'_j \geq 1 \quad \forall (i, j) \in E \quad (8.14)$$

$$v'_i \in \{0, 1\} \quad \forall i \in V' , \quad (8.15)$$

where V' is the set of vertices in G , E represents the set of edges in G and v'_i binary value indicates whether every vertex is either in the vertex cover or not. We note (V', E, K_1) an instance of the VCP.

Reduction to the PAT Problem. Consider a particular PAT problem with a single target voxel representing the tumor. This voxel is situated within the anatomy in such a way that the sequence of angles requires an increase in energy along the arc. This need stems from the fact that the target location in depth increases with the changing gantry angle position, as illustrated in Fig. 8.3. Each angle in this setup has a single available energy layer, with only one spot.

For each spot s , define a *restricting voxel* as the first OAR voxel v that, as we increase the weight of that spot, reaches its maximum dose constraint c_Ω as per Eq. 8.9. This definition is contingent on the assumption that no other spot contributes to the dose delivered to this voxel v . Referring to Fig. 8.3, the yellow pixels represent an example of restricting voxels on a 2D slice of a patient X-ray image.

Let K_2 denote the upper limit on the number of SU, and p represent the minimum required dose in the target voxel to satisfy the dose prescription (enforcing Eq. 8.8). Ensure that the restricting voxels limit the dose to the target voxel to $\frac{p}{K_2+1}$ per spot. This stipulates that for a set of $K_2 + 1$ spots, each spot must be at its maximum weight to achieve the target dose p . Consequently, this gives us a binary decision problem, since each spot either has zero weight or its maximum weight. Picking an energy layer (and by extension its associated spot) is thus equivalent to make a SU in the sequence.

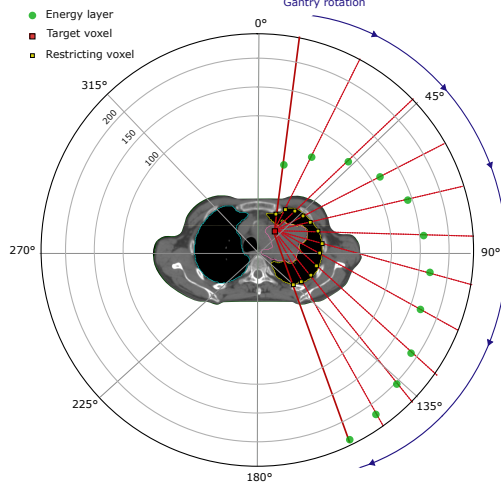


Figure 8.3: The PAT problem instance used in our reduction: One single voxel (red pixel on patient slice) is irradiated by an increasing energy sequence (green dots). A restricting voxel (yellow pixel, belongs to left lung) constraints the maximum spot weight of each spot from each energy layer (EL). Not picking an EL means removing a costly upward energy switching.

Given the set of binary decision variables y_i where $y_i = 1$ indicates the i -th energy layer is active (i.e., picked) and 0 otherwise, we derive the following system:

$$\sum_{i=1}^S y_i = K_2 + 1 \quad (8.16)$$

$$y_i + y_j \leq 1 \quad \forall i, j \text{ satisfying a mean dose constraint,} \quad (8.17)$$

where the first equality results from the fact that exactly $K_2 + 1$ spots are required in order to reach the target dose p . To create dose constraints of the form $y_i + y_j \leq 1$ for pairs of energy layers (for each edge in the VCP, Eq 8.14), we design an OAR only affected by layer i and layer j . This is visually represented by the green regions in Fig. 8.4 that uses the ribs as a potential OAR receiving dose contributions from two unique spots.

The mean dose in this OAR can be written as $\alpha_i y_i + \alpha_j y_j$, with α_i and α_j being positive constants. We assert this dose must be less than or equal to c (see Eq. 8.9). Assuming, with no loss of generality, that $\alpha_i \leq \alpha_j$, we get:

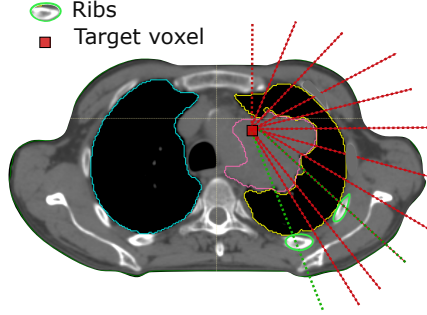


Figure 8.4: Mean dose constraint example: the ribs receive dose contribution only from two spots (highlighted in green)

$$y_i + \frac{\alpha_j}{\alpha_i} y_j \leq \frac{c}{\alpha_i} .$$

Consider the case where $c = \alpha_i$

$$y_i + \frac{\alpha_j}{\alpha_i} y_j \leq 1 .$$

This is true if and only if $y_i + y_j \leq 1$ since $\frac{\alpha_j}{\alpha_i} \leq 1$ and y can only be 0 or 1 which is now equivalent to the constraint formulated in Eq. 8.17.

Following the system established in Eq. 8.16-8.17, an instance of PAT will be denoted (f, S, C, K_2) where S is the number of spots, C is the number of mean dose constraints. We now prove that there exists a transformation $T(V', E, K_1) = (f, S, C, K_2)$, computed in polynomial time, such that

$$\text{VCP}(V', E, K_1) \leftrightarrow \text{PAT}(f, S, C, K_2) ,$$

that is, there exists a vertex cover for (V', E, K_1) *if and only if* there exists a solution to the PAT instance (f, S, C, K_2) , as described in program 8.6-8.12. We set $S = V'$ the number of energy layers (or spots) in our problem. Every vertex in the VCP is associated with an energy layer (or spot) with the subtlety that the *unused* energy layers (or spots) represent the *selected* vertices in the vertex cover, leading to $1 - v'_i = y_i$. Similarly, $C = E$ is the number of mean dose constraints and each mean dose constraint represent an edge in the VCP. Now, let us set $f(\mathbf{x}) = 0$ for the decision problem. The resulting transformations are:

$$y_i + y_j \leq 1 \implies 1 - v'_i + 1 - v'_j \leq 1 \implies v'_i + v'_j \geq 1 . \quad (8.18)$$

Additionally, the relationship between the summations yields

$$\sum_{i=1}^N y_i = K_2 + 1 \implies \sum_{i=1}^N v'_i = N - K_2 - 1 . \quad (8.19)$$

Setting $K_1 = N - K_2 - 1$, we obtain $K_2(K_1) = N - K_1 - 1$.

If an oracle could determine the PAT decision problem, it would also solve the VCP efficiently. The relationship between K_2 SU for PAT and $\sum^N v'_i$ vertex cover in the VCP implies that if the oracle finds a PAT solution with $K_2 = N - K_1 - 1$ SU, a solution to the corresponding VCP exists. If no PAT solution is identified by the oracle, then none addresses the VCP either.

We shall now establish that constructing the PAT instance (S, C, K_2) from the VCP instance (V', E, K_1) can be achieved in polynomial time. Assuming that the complexity of constructing the PAT instance is dominated by the beamlet matrix $\mathbf{A}$ and the OAR voxels set, we construct:

- The beamlet matrix $\mathbf{A}$, which has dimensions $V \times S$ with $V = V' + 1$ (accounting for one restricting voxel per vertex and an additional voxel for the target) and $S = V'$ (representing one spot per vertex). This results in a complexity of $\mathcal{O}(VS) = \mathcal{O}(V'^2)$.
- The OAR voxel set where one OAR corresponds to one edge in the VCP and each OAR contains exactly two elements leading to $\mathcal{O}(2E) = \mathcal{O}(E)$.

Combining these complexities results in a total complexity of $\mathcal{O}(V'^2 + E) = \mathcal{O}(V'^2 + V'^2) = \mathcal{O}(V'^2)$, given that E includes at most $\frac{1}{2}V'(V' - 1)$ edges in the graph. The construction can thus be made in polynomial time. Given that the Vertex Cover Problem is NP-complete, this proves the NP-hardness of the PAT problem. $\square$

We can also prove Corollary 8.3.1.1.

Corollary 8.3.1.1. *The proton arc therapy treatment planning problem is NP-complete.*

Proof. Verifying the validity of a solution $\mathbf{x}$ amounts to calculating the dose

in each voxel of the problem and checking constraints 8.8-8.12, which can be computed in polynomial $\mathcal{O}(VS)$ time, due to the matrix-vector multiplication. $\square$

While the proof presented in section 8.3.1 aimed at demonstrating that PAT is NP-hard when constraining the number of SU, it is possible that any variant where one wants to control the plan structure may be proven NP-hard too. The $\theta(y)$ function can be adapted for this purpose. In order to make NP-hardness proofs, there is no need here to further mathematically define how $\theta(y)$ can be computed, it is sufficient to define what it should return. For example, if one simply wants to constrain the treatment plan not to use more than K_2 spots in total, that is, here we simply define $\theta(y)$ to return the number of active spots. The formulation of our problem with one single spot per energy layer per direction is directly compatible with this new problem statement which is in fact even simpler. This obviously holds if one wants to constrain on the total number of energy layers in the plan.

Note that in our problem setup, the energy level would decrease in the spot sequence if the gantry rotates in the opposite direction. However, our proof remains valid for the case when the gantry is allowed to rotate in either direction. For example, in Fig. 8.3 one could simply extend the plan using mirror symmetry from the right hand side to the left hand side. That would ensure that there is an increase of energy when the gantry is on one side of the patient and a decrease on the other side, no matter which direction the gantry rotates.

8.3.2 Application to heavy ion arc therapy

The extension of arcing techniques to heavy ion therapies, using carbon ions for instance, has garnered significant attention due to their increased biological efficiency for radio-resistant cancer treatment [167, 166]. Consequently, we are led to question whether the NP-hardness proof that was previously demonstrated for proton arc therapy is also applicable to heavy ion arc therapy.

Heavy ions differ from protons in their interaction with matter, inflicting severe DNA damages via complex mechanisms. Multiple models have been proposed to understand and predict the radiobiological effect of heavy ions in matter. The Relative Biological Effectiveness (RBE) is a concept introduced to quantify the biological damage caused by a particular type of ionizing radiation compared to a reference radiation, usually X-rays [36]. The biological effective dose (BED), is thus calculated as the product of the absorbed dose (in Gy) and the RBE:

$$\text{BED} = \text{Absorbed dose (Gy)} \cdot \text{RBE} . \quad (8.20)$$

In proton therapy, an RBE value of 1.1 is commonly adopted, suggesting that protons are approximately 10% more biologically effective than photons per unit of physical dose. For heavy ion therapy, the RBE is significantly higher and fluctuates more noticeably throughout the irradiated volume, necessitating a more detailed RBE model. In heavy ion beam therapy, the Local Effect Model (LEM) is a method for converting the physical dose into an effective biological dose, accounting for the increased damage induced by the densely ionizing tracks of ions [251, 252].

In the context of spot-scanning delivery mode, we can extend our proof by considering the salient biological aspects and demonstrate Theorem 8.3.1.2.

Corollary 8.3.1.2. *The heavy ion arc therapy treatment planning problem is NP-hard.*

Proof. As we revisit the NP-hardness proof for proton arc problem, we realize that we cannot directly state that the restricting voxels for all spots constrain the dose to the target voxel to $\frac{p}{K_2+1}$, due to the variable RBE. We can elucidate this through a two-spot example.

Suppose the minimum target dose in the voxel is p . Let the previously determined RBE from each spot be denoted as $\text{RBE}_{\text{spot}_1}$ and $\text{RBE}_{\text{spot}_2}$. Due to factors such as different beam energies, $\text{RBE}_{\text{spot}_1}$ and $\text{RBE}_{\text{spot}_2}$ may not be equal. For instance, $\text{RBE}_{\text{spot}_1}$ might be 1.2, while $\text{RBE}_{\text{spot}_2}$ is 1.4.

To ensure equal biological dose contributions from both spots, we adjust the physical dose D_{spot_i} from each spot according to:

$$D_{\text{spot}_i} = (\text{RBE}_t / \text{RBE}_{\text{spot}_i}) \cdot \frac{p}{K_2 + 1} , \quad (8.21)$$

where RBE_t is the targeted RBE. This equation means that if the RBE from a certain spot is higher than the targeted RBE, then we need a lower physical dose given in that spot (and vice versa) to ensure that the product of RBE and dose is the same from all spots.

For instance, assuming $\text{RBE}_t = 1.3$, $p = 2$ Gy and $K_2 = 1$, the physical doses at the two spots in the target voxel are:

$$D_{\text{spot}_1} = (1.3/1.2) \cdot \frac{2 \text{ Gy}}{2} = 1.08 \text{ Gy}$$

$$D_{\text{spot}_2} = (1.3/1.4) \cdot \frac{2 \text{ Gy}}{2} = 0.93 \text{ Gy} .$$

We verify the biological dose contributions from the two spots as:

$$\text{RBE}_{\text{spot}_1} \cdot D_{\text{spot}_1} = 1.2 \cdot 1.08 \text{ Gy} = 1.3 \text{ Gy(RBE)}$$

$$\text{RBE}_{\text{spot}_2} \cdot D_{\text{spot}_2} = 1.4 \cdot 0.93 \text{ Gy} = 1.3 \text{ Gy(RBE)} .$$

Although the physical doses and RBE values differ for the two spots, they contribute equally to the total biological dose. This can be generalized to more than two spots, thereby providing a robust translation of the NP-hardness proof from proton to heavy ion arc therapy for spot-scanning delivery.

The remainder of the proof proceeds along the lines of what has been previously established for proton arc therapy. We can therefore claim that heavy ion arc therapy treatment planning is also a NP-hard problem. $\square$

8.4 Concluding remarks

In this study, we have presented a proof to demonstrate that the PAT problem is NP-hard, adding to the growing body of knowledge in the field of computational complexity theory related to radiotherapy planning [253, 254]. Our research not only confirms the intuition that optimizing PAT plans is a challenging task but it also raises crucial questions about the practicality of finding efficient algorithms to address this problem optimally.

The implications of our findings extend beyond theoretical computer science and have significant ramifications for the clinical application of proton arc therapy. As the demand for accurate radiation therapy continues to grow, the ability to generate high-quality treatment plans in a time-efficient manner becomes paramount. Our results suggest that, from a pure mathematical point of view, finding optimal proton arc therapy plans may be infeasible in practice for large, complex cases.

However, it is important to note that the NP-hardness of the proton arc problem should not discourage further research in this area. Instead, it should serve

as a catalyst for the development of heuristic approaches and approximation algorithms that can provide near-optimal solutions within acceptable time frames. Collaborative efforts between computer scientists, physicists, and clinicians will be essential to bridge the gap between the theoretical complexity and the clinical feasibility of proton arc therapy planning.

While this study predominantly focused on the PAT treatment planning problem, other variants may probably be proven as NP-hard too. This holds especially true if they possess certain constraints (or objectives) related to the structure of their solution and can be reduced from the PAT problem. This raises an intriguing query: Is there an existing or potentially novel variant of the problem where this might not hold true? Specifically, can we conceive a “*structurally constrained*” PAT with a non-trivial (i.e., non-constant) function $\theta(y)$, yet remains resistant to a polynomial reduction from the PAT problem? If such an exception does not exist, how might we go about proving this? How can we demonstrate that as long as $\theta(y)$ is not constant, a reduction from PAT is possible?

In conclusion, our study underscores the importance of recognizing the computational complexity of proton arc therapy planning, emphasizing the significant challenge it presents. While NP-hardness implies limitations on potential solutions, it is not insurmountable for the proton arc community. Heuristics such as the ELSA [26] and SPArc [17] algorithms have shown positive results. We can anticipate improved outcomes for cancer patients through the application of proton arc therapy by the development of other such algorithms.

Chapter 9

Conclusion and outlook

Chapter 9 marks the conclusion of this work and offers insights into future directions and perspectives.

Proton arc therapy represents a natural evolution from the fixed-angle IMPT modality in cancer radiation treatment, similar to how rotational IMRT arose from fixed-angle IMRT in photon therapy. However, the technical requirements for enabling this transition in proton therapy are drastically different and demanding at all levels. Hardware upgrades are required to enable the accurate delivery of proton radiation during the continuous rotation of the gigantic gantry, all while ensuring safety. New software challenges also arise as a direct consequence of hardware limitations. Recall the unique property of charged particle beams to stop in matter; energy modulation is required to cover the target volume in-depth. In most PBS-based systems, this is the role of the energy selection system (ESS) at the accelerator's exit and the complex magnet systems from the accelerator to the patient. However the energy change is not instantaneous and ranges between a fraction of a second and several seconds, depending on the system and energy change direction. This technical limitation becomes problematic when delivering a proton arc plan with hundreds of beam directions, each with numerous energy layer candidates. Consequently, traditional algorithms used for spot weight optimization in proton therapy cannot be directly applied to PAT treatment planning. They must be adapted or entirely re-imagined to account for the new degrees of freedom associated with PAT and make a thoughtful selection among them to deliver the plan efficiently and maximize potential clinical benefits.

The primary objective of this thesis was to better understand the optimization challenges inherent to PAT treatment planning and investigate potential solutions. Our aim was not to rigorously demonstrate the clinical value of PAT for various treatment sites. Instead, our focus was on studying the mathematical foundations of the problem and highlighting the key attributes that define an effective PAT optimizer.

The contributions can be summarized as follows:

- In January 2020, when starting this thesis, there were very limited optimization algorithms available for PAT planning. Actually, there were only three solutions to address specifically the ELSP and reduce the total number of EL in PAT plans: the mono-energetic and bi-energetic PMAT [153] and the SPArc [17] algorithms both published in 2016, and Battinelli's [144] 2019 masters thesis. Following these breakthroughs, more researchers began to explore the quest for an optimal PAT treatment planning solution, including our team. Our research journey was prompted by an important paper by Gu et al. [16], where they introduced a novel approach that implements a combined optimization framework. This framework simultaneously optimizes energy layer sparsity and sequencing alongside traditional spot

weight optimization. We embarked on this road with the aim of gaining a deeper understanding of the problem, enhancing the existing methods, and developing our own novel solvers. This led to the consequent Chapter 3 of this thesis. For this study, we implemented and compared three in-house PAT solvers that aim at directly solving the ELSP combined with the dose optimization problem. Although the comparison is mostly qualitative with no clinical implications, we believe the study contributed to the overall understanding of the PAT treatment planning problem. The PAT algorithms development was carried out within the context of the early stage of OpenTPS [19], which has since evolved into an advanced open-source research TPS for proton therapy.

- In the integrated framework proposed for solving both the ELSP and dose optimization in PAT planning, the notion of trade-offs naturally arises. However, in its current form, the planner is required to manually adjust the weights of each objective term in the total objective function to reach the desired balance. Alternatively, it is possible to incorporate a combination of hard and soft constraints, provided that the solver supports such problem statement. For instance, one could constrain the ELST to a maximum value. However, this approach is not straightforward, as there is uncertainty about whether the constraint might become overly restrictive and hinder the spot weight optimizer's ability to meet the dose objectives. Trade-offs in treatment planning have been a recurring topic in radiotherapy, given the inherent conflicts between the goals of effectively treating the target while sparing healthy tissues [116, 208, 255]. Chapter 4 of this thesis employed an existing method [25] to approximate a Pareto front for multi-objective optimization. We applied the method to evaluate the trade-off between the overall plan quality objective and the computed ELST in the PAT plans for several treatment sites. Although this presented major computational challenges, this study marked the first attempt to address the PAT problem as a bi-criteria optimization problem.
- During this thesis, tackling the computational demands of PAT plan optimization became a serious hurdle. While the Python programming language in the optimization simulations played a role, it was not the sole issue. Robust optimization was therefore not even considered because of a lack of computing efficiency. What added to this challenge was the fact that even industry experts moved away from combined optimization methods utilizing the whole set of degrees of freedom, favoring heuristic methods for PAT planning. Seeing this trend, we followed a similar approach. In chapter 5, we developed a method focused on pre-selecting EL, prior to the regular spot weight optimization, based on geometrical target coverage and a direct

estimate of the dynamic BDT. Our approach is unique in that it directly models BDT within the objective function and takes into account the need to spare the OAR. Other EL pre-selection techniques developed so far primarily focused on reducing energy SU [153, 26] or minimizing the total number of EL [159], thus minimizing indirectly the BDT. With this approach, we successfully generated robust PAT plans that met clinical objectives. We compared these plans with RaySearch's dynamic PAT optimizer prototype (ELSA) [26] and a standard IMPT plan. Our method achieved similar or slightly better plan quality in terms of OAR sparing, all while substantially reducing BDT compared to ELSA. The method builds upon existing EL pre-selection methods, offering distinct strengths due to its direct modeling of the BDT in the objective function and ability to spare OAR.

- As our research on combined optimization methods for PAT planning (i.e. dose-based EL selection) gained attention, we began collaborating more closely with industry experts, specifically RaySearch Laboratories. They had developed the aforementioned ELSA pre-selection technique, which was showcased for prostate cancer treatments and achieved very fast computation times [26]. However, ELSA, as all EL pre-selection methods (including ours presented in Chapter 5), has potential pitfalls. It pre-selects EL without considering dose information, sometimes resulting in suboptimal PAT plans failing to fulfill clinical goals. To make better use of PAT DOF, we brainstormed ways to combine ELSA's speed with dose-based EL selection effectively. In chapter 6, we attempted to expand the set of initial EL and introduce functions to penalize less useful EL during the dose optimization process. This approach involved modifying RayStation's source code in a research build to include a "cardinality" operator, which approximates the active number of EL for each discrete direction or in total. By minimizing such function along with the usual dose objectives, low MU weighted EL were penalized and filtered out at user-defined intervals. Unfortunately, the results were unsatisfactory, leading us to revisit the concept of EL filtering. However, this time, we approached it as a post-processing step, which is the focus of Chapter 7. The objective shifted from enhancing plan quality to improving the delivery efficiency of ELSA plans, leveraging the dose information available post-optimization. This work represents an important milestone in reconciling dose-based EL selection with geometry-based EL pre-selection, specifically aimed at accelerating delivery and optimization, and this in a commercial TPS software.
- The final contribution of this thesis is a pure mathematical proof that has important implications for research in PAT treatment planning. We have formally demonstrated that the decision problem associated with PAT plan-

ning, which involves determining if a plan can meet clinical goals while using a sequence of EL with fewer than x energy SU, is NP-hard. This classification places it within a category of computationally challenging problems, where there is no known algorithm capable of solving all instances of the problem in polynomial time. This implies that finding an optimal solution can require exponential time with increasing instance size. While this result was previously suspected in Chapter 3, we have now provided a comprehensive mathematical proof that identifies the SU constraint as the source of NP-hardness in the PAT problem. The proof was also extended for heavy ion arc treatment planning. As a result of these proofs, it becomes clear that using global optimization algorithms for this problem, which initially appeared well-suited due to the discrete modeling of EL selection, is not a practical approach. Instead, it directs researchers towards the development of approximation methods, heuristics, and problem-specific techniques to find high-quality solutions within a reasonable time frame. In practice, the community has already adopted such approaches, and now we have a formal basis for this choice, aligning theory with practical necessity.

Some perspectives for the work accomplished in this thesis will be discussed in the next sections by delving into several aspects.

The ideal PAT optimizer

In the context of PAT treatment planning, we need to strike a careful balance among three key aspects: plan quality, optimization time, and delivery time. This takeaway is essential for anyone looking to address the PAT planning problem effectively. Our study demonstrates that combined optimization of a surrogate for the BDT alongside spot weights can theoretically yield high-quality plans with reduced numbers of EL and SU. However, the computational demands for dose computation and optimization with all possible discrete angles and EL candidates are currently impractical. These demands often result in optimization runs taking more than an hour for a single non-robust run [16, 155]. So, does this imply that we should abandon these techniques and exclusively rely on EL pre-selection methods, which are by design more computationally efficient? This remains an open question. In my view, the full optimization framework should not be dismissed before thorough evaluation within high-performance computational environments, such as commercial TPS. However, this is not a straightforward task, as new objective functions and solvers would likely need implementation to handle EL sparsity and sequencing. During my three-month visit at RaySearch, as mentioned earlier, we made an attempt in this direction, which was summarized in Chapter 6. We formulated a new objective function that the convex solver could accommodate

to penalize and filter out the ‘worse’ EL during the spot weight optimization. The penalizing and filtering processes were applied to a plan that differed from a baseline plan by employing a half gantry angle spacing or allowing an additional arc rotation in the opposite direction after the first arc beam was delivered. This essentially doubled the pool of candidate EL, and we filtered out half of them to match the number in the baseline plan at the end of the optimization. We tested this method on oropharyngeal cancer patients whose baseline plans utilized a full 360° arc rotation with 360 EL. Surprisingly, we observed that the dose-based EL selection method had minimal impact on plan quality. Upon further investigation, we realized that the baseline plans might have already been of sufficiently high quality, rendering dose-based EL selection less influential. This has been further supported when we compared the baseline plans (i.e. dynamic arc) to discrete arc plans, which may be considered a gold standard in terms of plan quality for PAT planning. The baseline plans quality was already very close to the discrete plan one, suggesting that we may have already reached the optimal level of quality. In my opinion, this method may require further research for treatment sites where the baseline plan does not meet the desired quality. Nonetheless, the method, as it stands, substantially increases the BDT due to the finer gantry spacing or additional arc revolution. The potential benefits must, therefore, be carefully weighted against the increased delivery time. In this context, EL post-filtering strategies, as described in Chapter 7, may be found extremely useful in mitigating this increase. Moreover, when incorporating new objectives into an objective function, typically expressed as a linear weighted sum in the field, the challenge of manually tuning the weight of each objective adds complexity to PAT planning, which can lead to a tedious trial-and-error until finding the right balance of objective. Developing methods to solve the PAT problem is thus difficult because we aim to maintain user-friendly simplicity while providing the necessary flexibility for users to make informed choices. This complexity prompts the question: How can we navigate the trade-offs between plan quality and BDT effectively? This is where MCO interactive techniques, commonly known as Pareto surface navigation [208], can make a difference. MCO is not new to the radiotherapy community and dedicated features have already been implemented in some clinical TPS to study trade-offs between dose objectives among the target and nearby OAR. However, its usage seems to remain relatively underutilized and demonstrated only in a few study cases [256, 23, 257, 258] mostly centered on IMRT treatments. One of the possible explanations is that radiation oncologists, who are best equipped to make such decisions on trade-offs, typically get involved only during the final step of treatment plan approval and are not actively engaged in the plan optimization workflow due to their already heavy workload [259].

Now, let us consider the EL pre-selection methods. They effectively address

two major concerns in PAT planning: optimization time and delivery time. These methods are usually exceptionally fast because they do not rely on computationally intensive dose matrix multiplication to select the EL and restrict the subsequent spot weight optimization to the selected set of EL. Instead, they use simple objectives like geometrical coverage, which can be computed efficiently, to carry out the EL selection. Since they are heuristics, they can directly minimize a closer estimation of the full dynamic BDT, as demonstrated in Chapter 5, or surrogate of this time can be constrained through the plan's design. However, the key challenge that remains is ensuring high plan quality, which is probably the largest concern. In several cases, EL pre-selection methods could produce high-quality PAT plans that meet clinical objectives [26, 159]. But is this sufficient justification for choosing PAT over conventional IMPT? To address this question, it is essential to subject EL pre-selection methods to rigorous testing and assess their performance across larger patient cohorts for each disease site under evaluation. Currently, with limited available data, I believe we cannot definitively state whether dynamic PAT plans generated with EL pre-selection offer a clinical advantage over conventional planning with a small number of fixed-angle IMPT beams. This is an area that requires further investigation.

However, let us take a step back and draw a parallel with photon treatments. Tomotherapy (i.e. "slice therapy"), for instance, was developed in the late 1990s and received FDA clearance in 2002 [260], preceding VMAT, which became more prevalent in the late 2000s [124]. Tomotherapy combines technologies from IMRT and CT scanning to deliver highly conformal radiation treatments using a modulated fan beam and a moving couch, while VMAT involves modulated gantry rotation speed and dose rate. The introduction of VMAT was primarily driven by its compatibility with current linac machines and its faster treatment delivery compared to conventional IMRT and tomotherapy. It was not necessarily due to dosimetric advantages, as tomotherapy already offered highly conformal photon therapy but for a higher overall cost [261]. The conformity achieved with VMAT treatments compared to fixed-angle IMRT is typically equivalent, but thanks to a larger set of degrees of freedom, VMAT often improves OAR sparing [136, 137, 138, 139, 140, 141, 142]. With this perspective in mind, we could view PAT as a modality designed to expedite the treatment delivery workflow while maintaining high treatment quality, much like VMAT did for photon therapy. If the additional DOF allow us to further improve plan quality, that may be considered as a bonus.

Now, returning to the EL pre-selection method developed in the context of this thesis; this work was conducted concurrently with the filtering methods discussed in Chapter 7 and has only recently been completed. I realized that these two

approaches could be complementary. We observed that the ELSA EL pre-selection algorithm [26] studied here is highly dependent on the initial plan's design. It might be feasible to combine the concept of generating an ELSA plan with much finer gantry spacing (resulting in an excessive number of EL) and then filtering the energy layers using the SU gap method as well as eliminating irradiation directions that intersect selected OAR, enhancing their sparing and reducing delivery time simultaneously.

Another promising avenue for PAT optimization is the development of beamlet-free dose optimization, a technique that combines fast MC dose simulation with optimization [262]. In this approach, successive proton batches of randomly chosen spots are used to both accumulate the dose and optimize the spot weights through a stochastic coordinate descent algorithm [263]. By doing so, it bypasses the computational burden of pre-computing the large beamlet matrix and the heavy matrix multiplications required for calculating the total dose at each optimization iteration. As a result, it reduces dependence on the initial number of spots in the treatment plan. The actual extent of its benefit for IMPT treatment planning, particularly in terms of computational efficiency, is yet to be comprehensively demonstrated. It hinges on multiple factors, including the capabilities of the computing hardware and the specific characteristics of individual patient cases. Nevertheless, the beamlet-free approach could hold valuable promise, particularly for PAT plan optimization. This is because when considering all possible EL candidates in a PAT plan, the number of beamlets highly increases. In a small study conducted by our team [249], we demonstrated that the beamlet-free approach could enhance the computational efficiency of SPARC [17], addressing one of the major concerns associated with this method – the need to recompute the beamlet matrix after each splitting operation. Other PAT treatment planning approaches might also benefit from the advantages of the beamlet-free approach. However, the challenge of BDT modeling remains. Given that the beamlet-free approach relies on a gradient-based method and finite differences gradient approximation, it is unlikely to perform well with non-convex BDT models. To address this challenge, a method that tracks the activation status of EL and evaluates the spot contribution to a convex BDT approximation may be necessary. Thus, the choice between combined optimization and EL pre-selection keeps being an important decision in PAT treatment planning using the beamlet-free method.

An engineering or software challenge ?

While this thesis has predominantly focused on addressing the software-related challenges associated with ELST in PAT planning, it is legitimate to wonder whether efforts should be redirected towards resolving these challenges within accelerator engineering itself. Given the close collaboration with IBA, we know

they are well aware of the current software challenge that ELST imposes for arc planning and are willing to speed it up, especially when it comes to energy SU. However, the technical implications to improve ELST are becoming increasingly challenging. Accelerating the ELST largely depends on the specific beamline configuration and the type of magnets used. In gantry-based proton facilities combined with a cyclotron technology, the ESS is often installed upstream just after the extraction of the beam from the accelerator and all bending and focusing magnets are downstream [264, 265]. As a consequence, at each energy change, the current in each magnet along the beamline must be adjusted to ensure proper beam transport to the isocenter. This leads to a lag in the energy layer switching process which is more critical for SU than SD as completing a full magnet cycle, starting from the highest energy value and then descending to the desired value, is preferable for SU.

Research has been exploring achromatic magnets capable of a large momentum acceptance [264, 247]. The momentum acceptance refers to the range of energy variation that can occur without the need to retune the magnetic field in the magnets from the entire beamline. With an achromatic magnet, no matter what energy the beam has, it will be properly guided, eliminating the time issue associated with ELST [264, 247, 245, 244]. These sophisticated magnets come however with a bigger size, price and complexity to set up, hence the initial reluctance of vendors. Bi-directional energy modulation is another recent study that could merit investigation for PAT delivery [266], as it takes advantage of both magnet ramping directions to deliver an ascending followed by descending energy sequence, increasing the delivery efficiency. Additionally, Maradia et al. have demonstrated the effectiveness of dynamic ridge filters in broadening the Bragg Peaks, substantially reducing the number of required EL while maintaining optimal target conformity and sharp distal fall-off [248]. In another study, the same group also focuses on enhancing beamline transmission efficiency using a wedge instead of slits in the ESS [267]. This enables faster spot delivery with a higher dose rate and reducing overall treatment time. Although primarily applied to field-by-field IMPT delivery, these efforts hold great relevance for PAT, where the goal is to minimize EL and spots, ultimately reducing BDT while achieving high treatment quality.

Of course, several treatment machines and models have different ELST limitations [268]. In synchrotron-based facilities, where the beam energy is directly set at the accelerator, ELST is typically even longer compared to cyclotrons. This is due to the irradiation being done per spill or cycle [269], which means that after one EL is delivered, the remaining protons within the synchrotron ring must decelerate and be discarded before the next EL can be delivered. Conventional

synchrotrons with single-energy extraction are therefore not well-suited for arc treatments at first glance. However, multi-energy extraction (MEE) techniques have or are being developed for proton [270] or heavy ion therapy clinical use [271, 272]. MEE techniques allow more efficient spill change for each EL delivery, reaching down duration of hundreds of milliseconds by successively decelerating particles to the next planned energy. Energy degraders can also be installed downstream, before the patient in the beam nozzle [132], allowing for ultra-fast energy switching. However, particularly for protons, this can substantially broaden the pencil beam size, potentially compromising the lateral dose fall-off and penumbra [247]. This approach is therefore mostly considered for heavy ion arc systems [273], which are sharper and have less scattering. Nonetheless, there is hope that manufacturers will be able to further reduce the ELST in both directions (up and down), but this won't happen overnight and still has physics-related limitations. If they were to achieve durations of hundreds of milliseconds to align the SU time with the SD time, it would likely simplify the dynamic PAT optimization problem, essentially converting it into a discrete optimization one. In this scenario, the primary focus would be on reducing the total number of EL and spots in the plan without causing a severe degradation in plan quality. These alternative methods for EL filtering or sparsity have already been introduced in the field, offering a more straightforward approach to PAT optimization [144, 159]. In turn, it would probably render a portion of the research conducted in this thesis obsolete. This underscores the dynamic nature of science, which continually evolves alongside advancements in technology.

While much of our previous discussion have centered around the engineering challenges of gantry-based systems, the emerging interest in gantry-less systems cannot be overlooked. These systems use a fixed horizontal beam delivery setup, eliminating the complexities associated with moving a hefty gantry and its associated hardware. This naturally raises the question: How can one deliver PAT with a fixed beamline? Presently, two primary solutions are on the horizon. The first, still largely in the conceptual phase, involves modulating the field strengths of compact magnets using concentric coils. This allows for rapid adjustments in the incident angle of proton beams, effectively removing the need for gantry and patient rotational systems [274]. The second approach suggests upright patient positioning systems, a topic that has been thoroughly examined in recent literature [275, 89, 276]. As the name suggests, these systems position patients in a reproducible seated, semi-seated or even standing posture, complemented by dedicated image guidance systems. Combining the upright positioning with particle therapy is at the center of the table today as rotating the patient in a treatment chair is expected to be technologically simpler and more cost-effective than rotating a massive gantry [182]. This would obviously create a particular

synergistic compatibility for PAT treatments, in particular for facilities that do not have gantry systems. Additionally, commercial options for vertical imaging systems become available. Companies such as P-Cure (Shilat Industrial Zone, Israel) or LEO Cancer Care (Smallfield, Horley, Surrey, UK) now look at how designing a new radiotherapy therapy center with an upright positioning system or even retrofitting the upright positioner system to an existing facility. While the gantry-less approach does not address the ELST challenge for PAT, it offers solutions to other issues. It could make conventional PT and PAT treatments accessible to more patients, as it eliminates the need for expensive gantry systems, ultimately reducing the overall cost of PT treatments and systems maintenance. This is particularly relevant for heavy ion therapy [277], where gantries are scarce due to their cost, size, and advanced technology. However, the transition to upright positioning systems comes with its own set of challenges, both clinical and engineering-related as highlighted by Volz et al. [89]. Ensuring accurate and stable patient positioning is important, and there is still limited data available comparing patient anatomy in supine versus upright positions. Upright treatments may therefore not be suitable for all patients, potentially excluding those who cannot be accommodated by chair-based solutions. Nevertheless, if these challenges can be effectively addressed, upright positioning systems may potentially announce a paradigm shift in particle therapy.

A clinical perspective

While the primary focus of this thesis was not on demonstrating the clinical value of PAT, it is undoubtedly a fundamental question that follows the optimization of treatment plans. In this section, I will present some key clinical aspects of PAT.

First let us recall the primary motivation behind PAT: the DOF. The higher the number of DOF, the better a treatment can be tailored to meet clinical objectives. In particular, PAT may allow for higher plan quality by improving the dose distribution and pushing further the conformity, hence reducing the integral dose. The ability to irradiate from multiple angles may also help either avoid OAR entirely or better spare them by diluting high doses. Additionally, PAT may have the potential to enhance plan robustness by averaging out uncertainties across the many beam angles. Another key advantage is the potential simplification of treatment workflows, as a single arc beam with a single isocenter could be equivalent to treating with multiple IMPT fields using one or more isocenters.

So now, have we seen these advantages in silico treatment planning studies?

For dynamic PBS-PAT, there have been limited clinical results with robust plan-

ning, with the Beaumont group being the primary contributor to this research. They evaluated their PAT algorithm, SPArc [17, 18], across various treatment sites and observed promising outcomes, including consistent reductions in OAR doses and integral dose as predicted. According to their results, PAT also tends to enhance plan robustness, particularly for OAR. However, PAT typically does not result in large improvements in target CI, HI, or target volume robustness compared to IMPT plans, as detailed in Section 2.4.5. In fact, there are studies (other than SPArc-based) where target robustness was compromised [26], especially when evaluating fraction-wise target coverage with repeated CT images [278, 279, 186]. While robustness in PAT was initially considered as a potential advantage, achieving it seems challenging in practice. This may stem from interlaced spot maps from many directions, potentially involving more tangential incidence angles compared to IMPT plans with a few beams and stacked energy layers as hypothesized by Engwall et al. [158]. Consequently, robust optimization should be an essential component of any PAT optimization approach. PAT may also necessitate the use of adaptive workflow and strategies to enhance fraction-wise target coverage directly at the plan design step or during the plan optimization. Nevertheless, these findings should be taken with caution, given that PAT optimizers are still under development and may not harness sufficiently the degree of freedom provided by PAT.

It is important to underscore that these studies have involved relatively small patient cohorts, making it challenging to draw definitive conclusions. To determine the necessary dataset size for achieving statistical significance, a power analysis should be systematically performed [280], which depends on the statistical tests employed. Furthermore, comparing treatment modalities based on treatment planning alone is a complex task due to the multitude of factors involved, including patient-specific variations, biological effects, and the degeneracy of the planning process, which allows to generate the same dose distribution using various field arrangements and intensity maps [281]. While dose metrics bring valuable insights into the potential benefits of different techniques, they are not sufficient. Mazal et al. [182] insisted that future studies should systematically address clinical endpoints through models like TCP and NTCP when evaluating the merits of PAT. Clinical trials will also be desirable for confirming the impact of these planning-based findings.

While dynamic PAT optimization would benefit from more systematic studies involving multiple and larger patient cohorts but also independent studies from different research groups, it feels like there is some hesitation to move forward among some people in the field. This can be due to the recognized drawbacks of existing methods, such as computational demands and limited flexibility. As an

alternative, assessing the clinical value of PAT through discrete arcs or multi-field IMPT, which present less challenges to optimize, could provide a benchmark and a clearer understanding of PAT potential in terms of clinical benefits. Although discrete arc plans result in longer delivery times compared to dynamic approaches, they could serve as a transitional phase toward the adoption of dynamic arc therapies. This is especially relevant as discrete arcs could already be delivered at existing proton centers equipped with gantry rooms. Additionally, not all research or clinical facilities have access to or have developed their own dynamic arc optimization tools, making discrete arc a more accessible and quicker way to gather data. Discrete arcs also eliminate the dosimetric uncertainties from the continuous rotational delivery. Several studies from different research groups have already started to explore discrete arcs [158, 159] or multi-field IMPT [184, 186] and demonstrated impressive results, although the patient cohorts remain usually relatively small. Nevertheless, these studies represent a starting point for exploring the potential of PAT and determining its suitability for various cancer sites. It is worth noting that discrete arcs could remain relevant even as dynamic arc delivery becomes available. This could occur if technological advancements allow for “smoother” discrete arc delivery (i.e. step-and-shoot), through methods like auto beam sequencing [157], making it competitive in terms of delivery time with dynamic arcs.

In 2021, an important paper on the roadmap for proton therapy was authored by renowned experts but barely include any mention of PAT [282]. The rationale behind this omission was that the paper focused on technical advancements aimed at improving dose conformity, an aspect that PAT does not convincingly demonstrate in the current state-of-the-art. IMPT represents a cutting-edge advancement, excelling in its ability, to optimize highly conformal dose distributions while using a minimal number of beams. On the other hand, PAT appears somewhat contradictory because it employs many beam angles, potentially creating a non-negligible low-dose bath (i.e. low dose to high volume of healthy tissues). Research on clonogenic cell survival has indicated that RBE increases with lower doses in late-responding normal tissues [283]. The presence of low-dose regions thus raises concerns within the medical community regarding the potential risk of secondary cancers, particularly in pediatric patients [284, 285]. Previous studies [286, 287, 288, 289] have suggested that high total energy deposited in healthy tissues could contribute to the development of secondary cancers and may be a good estimator for quantifying cancer induction. However, the results observed with PAT planning thus far do not support the low-dose bath concern as integral dose reduction has consistently been observed compared to VMAT and/or IMPT [175, 174, 178, 179, 26, 184]. In addition, this reduction in integral dose appears to align with an equal or reduced risk of secondary malignancies, as indicated by

studies with PS-PAT [171, 169] or PBS-PAT [185] compared to VMAT, IMRT, or PS-PT depending on the study. Note that these cancer risk predictions rely on models that are still under investigation. The limited knowledge in this area is constrained by the amount of data collected within clinical cohorts and their follow-up periods.

Another aspect of concern with PAT is its potential impact on the patient's immune system, specifically regarding circulating lymphocytes in the bloodstream. Ionizing radiation is known to deplete circulating lymphocytes [290, 291], which has been associated with poor prognoses in various cancer types [292, 293, 294, 295]. This is especially relevant when irradiating brain tumors, where specific beam arrangement using fixed-angle IMPT can spare major cerebral arteries completely, while PAT may be more prone to expose them to some low-dose radiation. Therefore, from an immune response perspective, PAT may not be advantageous in such cases. Furthermore, because blood is in constant motion, assessing the impact on the immune system based solely on integral dose is insufficient. Factors like the transition time of the blood during the irradiation should also be considered. Tools and models are currently in development to estimate the dose to circulating blood in radiation therapy [296, 297]. It would be worthwhile to compare their evaluations between IMPT and PAT treatment plans to gain a better understanding of their respective impacts on the immune system.

Expanding on the topic of biological effects, an aspect disregarded in this thesis, but of extremely high importance, is the potential of PAT to finely shape the LET distribution used as a surrogate for RBE in optimization. Research has shown that the RBE of protons increases with LET [283]. This indicates that protons have their maximum biological impact when they come to a stop within the target volume, i.e. in the Bragg peak and the region of rapid dose fall-off. This can be a problem as high LET regions are often found in normal tissues distal to the target and shown to correlate with late toxicities [298, 46, 299]. While LET optimization is gaining consideration for IMPT treatment planning [300, 44, 301], it has limitations in fully confining high LET contributions within the target or moving them outside normal tissue [184]. PAT, with its increased DOF, offers the potential to leverage these limitations and move the high LET components within the target volume or at least moving them away from critical structures, sparing healthy tissue even more effectively [163, 161]. As a consequence, PAT could substantially impact clinical outcomes with LET optimization. However, it is important to acknowledge that the risks associated with high LET in OAR, such as potential side effects, are still not fully understood and require further research. Eventually, we will systematically incorporate the variable RBE model directly into optimization. This becomes especially important for heavy ions that have much

higher radiobiological effect prompting the development of heavy ion arc therapy to exploit the DOF for RBE-weighted dose optimization [167, 165]. Although variable RBE models are an ongoing area of research whose results should be approached with caution due to the large uncertainties and variabilities in RBE estimations and RBE-weighted doses [302].

Several treatment sites may particularly benefit from PAT, and these advantages can be quite specific. Brain cancer patients, for instance, could be one of the primary beneficiaries of PAT [303, 184]. Brain tumors often involve sensitive cognitive structures located in close proximity to the tumor, such as the brainstem whose radiation-induced injury can lead to life-threatening neurologic disorders [304, 305]. In conventional IMPT treatments, achieving proper target coverage while sparing these OAR can be challenging due to limited degrees of freedom and result in higher toxicity. There are also concerns regarding a potential connection between toxicity and LET hotspots in brain tumors [231], which PAT could potentially mitigate. This becomes especially critical for cases like chordoma [306], which require very high doses (greater than 70 Gy), or ependymoma (~ 60 Gy) located behind the brainstem [159, 307]. Clinical constraints typically restrict the dose the brainstem interior can receive to no more than 54 Gy [227], necessitating a compromise to effectively treat the tumor. Another potential application of PAT lies in the treatment of head and neck tumors. These treatments often involve complex field arrangements and are situated in anatomical regions with numerous sensitive organs, highly susceptible to radiation-induced toxicities. A study by de Jong et al. [187] suggested that discrete PAT treatment could expand the fraction of oropharyngeal cancer patients who could benefit from proton therapy in model-based patient selection settings [308]. Toxicities such as dysphagia (difficulty to swallow) or xerostomia (dry mouth) were shown to be further reduced by the use of PAT [157, 187]. Treatment sites characterized by challenging motion management (in particular for PT due to range uncertainties), could be promising candidates for PAT. These sites, often found in the thoracic region, are difficult to treat due to intra-fraction motion induced by breathing for instance. Thanks to the additional DOF, PAT may offer advantages in mitigating the so-called interplay effect [309, 310, 311] between the dynamic beam delivery and these intra-fraction movements, potentially resulting in more accurate and effective treatments [175, 176]. In turn, this may eliminate the need for rescanning techniques [312, 313, 314, 315]. Rescanning involves repeatedly scanning the energy layers of each beam to statistically average the effects of motion. By eliminating this step with PAT delivery, it could further enhance the speed of treatment delivery. Nevertheless, further research is required to demonstrate if the interplay effect may be systematically reduced.

Finally, when considering treatment delivery time and treatment workflow, it is common to draw parallels between PAT and VMAT. It is expected that PAT will streamline treatment procedures similar to how VMAT did compared to IMRT. However, there are important distinctions to keep in mind. IMPT requires fewer beams compared to IMRT [316] to achieve high dose conformity because of the different depth-dose curves of protons compared to photons. As a result, IMPT plans are typically already quite efficient in terms of total delivery time. The impact of arcing techniques on PT delivery may thus, not be as profound as observed for photon therapy. The treatment delivery time data reported in Table 2.1 in Section 2.4.5 generally support this viewpoint, although additional data is required to draw conclusions.

Regarding treatment workflow, this could be a great advantage for PAT [317], especially considering that the use of range shifters, which requires to be inserted and removed with a remote controller, may no longer be necessary to target shallow tumors with lower energy requirements [157, 179], reducing the overall treatment time. Moreover, manual intervention is often necessary in IMPT workflows to switch between beams, particularly when employing non-coplanar beams that require couch rotation and potentially intermediate re-imaging steps for alignment verification. It is worth noting that similar limitations could be observed with non-coplanar arcs, which may be better suited for specific tumor sites. However, differences between PAT and IMPT workflows may become less meaningful if we move towards auto beam sequencing, eliminating beam loading time and automating non-safety related user interactions with the treatment software [157].

The future of PAT

During my PhD research, I encountered various opinions about the potential benefits of PAT and whether the technological constraints, hardware and software, required to implement it are justified. The clinical results of PAT have been mixed thus far, which has led to reluctance in some quarters, in particular attributed to the low-dose bath. Additionally IMPT is already quite effective, thanks to its degeneracy. The emergence of FLASH PT [57] has also captured the attention of many, diverting focus away from PAT and even sparking debates about which modality is the most promising for the next generation of PT treatments [97]. This raises the question: Is PAT really worth the effort?

In my opinion, this is precisely what science is all about. As long as we have the resources and capabilities to investigate PAT, we should not hesitate to pursue it. The rationale behind PAT remains strong and should not be underestimated.

While challenges persist, they will also evolve alongside technological advancements. The goal is not to prove that PAT will completely revolutionize PT, replacing standard delivery methods for all tumor sites. Instead, our focus should be on identifying the specific treatment sites that could benefit the most from PAT delivery. Will PAT accelerate treatment delivery time? I have reservations about this, except in some specific disease sites. It will surely simplify the treatment workflow. Will it improve the dose metrics and show clinical significance? Possibly, but drawing definitive conclusions remains challenging as research on PAT optimizers is still ongoing. Will it have an impact on radiobiology? Most likely, given its high flexibility in managing high LET hotspots.

With the imminent arrival of IBA's DynamicArc® as the first product of its kind, we should gain further insights into the value of PAT. Furthermore, the prospect of adopting gantry-less systems could greatly enhance the clinical viability of PAT for proton therapy treatments. In general, PAT should be viewed as an innovative addition to the radiation therapy toolbox, with the potential to attract more patient referrals to PT [187] and ultimately contribute to the global democratization of PT [13]. In other words, it may broaden the range of patients who can benefit from PT, leading to better outcomes, and make PT more accessible to individuals worldwide. In the end, this is what matters the most; create the conditions that provide the best treatment for cancer patients.

Returning to the main topic of this thesis, the optimization of PAT treatments, we have accumulated substantial knowledge and proposed various solutions to address associated challenges. While the quest for the ideal PAT optimizer is not over, the immediate priority is to gather feedback from planners to refine the optimizer and generate clinical evidence using larger patient cohorts. Although minimizing the BDT (or a surrogate) of PAT plans is undoubtedly important and has been a central aspect in this thesis, it should be considered within the broader context of clinical treatment. BDT in PT constitutes approximately 20% of the overall treatment time [318], gaining a few percentage points by using a more realistic BDT model may not be the most critical factor to increase patient throughput. However, it could potentially benefit the treatment itself by reducing intrafraction uncertainties, but this still requires further demonstration.

While minimizing BDT can be achieved using various approximations and surrogates, there is also the question of how to report BDT evaluation results, as each institution usually employs its in-house BDT model. These models can range from simplistic to highly sophisticated, accounting for irradiation parameters and mechanical constraints related to spot-scanning and gantry rotation. Consequently, it becomes difficult to directly compare results across different research papers due

to this lack of standardization in BDT modeling. It is essential for the research community to establish common ground in terms of BDT modeling to facilitate meaningful comparisons between studies. Machine-specific models [147, 148, 150] and independent parametric BDT calculators, like ATOM [229], contribute to addressing this issue and should be further exploited.

In conclusion, PAT shows promise as an alternative to IMPT, offering potential benefits in terms of dose distribution, biological effects, and overall treatment workflow. However, its successful integration into clinical practice hinges on the optimization algorithms effectively balancing the triad of plan quality, delivery time, and computational efficiency. Designing an optimizer capable of harnessing the increase of degree of freedom in PAT is a mathematical challenge. Future research should focus on systematically assessing the clinical impact of PAT through large cohort independent studies and enhancing optimization methods based on feedbacks from planners. Only after conducting such evaluations can definitive conclusions be drawn about the efficacy of algorithms and clinical value of PAT. Finally, this thesis is also an open invitation to optimization experts from outside the field of radiotherapy to apply their expertise, contributing to the ongoing efforts to improve PAT optimization algorithms.

Appendix A

Supplementary material Chapter 3

A.1 (Hyper-)parameters

In this appendix, the computing performances of each algorithm are analyzed. The hyper-parameters are identified and carefully tuned to achieve the ArcPT goals. Several metrics are therefore evaluated. First, the layer sparsity, ideally aimed at 100%, *i.e.* at most one energy layer active per beam angle. Secondly, we seek to minimize as much as possible the number of energy switch-ups as well as the dose fidelity term to follow the physician's prescription. Finally, the homogeneity index ($HI = D5-D95$) is used to quantify the uniformity of the dose in the target. It should be as small as possible. In our problem set-up, only a single clinical objective on the target (uniform dose of 60 Gy) is implemented in the cost function. For the sake of the analysis, we show results only for the phantom data set created with 2-mm dose resolution (acceptable in clinic) and comprising 46 beams (step angle = 2°).

A.1.1 FISTArcPT

The FISTArcPT algorithm minimizes the cost function described in Eq. 3.18. This parsimonious function is built on several terms with specific purposes, each controlled by a weighting parameter. By varying the value of these parameters, different results can be obtained. The physicist can therefore fine-tune these parameters in order to achieve the primary goals of the treatment under consideration.

Analysis of weight λ

The first parameter to be analyzed is the weight of the group sparsity term λ . Recall that we aim at having no more than one energy layer active per beam, hence 100% in layer sparsity. In Fig. A.1a, it can be seen, that acceptable arc sparsity is obtained from $\lambda = 1$ with good dose homogeneity. Fig. A.1b shows that the dose fidelity follows the same pattern as the dose homogeneity in Fig. A.1a. This behavior is expected given the fact that the dose fidelity cost is a squared norm of the difference between total dose computed and the prescription. It also shows that λ should not be too low in order to turn off non-interesting layers that can potentially reduce the number of switch-ups (SU).

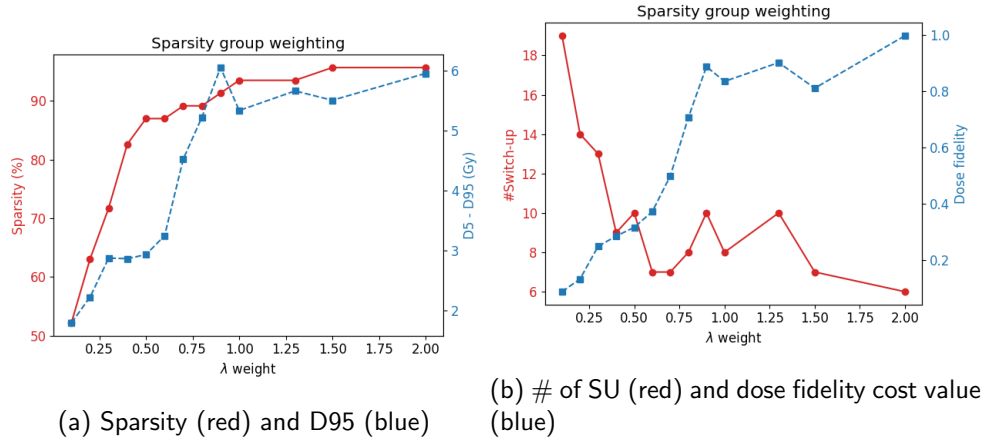


Figure A.1: FISTArcPT metric versus sparsity group weighting λ .

Analysis of weight ϵ

Surprisingly, ϵ has no real impact on this plan. Actually, the indicator function is not that useful in this specific case. The combination of dose fidelity, sparsity and energy sequencing already gives an acceptable arc plan. It is only when $\epsilon > 1.5$, that the plan is affected. In fact, layer sparsity decreases because the re-activation of layers starts to become more important than the removal.

Analysis of weight γ

Another important parameter to consider is the energy sequencing weight γ . Results are shown in Fig. A.2 with the sparsity group weight fixed at $\lambda = 1$. This parameter was really hard to tune due to the non-convex nature of the energy

sequencing term. Its impact was indeed more unpredictable. A value of $\gamma = 0.1$ was eventually picked following the results on Fig. A.2. If γ is too large, the gradient descent is influenced too much by the energy sequencing term and not enough by the treatment objectives (encoded in the fidelity term) and the sparsity in layers is difficult to obtain. While if γ is too small, the energy sequencing term will not have any impact anymore and the energy layer switching time will not be optimized. Additionally, it is noteworthy to point out the high correlation of ELST with the number of energy Switch-Up and the HI with the dose fidelity objective value, which can be visually observed from Fig. A.2.

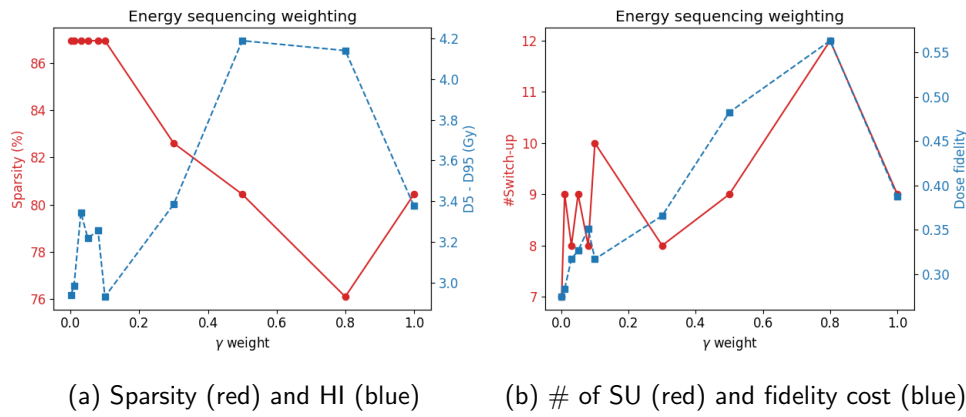


Figure A.2: FISTArcPT metric versus energy sequencing weighting γ .

From those results, we decided to fix values for the weighting parameters of the objective function for the next experiences. Table A.1 summarized the chosen parameter values.

Parameter names	λ	γ	ϵ
Value	1.	0.1	0.1

Table A.1: Tuned parameters in FISTArcPT algorithm.

A.1.2 Local Search

The local search algorithm, even though differing radically from the FISTArcPT, has also several inputs to tweak and orientate the search in the space of solutions. Due to the non-deterministic character of LS, each experiment was run 5 times. Average and standard deviations over these 5 runs are therefore reported.

Analysis of weight γ

As FISTArcPT, the local search algorithm has a weight controlling the importance of the energy sequencing term in the objective function. Fig. A.3 illustrates the impact of γ weight on the number of switch-ups and fidelity term. A value $\gamma = 50$ seems to be the best compromise to have an acceptable ELST and homogeneous dose in the target.

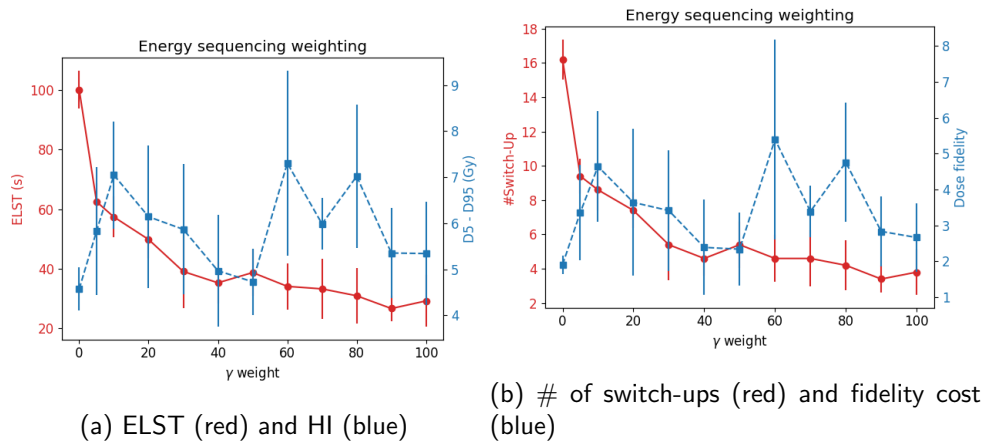


Figure A.3: LS metric versus energy sequencing weighting γ for 5 runs.

Minimum number of active energy layers per beam

We can set a minimum number of energy layers to be activated in each beam as a hard constraint in the configuration of the LS algorithm. Table A.2 shows a comparison where the minimum number of active layers in a beam is zero and situation like FISTArcPT, where the minimum is set to one. It is obvious that constraining the algorithm to activate at least one energy layer per control point is really restrictive and suppresses degrees of freedom resulting in higher cost, worse HI, and worse ELST.

	Minimum 0 active layer in each beam	Minimum 1 active layer in each beam
Fidelity cost	1.29 ± 0.27	1.66 ± 0.58
D5 - D95 (Gy)	7.04 ± 1.21	7.79 ± 1.66
ELST (s)	57.38 ± 6.86	96.66 ± 10.66
#Switch-Up	8.60 ± 1.20	15.00 ± 2.10

Table A.2: Impact of minimum number of active layer per beam for 5 runs.

Maximum number of active energy layer per beam

As we have set a minimum of active energy layers per beam, we can also set a superior limit on this number. Table A.3 summarizes the results. Allowing more than one layer active per beam gives better dosimetric outcomes and results in a decreased fidelity cost but at the expense of an increasing energy layer switching time.

	Max 1 active layer in each beam	Max 2 active layers in each beam	Max 3 active layers in each beam
Fidelity cost	1.29 ± 0.27	1.19 ± 0.57	0.69 ± 0.33
D5 - D95 (Gy)	7.04 ± 1.21	6.62 ± 1.55	4.81 ± 1.30
ELST (s)	57.38 ± 6.86	80.56 ± 13.36	95.36 ± 19.64
#Switch-Up	8.60 ± 1.20	11.20 ± 2.04	12.80 ± 2.79
Sparsity (%)	100.00 ± 0.00	49.13 ± 7.48	59.57 ± 10.69

Table A.3: Impact of maximum number of active layer per beam for 5 runs.

Initial temperature and cooling factor

This couple of parameters influence the behavior of local search and are used to play on the diversification/intensification trade-off. The higher (*resp.* lower) the initial temperature (*resp.* cooling factor), the more iterations between two resets and, therefore, the higher diversification. On the opposite, a smaller initial temperature and a higher cooling factor would increase the intensification. The initial temperature and cooling factor were set according to previous knowledge to 5 and 0.7 respectively.

Parameter names	γ	minEL in beams	maxEL in beams	Target weight	Max spot intensity
Value	50.0	0	1	100.0	20

Table A.4: Tuned parameters in LS algorithm.

From the results obtained in this section, we decided to fix the parameters for the next experiences. Table A.4 reports their values. In this table, the max spot intensity refers to maximum value a spot weight can take. We voluntarily limited it to 20 to avoid hot spot¹.

¹Edited for published version to better define the max spot intensity.

A.2 Convergence

For FISTArcPT, the stopping criterion used in the optimization is the maximum number of iterations. We set it to 1000 to get a good quality and deliverable plan and achieve good layer sparsity. Fig. A.4 displays the result. Except the indicator function, all terms involved in the objective function described in Eq. 3.18 are plotted. The total cost in general shows a rapid decrease and then a smooth convergence and is mainly influenced by the group sparsity term. The energy sequencing term is more chaotic but still converges after a few hundreds of iterations.

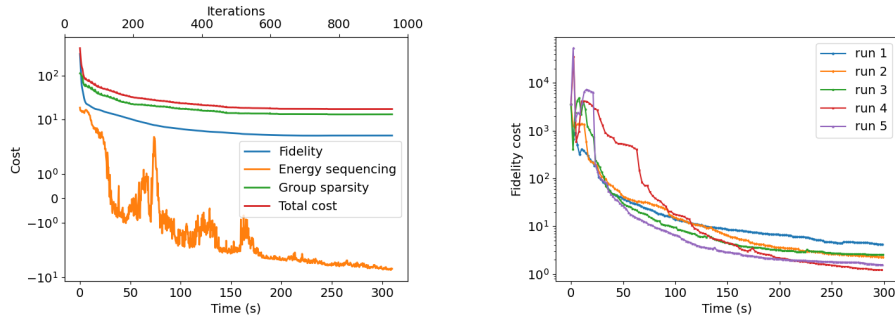


Figure A.4: Convergence of FISTArcPT Figure A.5: Convergence of 5 runs of LS-ArcPT with the same input parameters.

Regarding the local search algorithm, the objective function to be minimized (Eq. 3.22) is simply a quadratic function of the difference between the total dose delivered to the target and the prescription (omitting the OARs) to which the ELST and sparsity objectives are added. Fig. A.5 illustrates the evolution of the cost function for 5 runs. The stopping criterion is a time limit set to 5 minutes. As expected, the local search is really chaotic at the beginning, exploring solutions in a large neighborhood but little by little, converges smoothly towards an optimal solution along with the decrease of the temperature of the system. Compared to FISTA, the convergence is certainly slower but a rapid solution with specific properties could potentially be used as input to FISTA.

A.3 Problem scaling

The scale of the problem constitutes a major issue in ArcPT optimization. Depending on the resolution and size of the beamlet matrix, the algorithms will behave in different ways. We decided to pick 9 data samples and to analyze them: the two extremes and the middle one, *i.e.* phantoms with 1, 3, and 5 mm

resolution and with a plan composed of 10, 31, and 91 beams. For the sake of the comparison, we fixed the maximum number of iterations of FISTArcPT to 1000. This way, each test run of FISTA determines the time limit to be set in the LS and MIP algorithm input parameters (though MIP could stop before the time limit if the optimal solution was found).

As it could be expected, each algorithm behaves in a very different way and this is summarized in Tables A.5-A.6-A.7. FISTArcPT gets better results in terms of dose homogeneity (D5 - D95) for lower resolution data samples and smaller number of beams while on the other side, LS-ArcPT behaves in a complete opposite way with better results at high resolution with a high number of beams. For both algorithms, the ELST increases with the number of beams in a logical way and are quite compatible with each other. The MIP algorithm, on the other hand, does not follow the same pattern. Indeed, the maximum number of energy upwards switches has been constrained to 0 instead of being left as an optimization objective, hence the very low ELST for each test set. As it can be observed in Table A.7, the MIP gives excellent results in terms of dose homogeneity within the target volume and this, for any of the test sets. The 5-mm resolution data sets could achieve a better homogeneity index compared to other resolution sets because the MIP let us push the constraints on the target coverage harder ($[p_{\min T} = 60, p_{\max T} = 60.5]$ Gy instead of $[58, 62]$ Gy for the others). It can also be noticed that the boxes regarding the 1-mm resolution data set with 10 beams in the arc are empty. The reason is that the MIP could not find any feasible solution with the constraints given even though many combinations were tried.

In terms of optimization time, FISTArcPT behaves as expected because, a higher number of beams translates into a higher number of beamlets and then a more computationally expensive gradient. The MIP, for its part, was able to give an optimal solution in a shorter time than the duration limit set by the FISTA algorithm and this, for any of the test sets. Obviously, the resolution also plays a role in optimization time with higher resolution, *i.e.* higher number of voxels that translates in larger optimization time. Despite the fact that this claim is valid for any of the presented optimization algorithm, the MIP seems to be the one that could be the most impacted with the time exponentially with the problem size.

	Resolution	# beams		
		10	31	91
ELST (total)	1 mm	22.5"	32.9"	75.6"
	3 mm	29.8"	34.7"	40.2"
	5 mm	21.9"	34.1"	40.2"
Optimization time	1 mm	2435"	2595"	3640"
	3 mm	180"	500"	1800"
	5 mm	60"	160"	860"
D5 - D95 (Gy)	1 mm	5.0	4.7	5.5
	3 mm	3.1	3.4	4.0
	5 mm	1.6	2.0	2.6
Layer	1 mm	60.0	90.32	92.31
	3 mm	50.0	83.87	98.9
Sparsity (%)	5 mm	60.0	90.32	96.7

Table A.5: Summary of FISTArcPT experiences. The poorest resolution and the lowest number of beams produced the best homogeneity in the target volume.

	Resolution	# beams		
		10	31	91
ELST (total)	1 mm	25.0"	41.4"	56.6"
	3 mm	14.6"	16.4"	82.8"
	5 mm	13.4"	32.3"	55.4"
D5 - D95 (Gy)	1 mm	9.5	4.6	5.4
	3 mm	14.2	6.3	2.9
	5 mm	7.8	1.8	1.4

Table A.6: Summary of LS experiences.

	Resolution	# beams		
		10	31	91
ELST (total)	1 mm	/	3"	169.5"
	3 mm	2.4"	16.4"	3.6"
	5 mm	8.5"	3.6"	3.6"
Optimization time	1 mm	/	2041.1"	1357.8"
	3 mm	22.3"	65.8"	398.6"
	5 mm	2.9"	12.5"	97.8"
D5 - D95 (Gy)	1 mm	/	4.1	3.8
	3 mm	5.1	5.6	6.2
	5 mm	0.7	1.2	2.1
Layer	1 mm	/	100.0	78.0
	3 mm	100.0	87.1	100.0
Sparsity (%)	5 mm	80.0	100	100

Table A.7: Summary of MIP experiences.

Appendix B

Supplementary material Chapter 4

B.1 MIP model

We represent a specific sequence of energy layers (EL) as a path in a directed graph, from a source node 0^- which has outgoing directed edges $(0^-, n)$ to every EL $n \in E$, to a sink node 0^+ which has incoming edges from every EL, passing by a network of edges that link the ELs altogether: there is an edge (m, n) for each pair of ELs such that n belongs to a strictly higher beam than m . We note E^{edges} the set of all edges that constitute the graph, whereas $E^- = E \cup \{0^-\}$, $E^+ = E \cup \{0^+\}$, $E^\pm = E \cup \{0^-, 0^+\}$ denotes the nodes.

Let x_j and e_{mn} be decision variables such that

- $\forall j \in S : x_j$ determines the intensity attributed to a spot $j \in S$,
- e_{mn} are redundant variables determining whether some energy layer (EL) switch happens: $\forall m, n \in E^\pm : e_{mn} = 1$ if $(m, n) \in E^{\text{edges}}$ and layer $n \in E^+$ is activated directly after layer $m \in E^-$, otherwise $e_{mn} = 0$.

From [212], the proton arc therapy treatment optimization problem for finding a single optimal point on the Pareto front can be modelled as a weighted sum

using a mixed-integer two-index flow formulation

$$\begin{aligned}
 \min_{x_1, \dots, x_{|S|}} \quad & F(\mathbf{x}) := w_D f(\mathbf{d}) + w_T \text{BDT}^\approx(\mathbf{x}) \quad (\text{B.1}) \\
 \text{s.t. :} \quad & \sum_{n \in E^+} e_{0^-n} = \sum_{n \in E^-} e_{n0^+} = 1 \quad (\text{B.2}) \\
 \forall m \in E \quad & \sum_{l \in E^-} e_{lm} = \sum_{n \in E^+} e_{mn} \quad (\text{B.3}) \\
 \forall m \in E, \forall j \in S_m \quad & x_j \leq I \sum_{l \in E^-} e_{lm} \quad (\text{B.4}) \\
 \forall j \in S \quad & 0 \leq x_j \leq I \quad (\text{B.5}) \\
 \forall (m, n) \in E^{\text{edges}} \quad & e_{mn} \in \{0, 1\} \quad (\text{B.6})
 \end{aligned}$$

where the fidelity term $f(\mathbf{d})$ is represented as follows:

$$\begin{aligned}
 f(\mathbf{d}) &= \sum_{k \in \text{ROI}} u_k f_k(\mathbf{d}) \quad (\text{B.7}) \\
 \text{s.t.} \quad & f_k(\mathbf{d}) = \|(d_i - p_i)_+\|_1 \quad (\text{B.8}) \\
 & \mathbf{d} = \mathcal{A}\mathbf{x} \quad (\text{B.9})
 \end{aligned}$$

and the approximation of the irradiation time $\text{BDT}^\approx(\mathbf{x})$ is modeled as:

$$\begin{aligned}
 \text{BDT}^\approx(\mathbf{x}) &= \sum_{(m,n) \in E^{\text{edges}}} c_{mn} e_{mn} \\
 \text{s.t. :} \quad & \forall (m, n) \in E^{\text{edges}} \quad e_{mn} \in \{0, 1\} \\
 & c_{mn} = \begin{cases} 6.5 & \text{for upwards energy switching} \\ 1.6 & \text{for downwards energy switching} \\ 1.2 & \text{for constant energy} \end{cases}
 \end{aligned}$$

B.2 The BCO Pareto approximation algorithm

This appendix provides details on the BCO algorithm through a simple example as well as a full description of the equations derived for the Pareto approximation construction.

Following the path depicted in the binary tree search (dashed nodes in Figure 4.1), we give hereinafter a detailed description of the different steps within the

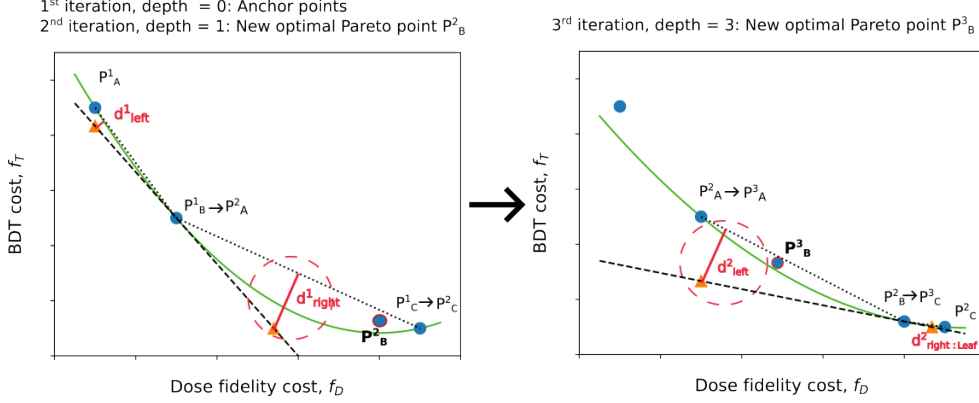


Figure B.1: Pareto front geometric construction. Left (iteration 1 & 2): Constructing depth 1 on the right hand side branch of the tree depicted in Figure 4.1. Right (iteration 3): Constructing depth 2. Blue: Pareto optimal points (the red encircled ones are selected for next iteration). Green: upper Pareto quadratic approximation. Distal points (orange) and their distance (red) to segment lines.

iterations. **Iteration 1.** The first iteration starts from the parent node where we compute the so-called anchor points P_A^1 , P_C^1 representing the individual minima of each objective. In the present study, these points are obtained by optimizing the global objective function with individual weights $\mathbf{w} = (w_D, w_T)$ for the dose fidelity objective f_D and the BDT $\approx$ objective f_T . These weights are set to $\mathbf{w}_A = (1, 0.001)$, $\mathbf{w}_C = (0.001, 1)$, respectively, to avoid too poor quality solutions in terms of dose coverage. Next, a third (interior) point P_B^1 is obtained with equal weights, $\mathbf{w}_B = (1, 1)$. From these three points, an initial upper Pareto quadratic approximation is obtained. Using the bounding segment lines between these points and the tangent line at the interior point, initial lower distal points and their distance to segment lines $P_A^1 P_B^1$ and $P_B^1 P_C^1$ (d_{left}^1 and d_{right}^1 , respectively) can be computed by solving the equation system: (see equations B.12 and B.13 in the formulas below). This step is illustrated geometrically on Figure B.1 (left). At this stage, the parent node has now two children nodes corresponding to each distance.

Iteration 2. In the next iteration $i = 2$, we pick the tree node associated with the currently largest d^{i-1} distance. In our example, d_{right}^1 (resulting from P_B^1 and P_C^1) is picked, the new weights to run to obtain a new Pareto optimal point (P_B^2) are obtained by averaging the weights of P_B^1 and P_C^1 . The same geometric construction as in iteration 1 is carried out to compute the segment lines d_{left}^2 and d_{right}^2 .

Iteration 3. As d_{right}^2 is lower than a predefined tolerance threshold, we then cut

the current branch (*leaf* node in tree). The third iteration thus continues with the tree node associated with d_{left}^2 . It generates a new point P_B^3 (Figure B.1, right) and so on.

These iterations (summarized in Algorithm 3) are repeated until all distances d^i fall below minimal distance tolerance, or the user-specified tree depth $I_{\max}$ is reached.

In order to mitigate the bias introduced by objectives having different magnitudes, a dimension scaling on the individual objective functions is performed whenever a new point $\mathbf{P}_B^i$ is computed, still following Craft et al. [25] method.

Formulas

The anchor points for input upper Pareto approximation are $\mathbf{P}_A = (x_A, y_A)$ for $f(\mathbf{d})$ and $\mathbf{P}_C = (x_C, y_C)$ for $\text{BDT}^\approx(\mathbf{x})$, the single interior point is $\mathbf{P}_B = (x_B, y_B)$. Then the upper Pareto quadratic approximation is

$$y = \frac{(x - x_C)(x - x_B)}{(x_A - x_C)(x_A - x_B)}y_A + \frac{(x - x_A)(x - x_B)}{(x_C - x_A)(x_C - x_B)}y_C + \frac{(x - x_A)(x - x_C)}{(x_B - x_A)(x_B - x_C)}y_B$$

The slope of its tangent line is

$$y' = \frac{(x - x_C) + (x - x_B)}{(x_A - x_C)(x_A - x_B)}y_A + \frac{(x - x_A) + (x - x_B)}{(x_C - x_A)(x_C - x_B)}y_C + \frac{(x - x_A) + (x - x_C)}{(x_B - x_A)(x_B - x_C)}y_B$$

The tangent line at the interior point has the slope

$$y'(x_B) = \frac{x_B - x_C}{(x_A - x_C)(x_A - x_B)}y_A + \frac{x_B - x_A}{(x_C - x_A)(x_C - x_B)}y_C + \frac{(x_B - x_A) + (x_B - x_C)}{(x_B - x_A)(x_B - x_C)}y_B$$

The tangent line at the interior point is

$$y - y_B = y'(x_B)(x - x_B)$$

The initial lower distal point (LDP) to facet $\mathbf{P}_A\mathbf{P}_B$ is

$$\mathbf{I}_{\text{left}} = (x_A, y_B + y'(x_B)(x_A - x_B)) \quad (\text{B.10})$$

The initial LDP to facet $\mathbf{P}_C\mathbf{P}_B$ is

$$\mathbf{I}_{\text{right}} = (x_B + \frac{y_C - y_B}{y'(x_B)}, y_C) \quad (\text{B.11})$$

The equation of facet $\mathbf{P}_A\mathbf{P}_B$ is

$$y - y_A = \frac{y_B - y_A}{x_B - x_A}(x - x_A)$$

that is, in standard form

$$(y_B - y_A)x - (x_B - x_A)y + y_A(x_B - x_A) - x_A(y_B - y_A) = 0$$

The distance from LDP $\mathbf{I}_{\text{left}}$ to facet $\mathbf{P}_A\mathbf{P}_B$ is

$$d_{\text{left}} = \frac{|(y_B - y_A)x_A - (x_B - x_A)[y_B - y'(x_B)(x_B - x_A)] + y_A(x_B - x_A) - x_A(y_B - y_A)|}{\sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}} \quad (\text{B.12})$$

Similarly, the equation of facet $\mathbf{P}_C\mathbf{P}_B$ is

$$(y_B - y_C)x - (x_B - x_C)y + y_C(x_B - x_C) - x_C(y_B - y_C) = 0$$

The distance from LDP $\mathbf{I}_{\text{right}}$ to facet $\mathbf{P}_C\mathbf{P}_B$ is

$$d_{\text{right}} = \frac{|(y_B - y_C)[x_B + \frac{y_C - y_B}{y'(x_B)}] - (x_B - x_C)y_C + y_C(x_B - x_C) - x_C(y_B - y_C)|}{\sqrt{(x_B - x_C)^2 + (y_B - y_C)^2}} \quad (\text{B.13})$$

B.3 Benchmark

For each disease site, IMPT and SPArc dose distributions are shown in Figure B.2. The DVHs are provided in Figure B.3. To give an appreciation of the MIP plan quality compared to IMPT and SPArc, a pastel shade band was added on top of the DVH curves, with a thickness proportional to the range of solutions between the MIP plans D_{t} and d_{T} .

In Section 4.3, we have argued that the plan parameters for the IMPT and SPArc plans had to be fine-tuned from those used in the MIP plans, in order to reach a sufficient target coverage. We provide additional material to support this assertion. In Table B.1, original plan parameters (used for the MIP plans) and tuned parameters are listed. Table B.2 reports the target coverage obtained for each plan and each patient, as well as the initial number of spots at plan creation and thus available to the optimizer.

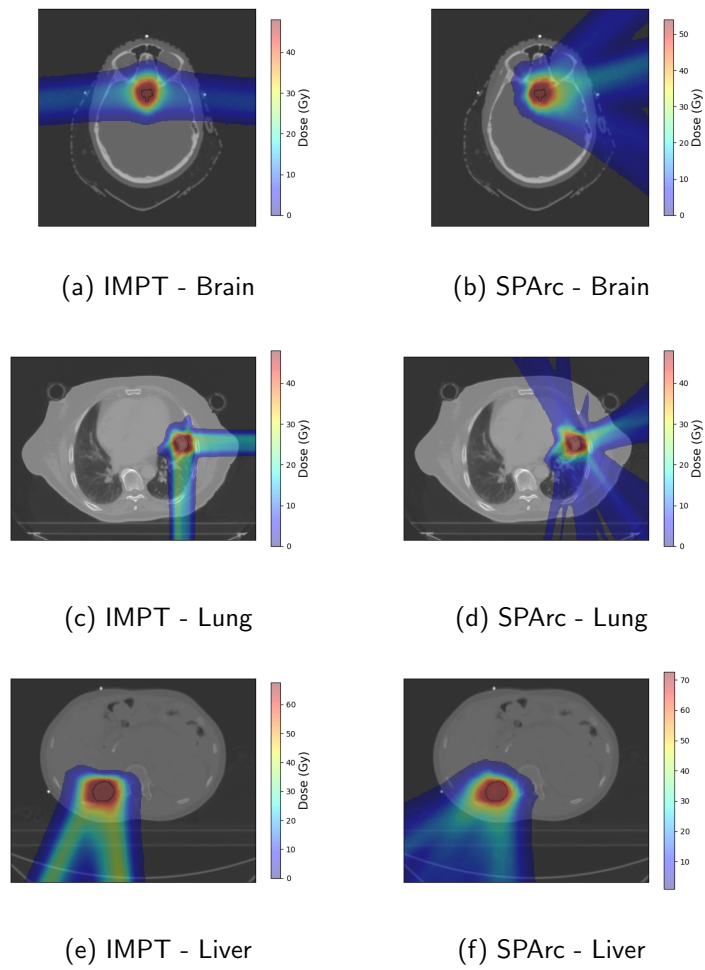


Figure B.2: Optimal dose distributions for 3 patients

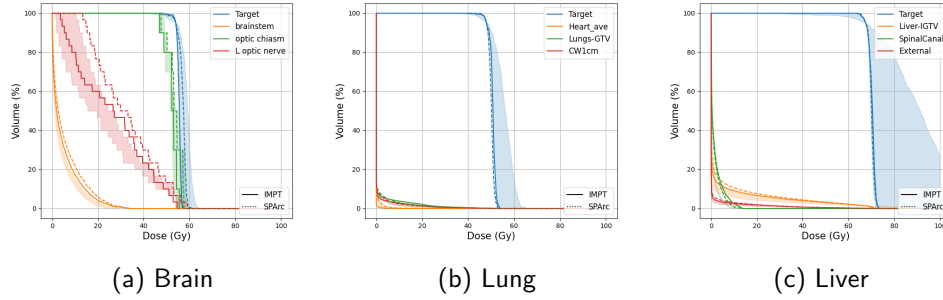


Figure B.3: DVH comparison between IMPT (solid line), SPARc (dashed line) and MIP-BCO solutions (filled area)

Plan parameters	Disease site	Original	Tuned
Spot spacing [mm]	Brain	7	5
	Lung		7
	Liver		5
Layer spacing [mm]	Brain	4	3
	Lung		4
	Liver		3
Target margin [mm]	Brain	6	7
	Lung		
	Liver		

Table B.1: Plan parameters for spot placement

	Modality	IMPT		SPARc	
	Plan parameters	Original	Tuned	Original	Tuned
D5-D98 (Gy)	Brain	7.5	3.4	10.9	6.3
	Lung	5.1		6.9	4.6
	Liver	18.5	5.8	11.8	5.7
# spots	Brain	459	1352	475	614
	Lung	312		472	550
	Liver	880	2481	1217	2698

Table B.2: Target coverage comparison of plans using original (MIP) and tuned plan parameters

Appendix C

Supplementary material Chapter 5

This appendix provides a detailed dosimetric comparison for organs-at-risk (OAR) from the study presented in Chapter 5.

C.1 Organs-at-risk dosimetry

The dosimetric comparison of OAR is shown in the **nominal** (Table C.1) and **worst-case** (Table C.2) scenarios. It compares results obtained with the in-house energy layer pre-selection (Flexi-Arc), the early layer and spot assignment (ELSA) pre-selection used to design proton arc plans as well to an IMPT plan. We observe that the Flexi-Arc plan with 16 sectors is the plan that achieves the lowest dose reduction to the brain stem and chiasm. For the rest of the OAR, it is more or less equivalent to the ELSA plan with 2 degree gantry angle spacing. The ocular system is much better spared in the IMPT plan compared to proton arc plans in general. Similar conclusions can be made for the worst-case scenario results.

	Flexi-Arc		ELSA		IMPT	
Gantry angle spacing (°)	1	1	4	3	2	
NSectors	8	16	5	5	8	
D0.03cc Brain stem interior (Gy)	51.1	49.4	52.0	53.3	51.5	51.1
D0.03cc Brain stem surface (Gy)	54.5	53.2	54.3	54.9	54.7	54.2
D0.03cc Chiasm (Gy)	55.3	54.8	55.2	55.6	55.3	55.5
D0.03cc Optic nerve L (Gy)	53.5	52.9	52.9	52.1	52.4	55.0
D0.03cc Optic nerve R (Gy)	50.9	51.0	53.5	49.5	50.3	53.1
D0.03cc Lens L (Gy)	0.7	0.2	0.5	0.3	0.7	0.0
D0.03cc Lens R (Gy)	0.4	0.2	0.9	1.1	0.3	0.0
D0.03cc Retina L (Gy)	3.0	1.6	3.4	2.9	2.9	0.0
D0.03cc Retina R (Gy)	1.2	0.9	3.1	2.5	1.1	0.0
D40% Hippocampus L (Gy)	1.0	0.9	3.5	3.3	1.1	0.2
D40% Hippocampus R (Gy)	2.1	2.3	4.0	3.2	2.6	0.6
Dmean Eye L (Gy)	0.9	0.6	1.0	1.0	1.0	0.0
Dmean Eye R (Gy)	0.5	0.3	1.0	1.0	0.5	0.0
Dmean Lacrimal gland L (Gy)	0.1	1.1	1.5	2.4	1.0	0.0
Dmean Lacrimal gland R (Gy)	1.0	0.6	0.6	0.6	1.0	0.0
Dmean Cochlea L (Gy)	1.0	0.7	2.0	2.0	1.0	0.0
Dmean Cochlea R (Gy)	0.9	0.8	2.0	2.0	0.9	0.0

Table C.1: Detailed dosimetric comparison for organs-at-risk in the **nominal** scenario. *Abbreviations* : D0.03 cc = near maximum dose to 0.3 cc of structure/organ, Dmean = mean dose, D40% = dose to 40% of the volume of both hippocampi, L = left, R = right.

	Flexi-Arc		ELSA		IMPT	
Gantry angle spacing (°)	1	1	4	3	2	
NSectors	8	16	5	5	8	
D0.03cc Brain stem interior (Gy)	53.8	52.4	53.4	54.6	54.2	53.9
D0.03cc Brain stem surface (Gy)	55.7	54.3	55.2	55.8	55.7	56.0
D0.03cc Chiasm (Gy)	55.8	55.7	55.9	56.3	55.9	55.8
D0.03cc Optic nerve L (Gy)	54.8	55.3	55.2	54.4	53.7	55.5
D0.03cc Optic nerve R (Gy)	54.0	53.3	56.5	52.8	54.0	54.5
D0.03cc Lens L (Gy)	1.0	0.3	0.9	0.7	1.0	0.0
D0.03cc Lens R (Gy)	0.0	0.3	1.2	1.2	0.5	0.0
D0.03cc Retina L (Gy)	5.3	2.5	4.2	4.9	3.8	0.0
D0.03cc Retina R (Gy)	2.3	1.8	4.2	3.3	1.5	0.0
D40% Hippocampus L (Gy)	1.6	1.5	4.7	4.8	1.8	0.6
D40% Hippocampus R (Gy)	3.5	3.3	5.3	4.5	3.9	1.3
Dmean Eye L (Gy)	1.2	0.8	1.5	1.6	1.4	0.0
Dmean Eye R (Gy)	0.7	0.4	1.3	1.2	0.6	0.0
Dmean Lacrimal gland L (Gy)	0.2	1.5	2.3	3.0	1.2	0.0
Dmean Lacrimal gland R (Gy)	1.4	0.9	0.7	0.9	1.4	0.0
Dmean Cochlea L (Gy)	1.6	1.3	2.8	2.9	1.9	0.0
Dmean Cochlea R (Gy)	1.9	1.5	2.8	3.5	1.6	0.0

Table C.2: Detailed dosimetric comparison for organs-at-risk in the **worst-case** scenario. Abbreviations : D0.03 cc = near maximum dose to 0.3 cc of structure/organ, Dmean = mean dose, D40% = dose to 40% of the volume of both hippocampi, L = left, R = right.

References

- [1] J. Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray. Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, 149(4):778–789, Aug. 2021.
- [2] J. M. Borrás, Y. Lievens, P. Dunscombe, M. Coffey, J. Malicki, J. Corral, C. Gasparotto, N. Defourny, M. Barton, R. Verhoeven, L. Van Eycken, M. Primic-Zakelj, M. Trojanowski, P. Stojan, and C. Grau. The optimal utilization proportion of external beam radiotherapy in European countries: An ESTRO-HERO analysis. *Radiotherapy and Oncology*, 116(1):38–44, July 2015.
- [3] P. Kallman, B. Lind, A. Eklof, and A. Brahme. Shaping of arbitrary dose distributions by dynamic multileaf collimation. *Physics in Medicine and Biology*, 33(11):1291–1300, Nov. 1988.
- [4] D. J. Convery and S. Webb. Generation of discrete beam-intensity modulation by dynamic multileaf collimation under minimum leaf separation constraints. *Physics in Medicine and Biology*, 43(9):2521–2538, Sept. 1998.
- [5] T. R. Mackie, T. Holmes, S. Swerdloff, P. Reckwerdt, J. O. Deasy, J. Yang, B. Paliwal, and T. Kinsella. Tomotherapy: A new concept for the delivery of dynamic conformal radiotherapy. *Medical Physics*, 20(6):1709–1719, 1993.
_eprint: <https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.596958>.
- [6] C. X. Yu. Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy. *Physics in Medicine and Biology*, 40(9):1435–1449, Sept. 1995.
- [7] M. Teoh, C. H. Clark, K. Wood, S. Whitaker, and A. Nisbet. Volumetric modulated arc therapy: a review of current literature and clinical use in

- practice. *The British Journal of Radiology*, 84(1007):967–996, Nov. 2011.
- [8] R. Miralbell, A. Lomax, L. Cella, and U. Schneider. Potential reduction of the incidence of radiation-induced second cancers by using proton beams in the treatment of pediatric tumors. *International Journal of Radiation Oncology*Biology*Physics*, 54(3):824–829, Nov. 2002.
- [9] B. C. Baumann, R. A. Lustig, S. Mazzoni, S. M. Grady, B. W. O'Malley, J. Y. K. Lee, J. G. Newman, J. M. Schuster, S. Both, A. Lin, J. F. Dorsey, and M. Alonso-Basanta. A prospective clinical trial of proton therapy for chordoma and chondrosarcoma: Feasibility assessment. *Journal of Surgical Oncology*, 120(2):200–205, Aug. 2019.
- [10] B. C. Baumann, N. Mitra, J. G. Harton, Y. Xiao, A. P. Wojcieszynski, P. E. Gabriel, H. Zhong, H. Geng, A. Doucette, J. Wei, P. J. O'Dwyer, J. E. Bekelman, and J. M. Metz. Comparative Effectiveness of Proton vs Photon Therapy as Part of Concurrent Chemoradiotherapy for Locally Advanced Cancer. *JAMA Oncology*, 6(2):237, Feb. 2020.
- [11] E. Rutherford. LXXIX. *The scattering of alpha and beta particles by matter and the structure of the atom. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 21(125):669–688, May 1911.
- [12] M. Durante and H. Paganetti. Nuclear physics in particle therapy: a review. *Reports on Progress in Physics*, 79(9):096702, Sept. 2016.
- [13] S. Yan, T. A. Ngoma, W. Ngwa, and T. R. Bortfeld. Global democratisation of proton radiotherapy. *The Lancet Oncology*, 24(6):e245–e254, June 2023.
- [14] H. Paganetti. *Proton Therapy Physics, Second Edition*. Series in Medical Physics and Biomedical Engineering. CRC Press, 2018.
- [15] B. G. L. Vanneste, E. J. Van Limbergen, E. N. Van Lin, J. G. H. Van Roermond, and P. Lambin. Prostate Cancer Radiation Therapy: What Do Clinicians Have to Know? *BioMed Research International*, 2016:1–14, 2016.
- [16] W. Gu, D. Ruan, Q. Lyu, W. Zou, L. Dong, and K. Sheng. A novel energy layer optimization framework for spot-scanning proton arc therapy. *Medical Physics*, 47(5):2072–2084, 2020. Number: 5 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/mp.14083>.
- [17] X. Ding, X. Li, J. M. Zhang, P. Kabolizadeh, C. Stevens, and D. Yan. Spot-

- Scanning Proton Arc (SPArc) Therapy: The First Robust and Delivery-Efficient Spot-Scanning Proton Arc Therapy. *International Journal of Radiation Oncology*Biology*Physics*, 96(5):1107–1116, Dec. 2016. Number: 5.
- [18] G. Liu, X. Li, L. Zhao, W. Zheng, A. Qin, S. Zhang, C. Stevens, D. Yan, P. Kabolizadeh, and X. Ding. A novel energy sequence optimization algorithm for efficient spot-scanning proton arc (SPArc) treatment delivery. *Acta Oncologica*, 59(10):1178–1185, Oct. 2020. Number: 10.
- [19] S. Wuyckens, D. Dasnoy, G. Janssens, V. Hamaide, M. Huet, E. Loÿen, G. R. de Hertaing, B. Macq, E. Sterpin, J. A. Lee, K. Souris, and S. Deffet. OpenTPS – Open-source treatment planning system for research in proton therapy. 2023. Publisher: arXiv Version Number: 1.
- [20] A. Brahme. Treatment Optimization Using Physical and Radiobiological Objective Functions. In A. R. Smith, editor, *Radiation Therapy Physics*, pages 209–246. Springer Berlin Heidelberg, Berlin, Heidelberg, 1995.
- [21] D. Craft, T. Halabi, and T. Bortfeld. Exploration of tradeoffs in intensity-modulated radiotherapy. *Physics in Medicine and Biology*, 50(24):5857–5868, Dec. 2005. Number: 24.
- [22] D. Craft, T. Halabi, H. A. Shih, and T. Bortfeld. An Approach for Practical Multiobjective IMRT Treatment Planning. *International Journal of Radiation Oncology*Biology*Physics*, 69(5):1600–1607, Dec. 2007.
- [23] D. L. Craft, T. S. Hong, H. A. Shih, and T. R. Bortfeld. Improved Planning Time and Plan Quality Through Multicriteria Optimization for Intensity-Modulated Radiotherapy. *International Journal of Radiation Oncology, Biology, Physics*, 82(1):e83–e90, Jan. 2012. Publisher: Elsevier.
- [24] U. Oelfke and T. Bortfeld. Inverse planning for photon and proton beams. *Medical Dosimetry*, 26(2):113–124, 2001.
- [25] D. L. Craft, T. F. Halabi, H. A. Shih, and T. R. Bortfeld. Approximating convex Pareto surfaces in multiobjective radiotherapy planning. *Medical Physics*, 33(9):3399–3407, 2006. _eprint: <https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.2335486>.
- [26] E. Engwall, C. Battinelli, V. Wase, O. Marthin, L. Glimelius, R. Bokrantz, B. Andersson, and A. Fredriksson. Fast robust optimization of proton PBS

- arc therapy plans using early energy layer selection and spot assignment. *Physics in Medicine & Biology*, 67(6):065010, Mar. 2022. Number: 6.
- [27] W. H. Organization. Ionizing radiation, health effects and protective measures, Mar. 2016.
- [28] B. G. C. V. Wouters, Marianne Koritzinsky. Irradiation-induced damage and the DNA damage response. In *Basic Clinical Radiobiology*. CRC Press, 5 edition, 2018. Num Pages: 12.
- [29] J. H. Folley, W. Borges, and T. Yamawaki. Incidence of leukemia in survivors of the atomic bomb in Hiroshima and Nagasaki, Japan. *The American Journal of Medicine*, 13(3):311–321, Sept. 1952.
- [30] G. Calais, M. Alfonsi, E. Bardet, C. Sire, T. Germain, P. Bergerot, B. Rhein, J. Tortochaux, P. Oudinot, and P. Bertrand. Randomized Trial of Radiation Therapy Versus Concomitant Chemotherapy and Radiation Therapy for Advanced-Stage Oropharynx Carcinoma. *JNCI: Journal of the National Cancer Institute*, 91(24):2081–2086, Dec. 1999.
- [31] J. Green, J. Kirwan, J. Tierney, P. Symonds, L. Fresco, C. Williams, and M. Collingwood. Concomitant chemotherapy and radiation therapy for cancer of the uterine cervix. In The Cochrane Collaboration, editor, *The Cochrane Database of Systematic Reviews*, page CD002225. John Wiley & Sons, Ltd, Chichester, UK, Oct. 2001.
- [32] J. S. Cooper, T. F. Pajak, A. A. Forastiere, J. Jacobs, B. H. Campbell, S. B. Saxman, J. A. Kish, H. E. Kim, A. J. Cmelak, M. Rotman, M. Machtay, J. F. Ensley, K. C. Chao, C. J. Schultz, N. Lee, and K. K. Fu. Postoperative Concurrent Radiotherapy and Chemotherapy for High-Risk Squamous-Cell Carcinoma of the Head and Neck. *New England Journal of Medicine*, 350(19):1937–1944, May 2004.
- [33] S. Y. Wu, L. Singer, L. Boreta, M. A. Garcia, S. E. Fogh, and S. E. Braunstein. Palliative radiotherapy near the end of life. *BMC Palliative Care*, 18(1):29, Dec. 2019.
- [34] M. E. Ray and M. S. Sabel. Principles of Radiation Therapy. In *Essentials of Surgical Oncology*, pages 65–79. Elsevier, 2007.
- [35] P. Rancoita and C. Leroy. *Principles Of Radiation Interaction In Matter And Detection (4th Edition)*. World Scientific Publishing Company, 2015.

- [36] M. C. Joiner and A. J. v. d. Kogel. *Basic Clinical Radiobiology*. CRC Press, Aug. 2018. Google-Books-ID: kF0PEAAAQBAJ.
- [37] R. R. Wilson. Radiological Use of Fast Protons. *Radiology*, 47(5):487–491, 1946.
- [38] B. Marchand, D. Prieels, B. Bauvir, C. Cyclotron, R. Sepulchre, and M. Gérard. IBA proton pencil beam scanning: An innovative solution for cancer treatment. *Proc. 7th EPAC (Vienna)*, Jan. 2000.
- [39] C. Bula, D. Meer, and E. Pedroni. System for the delivery of proton therapy by pencil beam scanning of a predeterminable volume within a patient, Apr. 2015.
- [40] S. E. Combs and J. Debus. Treatment with heavy charged particles: Systematic review of clinical data and current clinical (comparative) trials. *Acta Oncologica*, 52(7):1272–1286, Oct. 2013.
- [41] T. Kamada, H. Tsujii, E. A. Blakely, J. Debus, W. De Neve, M. Durante, O. Jäkel, R. Mayer, R. Orecchia, R. Pötter, S. Vatnitsky, and W. T. Chu. Carbon ion radiotherapy in Japan: an assessment of 20 years of clinical experience. *The Lancet Oncology*, 16(2):e93–e100, Feb. 2015.
- [42] Y. Z. Szeto, M. G. Witte, S. R. Van Kranen, J.-J. Sonke, J. Belderbos, and M. Van Herk. Effects of anatomical changes on pencil beam scanning proton plans in locally advanced NSCLC patients. *Radiotherapy and Oncology*, 120(2):286–292, Aug. 2016.
- [43] E. J. Hall and A. J. Giaccia. *Radiobiology for the Radiologist*. Lippincott Williams & Wilkins, 2006. Google-Books-ID: 6HhJwRyqBzgC.
- [44] H. Paganetti, E. Blakely, A. Carabe-Fernandez, D. J. Carlson, I. J. Das, L. Dong, D. Grosshans, K. D. Held, R. Mohan, V. Moiseenko, A. Niemierko, R. D. Stewart, and H. Willers. Report of the AAPM TG -256 on the relative biological effectiveness of proton beams in radiation therapy. *Medical Physics*, 46(3), Mar. 2019.
- [45] F. Guan, L. Bronk, U. Titt, S. H. Lin, D. Mirkovic, M. D. Kerr, X. R. Zhu, J. Dinh, M. Sobieski, C. Stephan, C. R. Peeler, R. Taleei, R. Mohan, and D. R. Grosshans. Spatial mapping of the biologic effectiveness of scanned particle beams: towards biologically optimized particle therapy. *Scientific Reports*, 5(1):9850, May 2015.

- [46] C. R. Peeler, D. Mirkovic, U. Titt, P. Blanchard, J. R. Gunther, A. Mahajan, R. Mohan, and D. R. Grosshans. Clinical evidence of variable proton biological effectiveness in pediatric patients treated for ependymoma. *Radiotherapy and Oncology*, 121(3):395–401, Dec. 2016.
- [47] H. Paganetti. Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer. *Physics in Medicine and Biology*, 59(22):R419–R472, Nov. 2014.
- [48] R. Mohan, C. R. Peeler, F. Guan, L. Bronk, W. Cao, and D. R. Grosshans. Radiobiological issues in proton therapy. *Acta Oncologica*, 56(11):1367–1373, Nov. 2017.
- [49] D. Haas-Kogan, D. Indelicato, H. Paganetti, N. Esiashvili, A. Mahajan, T. Yock, S. Flampouri, S. MacDonald, M. Fouladi, K. Stephen, J. Kalapurakal, S. Terezakis, H. Kooy, D. Grosshans, M. Makrigiorgos, K. Mishra, T. Y. Poussaint, K. Cohen, T. Fitzgerald, V. Gondi, A. Liu, J. Michalski, D. Mirkovic, R. Mohan, S. Perkins, K. Wong, B. Vikram, J. Buchsbaum, and L. Kun. National Cancer Institute Workshop on Proton Therapy for Children: Considerations Regarding Brainstem Injury. *International Journal of Radiation Oncology*Biology*Physics*, 101(1):152–168, May 2018.
- [50] E. R. Watson, K. E. Hainan, S. Dische, M. I. Saunders, I. S. Cade, J. B. McEwen, G. Wiernik, D. J. D. Perrins, and I. Sutherland. Hyperbaric oxygen and radiotherapy: a Medical Research Council trial in carcinoma of the cervix. *The British Journal of Radiology*, 51(611):879–887, Nov. 1978.
- [51] C. M. Hoff, H. S. Hansen, M. Overgaard, C. Grau, J. Johansen, J. Bentzen, and J. Overgaard. The importance of haemoglobin level and effect of transfusion in HNSCC patients treated with radiotherapy – Results from the randomized DAHANCA 5 study. *Radiotherapy and Oncology*, 98(1):28–33, Jan. 2011.
- [52] J.-P. Pignon, A. L. Maître, E. Maillard, and J. Bourhis. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): An update on 93 randomised trials and 17,346 patients. *Radiotherapy and Oncology*, 92(1):4–14, July 2009.
- [53] J. A. Bonner, P. M. Harari, J. Giralt, R. B. Cohen, C. U. Jones, R. K. Sur, D. Raben, J. Baselga, S. A. Spencer, J. Zhu, H. Youssoufian, E. K. Rowinsky, and K. K. Ang. Radiotherapy plus cetuximab for locoregionally

- advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. *The Lancet Oncology*, 11(1):21–28, Jan. 2010.
- [54] K. K. Ang, Q. Zhang, D. I. Rosenthal, P. F. Nguyen-Tan, E. J. Sherman, R. S. Weber, J. M. Galvin, J. A. Bonner, J. Harris, A. K. El-Naggar, M. L. Gillison, R. C. Jordan, A. A. Konski, W. L. Thorstad, A. Trotti, J. J. Beitler, A. S. Garden, W. J. Spanos, S. S. Yom, and R. S. Axelrod. Randomized Phase III Trial of Concurrent Accelerated Radiation Plus Cisplatin With or Without Cetuximab for Stage III to IV Head and Neck Carcinoma: RTOG 0522. *Journal of Clinical Oncology*, 32(27):2940–2950, Sept. 2014.
- [55] M. R. Horsman, L. S. Mortensen, J. B. Petersen, M. Busk, and J. Overgaard. Imaging hypoxia to improve radiotherapy outcome. *Nature Reviews Clinical Oncology*, 9(12):674–687, Dec. 2012.
- [56] S. I. Chung, D. K. Smart, E. J. Chung, and D. E. Citrin. Radioprotection as a Method to Enhance the Therapeutic Ratio of Radiotherapy. In P. J. Tofilon and K. Camphausen, editors, *Increasing the Therapeutic Ratio of Radiotherapy*, pages 79–102. Springer International Publishing, Cham, 2017. Series Title: Cancer Drug Discovery and Development.
- [57] J. R. Hughes and J. L. Parsons. FLASH Radiotherapy: Current Knowledge and Future Insights Using Proton-Beam Therapy. *International Journal of Molecular Sciences*, 21(18):6492, Sept. 2020.
- [58] A. Peeters, J. P. Grutters, M. Pijls-Johannesma, S. Reimoser, D. De Ruyscher, J. L. Severens, M. A. Joore, and P. Lambin. How costly is particle therapy? Cost analysis of external beam radiotherapy with carbon-ions, protons and photons. *Radiotherapy and Oncology*, 95(1):45–53, Apr. 2010.
- [59] T. R. Bortfeld and J. S. Loeffler. Three ways to make proton therapy affordable. *Nature*, 549(7673):451–453, Sept. 2017.
- [60] J. M. Edmund and T. Nyholm. A review of substitute CT generation for MRI-only radiation therapy. *Radiation Oncology*, 12(1):28, Dec. 2017.
- [61] B. J. Pichler, A. Kolb, T. Nägele, and H.-P. Schlemmer. PET/MRI: Paving the Way for the Next Generation of Clinical Multimodality Imaging Applications. *Journal of Nuclear Medicine*, 51(3):333–336, Mar. 2010.
- [62] J.-P. Dillenseger, E. Moerschel, and C. Zorn. *Guide des technologies de*

- l'imagerie médicale et de la radiothérapie: Quand la théorie éclaire la pratique.* Elsevier Masson, Oct. 2016.
- [63] Y. Kwong, A. O. Mel, G. Wheeler, and J. M. Troupis. Four-dimensional computed tomography (4DCT): A review of the current status and applications: 4DCT review. *Journal of Medical Imaging and Radiation Oncology*, 59(5):545–554, Oct. 2015.
 - [64] J.-J. Sonke, M. Aznar, and C. Rasch. Adaptive Radiotherapy for Anatomical Changes. *Seminars in Radiation Oncology*, 29(3):245–257, July 2019.
 - [65] M. Van Herk, P. Remeijer, C. Rasch, and J. V. Lebesque. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, 47(4):1121–1135, July 2000.
 - [66] V. T. Taasti, L. Hong, J. S. Shim, J. O. Deasy, and M. Zarepisheh. Automating proton treatment planning with beam angle selection using Bayesian optimization. *Medical Physics*, 47(8):3286–3296, Aug. 2020.
 - [67] W. Cao, H. Rocha, R. Mohan, G. Lim, H. M. Goudarzi, B. C. Ferreira, and J. M. Dias. Reflections on beam configuration optimization for intensity-modulated proton therapy. *Physics in Medicine and Biology*, 67(13), July 2022. Publisher: IOP Publishing Ltd.
 - [68] U. Schneider, E. Pedroni, and A. Lomax. The calibration of CT Hounsfield units for radiotherapy treatment planning. *Physics in Medicine and Biology*, 41(1):111–124, Jan. 1996.
 - [69] L. Hong, M. Goitein, M. Bucciolini, R. Comiskey, B. Gottschalk, S. Rosenthal, C. Serago, and M. Urie. A pencil beam algorithm for proton dose calculations. *Physics in Medicine and Biology*, 41(8):1305–1330, Aug. 1996. Publisher: IOP Publishing.
 - [70] M. Fippel and M. Soukup. A Monte Carlo dose calculation algorithm for proton therapy. *Medical Physics*, 31(8):2263–2273, July 2004.
 - [71] X. Jia, J. Schümann, H. Paganetti, and S. B. Jiang. GPU-based fast Monte Carlo dose calculation for proton therapy. *Physics in Medicine and Biology*, 57(23):7783–7797, Dec. 2012.
 - [72] K. Souris, J. A. Lee, and E. Sterpin. Fast multi-purpose Monte Carlo

- simulation for proton therapy using multi- and many-core CPU architectures. *Medical Physics*, 1700:1–23, 2016.
- [73] A. Schiavi, M. Senzacqua, S. Pioli, A. Mairani, G. Magro, S. Molinelli, M. Ciocca, G. Battistoni, and V. Patera. Fred: a GPU-accelerated fast-Monte Carlo code for rapid treatment plan recalculation in ion beam therapy. *Physics in Medicine & Biology*, 62(18):7482–7504, Sept. 2017.
- [74] S. Huang, K. Souris, S. Li, M. Kang, A. M. Barragan Montero, G. Janssens, A. Lin, E. Garver, C. Ainsley, P. Taylor, Y. Xiao, and L. Lin. Validation and application of a fast Monte Carlo algorithm for assessing the clinical impact of approximations in analytical dose calculations for pencil beam scanning proton therapy. *Medical Physics*, 45(12):5631–5642, Dec. 2018.
- [75] F. Fracchiolla, E. Engwall, M. Janson, F. Tamm, S. Lorentini, F. Fellin, M. Bertolini, C. Algranati, R. Righetto, P. Farace, M. Amichetti, and M. Schwarz. Clinical validation of a GPU-based Monte Carlo dose engine of a commercial treatment planning system for pencil beam scanning proton therapy. *Physica Medica*, 88:226–234, Aug. 2021.
- [76] L. Engberg, A. Forsgren, K. Eriksson, and B. Hårdemark. Explicit optimization of plan quality measures in intensity-modulated radiation therapy treatment planning. *Medical Physics*, 44(6):2045–2053, June 2017.
- [77] H. Liu, Y. Chen, and B. Lu. A new inverse planning formalism with explicit DVH constraints and kurtosis-based dosimetric criteria. *Physics in Medicine & Biology*, 63(18):185015, Sept. 2018.
- [78] B. Schaffner and E. Pedroni. The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between CT-HU and proton stopping power. *Physics in Medicine and Biology*, 43(6):1579–1592, June 1998.
- [79] J. Unkelbach, T. Bortfeld, B. C. Martin, and M. Soukup. Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning. *Medical Physics*, 36(1):149–163, 2009. eprint: <https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.3021139>.
- [80] A. Fredriksson, A. Forsgren, and B. Hårdemark. Minimax optimization for handling range and setup uncertainties in proton therapy: Minimax optimization for handling uncertainties in proton therapy. *Medical Physics*, 38(3):1672–1684, Mar. 2011. Number: 3.

- [81] E. Sterpin, S. T. Rivas, F. V. d. Heuvel, B. George, J. A. Lee, and K. Souris. Development of robustness evaluation strategies for enabling statistically consistent reporting. *Physics in Medicine & Biology*, 66(4):045002, Feb. 2021. Publisher: IOP Publishing.
- [82] K. Souris, A. Barragan Montero, G. Janssens, D. Di Perri, E. Sterpin, and J. A. Lee. Technical Note: Monte Carlo methods to comprehensively evaluate the robustness of 4D treatments in proton therapy. *Medical Physics*, 46(10):4676–4684, 2019. eprint: <https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.13749>.
- [83] L. B. Marks, E. D. Yorke, A. Jackson, R. K. Ten Haken, L. S. Constine, A. Eisbruch, S. M. Bentzen, J. Nam, and J. O. Deasy. Use of Normal Tissue Complication Probability Models in the Clinic. *International Journal of Radiation Oncology*Biology*Physics*, 76(3):S10–S19, Mar. 2010.
- [84] B. Arjomandy, N. Sahoo, X. Ding, and M. Gillin. Use of a two-dimensional ionization chamber array for proton therapy beam quality assurance: 2D IC array for proton beam QA. *Medical Physics*, 35(9):3889–3894, Aug. 2008.
- [85] L. Lin, M. Kang, T. D. Solberg, T. Mertens, C. Baumer, C. G. Ainsley, and J. E. McDonough. Use of a novel two-dimensional ionization chamber array for pencil beam scanning proton therapy beam quality assurance. *Journal of Applied Clinical Medical Physics*, 16(3):270–276, May 2015.
- [86] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy. A technique for the quantitative evaluation of dose distributions. *Medical Physics*, 25(5):656–661, May 1998.
- [87] L. T. Chan, Y. I. Tan, P. W. Tan, Y. F. Leong, J. S. Khor, M. W. Teh, J. F. L. Cruz, S. P. Baggarley, K. H. Ooi, and Y. H. Leong. Comparing log file to measurement-based patient-specific quality assurance. *Physical and Engineering Sciences in Medicine*, 46(1):303–311, Mar. 2023.
- [88] W. Hsi, Y. Kang, X. Ding, J. Iannello, A. Mascia, N. Schreuder, J. Smith, O. Zeidan, and Y. Zheng. MO-A-213AB-03: Commissioning of a Clinical Chair for Patients Treated in the Seated Position Using an Inclined Beam Line Treatment Room. *Medical Physics*, 39(6Part20):3859–3859, June 2012.
- [89] L. Volz, Y. Sheng, M. Durante, and C. Graeff. Considerations for Upright Particle Therapy Patient Positioning and Associated Image Guidance. *Front-*

- tiers in Oncology*, 12:930850, July 2022.
- [90] A.-C. Knopf, T. S. Hong, and A. Lomax. Scanned proton radiotherapy for mobile targets—the effectiveness of re-scanning in the context of different treatment planning approaches and for different motion characteristics. *Physics in Medicine and Biology*, 56(22):7257–7271, Nov. 2011.
- [91] L. Lin, K. Souris, M. Kang, A. Glick, H. Lin, S. Huang, K. Stützer, G. Janssens, E. Sterpin, J. A. Lee, T. D. Solberg, J. E. McDonough, C. B. Simone, and E. Ben-Josef. Evaluation of motion mitigation using abdominal compression in the clinical implementation of pencil beam scanning proton therapy of liver tumors. *Medical Physics*, 44(2):703–712, Feb. 2017.
- [92] V. Favaudon, L. Caplier, V. Monceau, F. Pouzoulet, M. Sayarath, C. Fouillade, M.-F. Poupon, I. Brito, P. Hupé, J. Bourhis, J. Hall, J.-J. Fontaine, and M.-C. Vozenin. Ultrahigh dose-rate FLASH irradiation increases the differential response between normal and tumor tissue in mice. *Science Translational Medicine*, 6(245), July 2014.
- [93] S. Jolly, H. Owen, M. Schippers, and C. Welsch. Technical challenges for FLASH proton therapy. *Physica Medica*, 78:71–82, Oct. 2020.
- [94] B. Rothwell, M. Lowe, E. Traneus, M. Krieger, and J. Schuemann. Treatment planning considerations for the development of FLASH proton therapy. *Radiotherapy and Oncology*, 175:222–230, Oct. 2022.
- [95] S. Deffet, K. Souris, and E. Sterpin. Optimization of patient-specific range modulators for conformal FLASH proton therapy. 2023. Publisher: arXiv Version Number: 1.
- [96] Y. Simeonov, U. Weber, C. Schuy, R. Engenhardt-Cabillic, P. Penchev, V. Flatten, and K. Zink. Development, Monte Carlo simulations and experimental evaluation of a 3D range-modulator for a complex target in scanned proton therapy. *Biomedical Physics & Engineering Express*, 8(3):035006, May 2022.
- [97] M. Kang, X. Ding, and Y. Rong. FLASH instead of proton arc therapy is a more promising advancement for the next generation proton radiotherapy. *Journal of Applied Clinical Medical Physics*, 24(8):e14091, Aug. 2023.
- [98] Quasi-Newton Methods. In *Numerical Optimization*, pages 135–163. Springer New York, 2006. Series Title: Springer Series in Operations Re-

search and Financial Engineering.

- [99] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, Mass, 2nd edition edition, Sept. 1999.
- [100] N. Parikh. Proximal Algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.
- [101] R. Tibshirani. Regression Shrinkage and Selection via The Lasso: A Retrospective. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(3):273–282, June 2011.
- [102] P. L. Combettes and V. R. Wajs. Signal Recovery by Proximal Forward-Backward Splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, Jan. 2005.
- [103] N. G. Polson, J. G. Scott, and B. T. Willard. Proximal Algorithms in Statistics and Machine Learning, May 2015. arXiv:1502.03175 [cs, stat].
- [104] A. Chambolle, R. De Vore, Nam-Yong Lee, and B. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3):319–335, Mar. 1998.
- [105] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. 2003. Publisher: arXiv Version Number: 2.
- [106] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. Publisher: SIAM.
- [107] M. Sniedovich. *Dynamic Programming: Foundations and Principles, Second Edition*. CRC Press, Sept. 2010.
- [108] E. L. Lawler and D. E. Wood. Branch-and-Bound Methods: A Survey. *Operations Research*, 14(4):699–719, 1966. _eprint: <https://doi.org/10.1287/opre.14.4.699>.
- [109] P. Siarry, editor. *Metaheuristics*. Springer International Publishing, Cham, 2016.
- [110] R. J. Vanderbei. *Linear Programming: Foundations and Extensions*, volume

- 196 of *International Series in Operations Research & Management Science*. Springer US, Boston, MA, 2014.
- [111] K. G. Murty. *Linear Programming*. Wiley, 1984. Google-Books-ID: ibQJ-vAEACAAJ.
- [112] Sequential Quadratic Programming. In *Numerical Optimization*, pages 529–562. Springer New York, 2006. Series Title: Springer Series in Operations Research and Financial Engineering.
- [113] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, Apr. 2002. Number: 2.
- [114] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
- [115] A. Corana, M. Marchesi, C. Martini, and S. Ridella. Minimizing multimodal functions of continuous variables with the “simulated annealing” algorithm—Corrigenda for this article is available here. *ACM Transactions on Mathematical Software*, 13(3):262–280, Sept. 1987.
- [116] S. Breedveld, D. Craft, R. v. Haveren, and B. Heijmen. Multi-criteria optimization and decision-making in radiotherapy. *European Journal of Operational Research*, 277(1):1–19, 2019.
- [117] Y. Haimes, L. Lasdon, and D. Wismer. On a Bicriterion Formulation of the Problems of Integrated System Identification and System Optimization. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-1(3):296–297, July 1971.
- [118] M. Schniederjans. *Goal Programming: Methodology and Applications: Methodology and Applications*. Springer Science & Business Media, Dec. 2012.
- [119] A. Abraham and R. Goldberg. *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*. Springer Science & Business Media, Sept. 2005.
- [120] I. Wegener. *Complexity Theory*. Springer-Verlag, Berlin/Heidelberg, 2005.
- [121] D. E. Knuth. *The Art of Computer Programming: Sorting and Searching, Volume 3*. Addison-Wesley Professional, Apr. 1998.

- [122] D. Jungnickel. *Graphs, Networks and Algorithms*. Springer Science & Business Media, Sept. 2007.
- [123] O. Goldreich. *P, NP, and NP-Completeness: The Basics of Computational Complexity*. Cambridge University Press, Aug. 2010.
- [124] K. Otto. Volumetric modulated arc therapy: IMRT in a single gantry arc: Single arc radiation therapy. *Medical Physics*, 35(1):310–317, Dec. 2007. Number: 1.
- [125] C. C. Ling, P. Zhang, Y. Archambault, J. Bocanek, G. Tang, and T. LoSasso. Commissioning and Quality Assurance of RapidArc Radiotherapy Delivery System. *International Journal of Radiation Oncology*Biology*Physics*, 72(2):575–581, Oct. 2008.
- [126] J. Deasy, T. Mackie, and P. DeLuca. Conformal proton tomotherapy using distal-edge tracking. *Radiotherapy and Oncology*, 37(3):S43, 1995. Place: Ireland INIS Reference Number: 34044555.
- [127] G. Sandison, E. Papiez, C. Bloch, and J. Morphis. Phantom assessment of lung dose from proton arc therapy. *International Journal of Radiation Oncology*Biology*Physics*, 38(4):891–897, July 1997. Number: 4.
- [128] J. Seco, G. Gu, T. Marcelos, H. Kooy, and H. Willers. 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. *International Journal of Radiation Oncology*Biology*Physics*, 87(1):188–194, 2013.
- [129] D. Sánchez-Parcerisa, J. C. Pourbaix, C. G. Ainsley, D. Dolney, and A. Carabe. Fast range switching of passively scattered proton beams using a modulation wheel and dynamic beam current modulation. *Physics in Medicine and Biology*, 59(7):N19–N26, Apr. 2014.
- [130] X. Ding, A. Witztum, X. Liang, M. Reiche, H. Lin, B. Teo, L. Yin, J. Fiene, J. McDonough, and A. Kassaei. WE-F-16A-03: 3D Printer Application in Proton Therapy: A Novel Method to Deliver Passive-Scattering Proton Beams with a Fixed Range and Modulation for SRS and SRT. *Medical Physics*, 41(6Part30):514–514, June 2014.
- [131] W. Zou, T. Fisher, M. Zhang, L. Kim, T. Chen, V. Narra, B. Swann, R. Singh, R. Siderit, L. Yin, B.-k. K. Teo, M. McKenna, J. McDonough, and

- Y. J. Ning. Potential of 3D printing technologies for fabrication of electron bolus and proton compensators. *Journal of Applied Clinical Medical Physics*, 16(3):90–98, May 2015.
- [132] E. Pedroni, R. Bacher, H. Blattmann, T. Böhringer, A. Coray, A. Lomax, S. Lin, G. Munkel, S. Scheib, U. Schneider, and A. Tournovsky. The 200-MeV proton therapy project at the Paul Scherrer Institute: Conceptual design and practical realization. *Medical Physics*, 22(1):37–53, Jan. 1995.
- [133] E. Pedroni, S. Scheib, T. Böhringer, A. Coray, M. Grossmann, S. Lin, and A. Lomax. Experimental characterization and physical modelling of the dose distribution of scanned proton pencil beams. *Physics in Medicine and Biology*, 50(3):541–561, Feb. 2005.
- [134] G. Macchia, F. Deodato, S. Cilla, S. Cammelli, A. Guido, M. Ferioli, G. Siepe, V. Valentini, A. G. Morganti, and G. Ferrandina. Volumetric modulated arc therapy for treatment of solid tumors: current insights. *OncoTargets and Therapy*, Volume 10:3755–3772, July 2017.
- [135] S. O. Hunte, C. H. Clark, N. Zyuzikov, and A. Nisbet. Volumetric modulated arc therapy (VMAT): a review of clinical outcomes—what is the clinical evidence for the most effective implementation? *The British Journal of Radiology*, 95(1136):20201289, Aug. 2022.
- [136] P. Zhang, L. Happersett, M. Hunt, A. Jackson, M. Zelefsky, and G. Mageras. Volumetric Modulated Arc Therapy: Planning and Evaluation for Prostate Cancer Cases. *International Journal of Radiation Oncology*Biology*Physics*, 76(5):1456–1462, Apr. 2010.
- [137] F. Kjær-Kristoffersen, L. Ohlhues, J. Medin, and S. Korreman. RapidArc volumetric modulated therapy planning for prostate cancer patients. *Acta Oncologica*, 48(2):227–232, Jan. 2009.
- [138] S. Vieillot, D. Azria, C. Lemanski, C. L. Moscardo, S. Gourgou, J.-B. Dubois, N. Aillères, and P. Fenoglietto. Plan comparison of volumetric-modulated arc therapy (RapidArc) and conventional intensity-modulated radiation therapy (IMRT) in anal canal cancer. *Radiation Oncology*, 5(1):92, Dec. 2010.
- [139] W. Duthoy, W. De Gersem, K. Vergote, M. Coghe, T. Boterberg, Y. De Deene, C. De Wagter, S. Van Belle, and W. De Neve. Whole abdominopelvic radiotherapy (WAPRT) using intensity-modulated arc therapy

- (IMAT): first clinical experience. *International Journal of Radiation Oncology*Biophysics*, 57(4):1019–1032, Nov. 2003.
- [140] W. Duthoy, W. De Gersem, K. Vergote, T. Boterberg, C. Derie, P. Smeets, C. De Wagter, and W. De Neve. Clinical implementation of intensity-modulated arc therapy (IMAT) for rectal cancer. *International Journal of Radiation Oncology*Biophysics*, 60(3):794–806, Nov. 2004.
- [141] E. Wong, D. P. D'Souza, J. Z. Chen, M. Lock, G. Rodrigues, T. Coad, K. Trenka, M. Mulligan, and G. S. Bauman. Intensity-modulated arc therapy for treatment of high-risk endometrial malignancies. *International Journal of Radiation Oncology*Biophysics*, 61(3):830–841, Mar. 2005.
- [142] L. Cozzi, K. A. Dinshaw, S. K. Shrivastava, U. Mahantshetty, R. Engineer, D. D. Deshpande, S. Jamema, E. Vanetti, A. Clivio, G. Nicolini, and A. Fogliata. A treatment planning study comparing volumetric arc modulation with RapidArc and fixed field IMRT for cervix uteri radiotherapy. *Radiotherapy and Oncology*, 89(2):180–191, Nov. 2008.
- [143] H. Paganetti, A. Niemierko, M. Ancukiewicz, L. E. Gerweck, M. Goitein, J. S. Loeffler, and H. D. Suit. Relative biological effectiveness (RBE) values for proton beam therapy. *International Journal of Radiation Oncology*Biophysics*, 53(2):407–421, June 2002.
- [144] C. Battinelli. Proton Arc Therapy Optimization. Master's thesis, KTH, Optimization and Systems Theory, 2019.
- [145] X. Li, G. Liu, G. Janssens, O. D. Wilde, V. Bossier, X. Lerot, A. Pouppez, D. Yan, C. Stevens, P. Kabolizadeh, and X. Ding. The first prototype of spot-scanning proton arc treatment delivery. *Radiotherapy and Oncology*, 137:130–136, 2019.
- [146] G. Janssens, L. Jacques, J. O. D. Xivry, X. Geets, and B. MacQ. Diffeomorphic registration of images with variable contrast enhancement. *International Journal of Biomedical Imaging*, 2011. ISBN: 1687-4188 (ISSN)\r1687-4196 (e-ISSN).
- [147] J. Shen, E. Tryggestad, J. E. Younkin, S. R. Keole, K. M. Furutani, Y. Kang, M. G. Herman, and M. Bues. Technical Note: Using experimentally determined proton spot scanning timing parameters to accurately model beam delivery time. *Medical Physics*, 44(10):5081–5088, Oct. 2017.

- [148] T. Pfeiler, C. Bäumer, E. Engwall, D. Geismar, B. Spaan, and B. Timmermann. Experimental validation of a 4D dose calculation routine for pencil beam scanning proton therapy. *Zeitschrift für Medizinische Physik*, 28(2):121–133, Apr. 2018.
- [149] L. Zhao, G. Liu, W. Zheng, J. Shen, A. Lee, D. Yan, R. Deraniyagala, C. Stevens, X. Li, S. Tang, and X. Ding. Building a precise machine-specific time structure of the spot and energy delivery model for a cyclotron-based proton therapy system. *Physics in Medicine & Biology*, 67(1):01NT01, Jan. 2022. Number: 1.
- [150] L. Zhao, G. Liu, S. Chen, J. Shen, W. Zheng, A. Qin, D. Yan, X. Li, and X. Ding. Developing an accurate model of spot-scanning treatment delivery time and sequence for a compact superconducting synchrocyclotron proton therapy system. *Radiation Oncology*, 17(1), 2022.
- [151] S. Rana, J. Bennouna, A. N. Gutierrez, and A. B. Rosenfeld. Impact of magnetic field regulation in conjunction with the volumetric repainting technique on the spot positions and beam range in pencil beam scanning proton therapy. *Journal of Applied Clinical Medical Physics*, 21(11):124–131, Nov. 2020.
- [152] A. Carabe-Fernandez, M. Kirk, D. Sanchez-Parcerisa, M. Fager, B. Burgdorf, M. Stowe, and T. Solberg. SU-E-T-640: Proton Modulated Arc Therapy Using Scanned Pencil Beams. *Medical Physics*, 42(6Part22):3483–3483, June 2015.
- [153] D. Sanchez-Parcerisa, M. Kirk, M. Fager, B. Burgdorf, M. Stowe, T. Solberg, and A. Carabe. Range optimization for mono- and bi-energetic proton modulated arc therapy with pencil beam scanning. *Physics in Medicine and Biology*, 61(21):N565–N574, Nov. 2016. Number: 21.
- [154] U. Langner, J. Eley, M. Guerrero, K. Langen, S. Mossahebi, M. Zhu, W. D’Souza, and B. Yi. A method to deliver energy modulated planar proton arc therapy (EMPPAT). *Journal of Proton Therapy*, 3:1–12, 2017.
- [155] G. Zhang, H. Shen, Y. Lin, R. C. Chen, Y. Long, and H. Gao. Energy layer optimization via energy matrix regularization for proton spot-scanning arc therapy. *Medical Physics*, 49(9):5752–5762, 2022. Number: 9. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/mp.15836>.
- [156] G. Zhang, Y. Long, Y. Lin, R. C. Chen, and H. Gao. A treatment plan

- optimization method with direct minimization of number of energy jumps for proton arc therapy. *Physics in Medicine & Biology*, 68(8):085001, Apr. 2023. Number: 8.
- [157] B. A. De Jong, C. Battinelli, J. Free, D. Wagenaar, E. Engwall, G. Janssens, J. A. Langendijk, E. W. Korevaar, and S. Both. Spot scanning proton arc therapy reduces toxicity in oropharyngeal cancer patients. *Medical Physics*, 50(3):1305–1317, Mar. 2023. Number: 3.
- [158] E. Engwall, O. Marthin, V. Wase, J. Sundström, V. Mikhalev, B. A. De Jong, J. A. Langendijk, H. Melb  us, B. Andersson, E. W. Korevaar, S. Both, R. Bokrantz, L. Glimelius, and A. Fredriksson. Partitioning of discrete proton arcs into interlaced subplans can bring proton arc advances to existing proton facilities. *Medical Physics*, page mp.16617, July 2023.
- [159] W. Cao, Y. Li, X. Zhang, F. Poenisch, P. Yepes, N. Sahoo, D. Grosshans, S. McGovern, G. B. Gunn, S. J. Frank, and X. R. Zhu. Intensity modulated proton arc therapy via geometry-based energy selection for ependymoma. *Journal of Applied Clinical Medical Physics*, 24(7):e13954, July 2023. Number: 7.
- [160] A. Carabe, I. V. Karagounis, K. Huynh, A. Bertolet, N. Fran  ois, M. M. Kim, A. Maity, E. Abel, and R. Dale. Radiobiological effectiveness difference of proton arc beams versus conventional proton and photon beams. *Physics in Medicine & Biology*, 65(16):165002, Aug. 2020. Number: 16.
- [161] A. Bertolet and A. Carabe. Proton monoenergetic arc therapy (PMAT) to enhance LETd within the target. *Physics in Medicine & Biology*, 65(16):165006, Aug. 2020. Number: 16.
- [162] H. Henjum, J. Tjel  ta, L. F. Fj  ra, S. Pilskog, C. H. Stokkev  g, E. Lyngholm, A. H. Handeland, and K. S. Ytre-Hauge. Influence of beam pruning techniques on LET and RBE in proton arc therapy. *Frontiers in Oncology*, 13:1155310, Sept. 2023.
- [163] X. Li, X. Ding, W. Zheng, G. Liu, G. Janssens, K. Souris, A. M. Barrag  n-Montero, D. Yan, C. Stevens, and P. Kabolizadeh. Linear Energy Transfer Incorporated Spot-Scanning Proton Arc Therapy Optimization: A Feasibility Study. *Frontiers in Oncology*, 11:698537, July 2021.
- [164] L. Glimelius, O. Marthin, V. Wase, J. Sundstr  m, E. Engwall, A. Fredriksson, H. Melb  us, F. Tamm, B. Andersson, J.   den, and R. Bokrantz. Ro-

- bust LET optimization of proton arcs can substantially reduce high LET in critical structures, 2023. ESTRO 2023.
- [165] S. Mein, B. Kopp, T. Tessonnier, J. Liermann, A. Abdollahi, J. Debus, T. Haberer, and A. Mairani. Spot-scanning hadron arc (SHArc) therapy: A proof of concept using single- and multi-ion strategies with helium, carbon, oxygen, and neon ions. *Medical Physics*, 49(9):6082–6097, Sept. 2022.
- [166] S. Mein, T. Tessonnier, B. Kopp, C. Schömers, S. Harrabi, A. Abdollahi, J. Debus, T. Haberer, and A. Mairani. Biological Dose Optimization for Particle Arc Therapy Using Helium and Carbon Ions. *International Journal of Radiation Oncology*Biophysics*, 114(2):334–348, Oct. 2022.
- [167] S. Mein, T. Tessonnier, B. Kopp, S. Harrabi, A. Abdollahi, J. Debus, T. Haberer, and A. Mairani. Spot-Scanning Hadron Arc (SHArc) Therapy: A Study With Light and Heavy Ions. *Advances in Radiation Oncology*, 6(3):100661, May 2021.
- [168] R. T. Flynn, D. L. Barbee, T. R. Mackie, and R. Jeraj. Comparison of intensity modulated x-ray therapy and intensity modulated proton therapy for selective subvolume boosting: a phantom study. *Physics in Medicine and Biology*, 52(20):6073–6091, Oct. 2007.
- [169] L. A. Rechner, R. M. Howell, R. Zhang, C. Etzel, A. K. Lee, and W. D. Newhauser. Risk of radiogenic second cancers following volumetric modulated arc therapy and proton arc therapy for prostate cancer. *Physics in Medicine and Biology*, 57(21):7117–7132, Nov. 2012.
- [170] A. Hecht. SU-FF-T-136: Improved Dose Characteristics From Proton Arc Therapy. *Medical Physics*, 36(6Part10):2551–2551, June 2009.
- [171] J.-E. Rah, G.-Y. Kim, D. H. Oh, T. H. Kim, J. W. Kim, D. Y. Kim, S. Y. Park, and D. Shin. 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. *Radiation Oncology*, 11(1):140, Dec. 2016.
- [172] J. P. Blanco Kiely and B. M. White. Dosimetric feasibility of single-energy proton modulated arc therapy for treatment of chordoma at the skull base. *Acta Oncologica*, 55(9-10):1243–1245, Oct. 2016.
- [173] X. Ding, X. Li, A. Qin, J. Zhou, D. Yan, C. Stevens, D. Krauss, and P. Kab-

- olizadeh. 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 Oncologica*, 57(3):435–437, Mar. 2018. Number: 3.
- [174] X. Li, P. Kabolizadeh, D. Yan, A. Qin, J. Zhou, Y. Hong, T. Guerrero, I. Grills, C. Stevens, and X. Ding. Improve dosimetric outcome in stage III non-small-cell lung cancer treatment using spot-scanning proton arc (SPArc) therapy. *Radiation Oncology*, 13(1):35, Dec. 2018. Number: 1.
- [175] G. Liu, L. Zhao, A. Qin, I. Grills, R. Deraniyagala, C. Stevens, S. Zhang, D. Yan, X. Li, and X. Ding. Lung Stereotactic Body Radiotherapy (SBRT) Using Spot-Scanning Proton Arc (SPArc) Therapy: A Feasibility Study. *Frontiers in Oncology*, 11, 2021.
- [176] S. Chang, G. Liu, L. Zhao, J. T. Dilworth, W. Zheng, S. Jawad, D. Yan, P. Chen, C. Stevens, P. Kabolizadeh, X. Li, and X. Ding. Feasibility study: spot-scanning proton arc therapy (SPArc) for left-sided whole breast radiotherapy. *Radiation Oncology*, 15(1):232, Dec. 2020. Number: 1.
- [177] X. Ding, J. Zhou, X. Li, K. Blas, G. Liu, Y. Wang, A. Qin, P. Chinnaiyan, D. Yan, C. Stevens, I. Grills, and P. Kabolizadeh. Improving dosimetric outcome for hippocampus and cochlea sparing whole brain radiotherapy using spot-scanning proton arc therapy. *Acta Oncologica*, 58(4):483–490, Apr. 2019. Number: 4.
- [178] G. Liu, X. Li, A. Qin, J. Zhou, W. Zheng, L. Zhao, J. Han, S. Zhang, D. Yan, C. Stevens, I. Grills, and X. Ding. Is proton beam therapy ready for single fraction spine SBRS? – a feasibility study to use spot-scanning proton arc (SPArc) therapy to improve the robustness and dosimetric plan quality. *Acta Oncologica*, 60(5):653–657, May 2021. Number: 5.
- [179] G. Liu, X. Li, A. Qin, W. Zheng, D. Yan, S. Zhang, C. Stevens, P. Kabolizadeh, and X. Ding. Improve the dosimetric outcome in bilateral head and neck cancer (HNC) treatment using spot-scanning proton arc (SPArc) therapy: a feasibility study. *Radiation Oncology*, 15(1):21, Dec. 2020. Number: 1.
- [180] B. Yi, S. Mossahebi, A. Modiri, E. M. Nichols, M. Guerrero, N. Lamichhane, and P. Mohindra. Proton Arc Therapy vs Interstitial HDR Brachytherapy in Gynecologic Cancer with Parametrial/pelvic Side Wall Extension. *International Journal of Particle Therapy*, 9(2):31–39, Sept. 2022.

- [181] A. Modiri, S. Mossahebi, P. Mohindra, R. Miller, A. Sawant, S. Chen, and B. Yi. Enhanced Sparing of Organs at Risk with the Single-Energy-Modulated Proton Arc Therapy (SEM-PAT); A Proof of Concept in Central Lung and Pelvis . (Abstract). *Medical Physics*, (6):E765, 2020.
- [182] A. Mazal, J. A. Vera Sanchez, D. Sanchez-Parcerisa, J. M. Udias, S. España, V. Sanchez-Tembleque, L. M. Fraile, P. Bragado, A. Gutierrez-Uzquiza, N. Gordillo, G. Garcia, J. Castro Novais, J. M. Perez Moreno, L. Mayorga Ortiz, A. Ilundain Idoate, M. Cremades Sendino, C. Ares, R. Miralbell, and N. Schreuder. Biological and Mechanical Synergies to Deal With Proton Therapy Pitfalls: Minibeams, FLASH, Arcs, and Gantryless Rooms. *Frontiers in Oncology*, 10:613669, Jan. 2021.
- [183] A.-C. Wéra, M. Chocan Vera, H. Ozan, E. Engwall, V. Wase, O. Marthin, J. Sundström, S. Wuyckens, K. Haustermans, A. M. Barragán-Montero, K. Souris, J. A. Lee, and E. Sterpin. Proton arc therapy for esophageal cancer: is the dosimetric benefit worth the effort? *Submitted to Physica Medica*, 2023.
- [184] L. Toussaint, D. J. Indelicato, K. S. Holgersen, J. B. B. Petersen, C. H. Stokkevåg, Y. Lassen-Ramshad, O. Casares-Magaz, A. Vestergaard, and L. P. Muren. Towards proton arc therapy: physical and biologically equivalent doses with increasing number of beams in pediatric brain irradiation. *Acta Oncologica*, 58(10):1451–1456, Oct. 2019.
- [185] L. Toussaint, D. J. Indelicato, L. P. Muren, and C. H. Stokkevåg. Risk of second primary cancer from proton arc therapy of pediatric brain tumors. *Phys Imaging Radiat Oncol.*, (27):100480, Aug. 2023.
- [186] R. Argota-Perez, U. V. Elstrøm, K. Jensen, and S. S. Korreman. Robustness of multifield pseudo-arc proton therapy for sinonasal cancer. *Acta Oncologica*, 62(5):508–512, May 2023.
- [187] B. A. De Jong, E. W. Korevaar, A. Maring, C. I. Werkman, D. Scandurra, G. Janssens, S. Both, and J. A. Langendijk. Proton arc therapy increases the benefit of proton therapy for oropharyngeal cancer patients in the model based clinic. *Radiotherapy and Oncology*, 184:109670, July 2023.
- [188] J. A. Langendijk, P. Lambin, D. De Ruyscher, J. Widder, M. Bos, and M. Verheij. Selection of patients for radiotherapy with protons aiming at reduction of side effects: The model-based approach. *Radiotherapy and Oncology*, 107(3):267–273, June 2013. Number: 3.

- [189] S. Wuyckens. ArcPT Benchmark, 2022. Open Data @ UCLouvain.
- [190] T. Kanai, K. Kawachi, Y. Kumamoto, H. Ogawa, T. Yamada, H. Matsuzawa, and T. Inada. Spot scanning system for proton radiotherapy. *Medical Physics*, 7(4):365–369, 1980.
- [191] A. Lomax. Intensity modulation methods for proton radiotherapy. *Physics in Medicine and Biology*, 44(1):185–205, Jan. 1999. Publisher: IOP Publishing.
- [192] G. Liu, L. Zhao, W. Zheng, A. Qin, D. Yan, G. Janssens, J. Shen, C. Stevens, P. Kabolizadeh, and X. Ding. Develop An Accurate Model of Spot-Scanning Treatment Delivery Time for a Compact Superconducting Synchrocyclotron Proton System. July 2020.
- [193] P. Kalendralis, A. Traverso, Z. Shi, I. Zhovannik, R. Monshouwer, M. P. A. Starmans, S. Klein, E. Pfaehler, R. Boellaard, A. Dekker, and L. Wee. Multicenter CT phantoms public dataset for radiomics reproducibility tests. *Medical Physics*, 46(3):1512–1518, 2019.
- [194] M. u. Rehman, K. Erhart, J. Kielbasa, S. L. Meeks, Z. Li, T. Willoughby, N. Ramakrishna, K. Stephenson, T. S. Rahman, P. Kelly, and O. Zeidan. An optimized approach for robust spot placement in proton pencil beam scanning. *Physics in Medicine & Biology*, 64(23):235016, Dec. 2019. Publisher: IOP Publishing.
- [195] A. W. Lee, W. T. Ng, J. J. Pan, C.-L. Chiang, S. S. Poh, H. C. Choi, Y. C. Ahn, H. AlHussain, J. Corry, C. Grau, V. Grégoire, K. J. Harrington, C. S. Hu, D. L. Kwong, J. A. Langendijk, Q. T. Le, N. Y. Lee, J. C. Lin, T. X. Lu, W. M. Mendenhall, B. O’Sullivan, E. Ozyar, L. J. Peters, D. I. Rosenthal, G. Sanguineti, Y. L. Soong, Y. Tao, S. S. Yom, and J. T. Wee. International Guideline on Dose Prioritization and Acceptance Criteria in Radiation Therapy Planning for Nasopharyngeal Carcinoma. *International Journal of Radiation Oncology*Biophysics*, 105(3):567–580, 2019.
- [196] A. Beck. *First-Order Methods in Optimization*. SIAM-Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2017.
- [197] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani. A Sparse-Group Lasso. *Journal of Computational and Graphical Statistics*, 22(2):231–245, 2013. Publisher: Taylor & Francis.

- [198] M. Defferrard, R. Pena, and N. Perraudin. PyUNLocBoX: Optimization by Proximal Splitting.
- [199] H. Raguet, J. Fadili, and G. Peyré. A Generalized Forward-Backward Splitting. *SIAM Journal on Imaging Sciences*, 6(3):1199–1226, Jan. 2013.
- [200] C. Niculescu and L.-E. Persson. *Convex Functions and Their Applications: A Contemporary Approach*. Jan. 2006.
- [201] D. Pisinger and S. Ropke. Large neighborhood search. In *Handbook of metaheuristics*, pages 99–127. Springer, 2019.
- [202] F. A. Potra and S. J. Wright. Interior-point methods. *Journal of Computational and Applied Mathematics*, 124(1):281–302, 2000.
- [203] E. Shaw, R. Kline, M. Gillin, L. Souhami, A. Hirschfeld, R. Dinapoli, and L. Martin. Radiation therapy oncology group: Radiosurgery quality assurance guidelines. *International Journal of Radiation Oncology*Biophysics*, 27(5):1231–1239, 1993.
- [204] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing*, 16:1190–1208, Sept. 1995. Publisher: Society for Industrial and Applied Mathematics Publications.
- [205] G. Laporte and F. V. Louveaux. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations research letters*, 13(3):133–142, 1993. Publisher: Elsevier.
- [206] K.-H. Küfer, H. W. Hamacher, and T. Bortfeld. A multicriteria optimization approach for inverse radiotherapy planning. 2000.
- [207] Y. Yu, J. Zhang, G. Cheng, M. Schell, and P. Okunieff. Multi-objective optimization in radiotherapy: Applications to stereotactic radiosurgery and prostate brachytherapy. *Artificial intelligence in medicine*, 19:39–51, June 2000.
- [208] D. Craft, D. McQuaid, J. Wala, W. Chen, E. Salari, and T. Bortfeld. Multi-criteria VMAT optimization. *Medical Physics*, 39(2):686–696, 2012. eprint: <https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.3675601>.
- [209] G. K. Bahr, J. G. Kereiakes, H. Horwitz, R. Finney, J. Galvin, and K. Goode. The Method of Linear Programming Applied to Radia-

- tion Treatment Planning. *Radiology*, 91(4):686–693, 1968. _eprint: <https://doi.org/10.1148/91.4.686>.
- [210] T. Bortfeld. Optimized planning using physical objectives and constraints. *Seminars in Radiation Oncology*, 9(1):20–34, 1999.
- [211] H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, and A. Kumar. A New Linear Programming Approach to Radiation Therapy Treatment Planning Problems. *Operations Research*, 54(2):201–216, Apr. 2006.
- [212] S. Wuyckens, M. Saint-Guillain, G. Janssens, L. Zhao, X. Li, X. Ding, E. Sterpin, J. A. Lee, and K. Souris. Treatment planning in arc proton therapy: Comparison of several optimization problem statements and their corresponding solvers. *Computers in Biology and Medicine*, 148:105609, Sept. 2022.
- [213] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2022.
- [214] J. Clausen. Branch and Bound Algorithms-Principles and Examples. 2003.
- [215] D. R. Morrison, S. H. Jacobson, J. J. Sauppe, and E. C. Sewell. Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning. *Discrete Optimization*, 19:79–102, 2016.
- [216] R. Marsten, R. Subramanian, M. Saltzman, I. Lustig, and D. Shanno. Interior Point Methods for Linear Programming: Just Call Newton, Lagrange, and Fiacco and McCormick! *Interfaces*, 20(4):105–116, 1990. _eprint: <https://doi.org/10.1287/inte.20.4.105>.
- [217] R. S. Burachik, C. Y. Kaya, and M. M. Rizvi. Algorithms for Generating Pareto Fronts of Multi-objective Integer and Mixed-Integer Programming Problems, 2019.
- [218] F. Albertini, E. B. Hug, and A. J. Lomax. Is it necessary to plan with safety margins for actively scanned proton therapy? *Physics in Medicine and Biology*, 56(14):4399–4413, June 2011. Publisher: IOP Publishing.
- [219] A. Fredriksson. A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization. *Medical Physics*, 39(8):5169–5181, 2012. _eprint: <https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.4737113>.
- [220] S. Ahmed, A. Shapiro, and E. Shapiro. The sample average approxima-

- tion method for stochastic programs with integer recourse. *Submitted for publication*, pages 1–24, 2002.
- [221] H. Kahn and A. W. Marshall. Methods of reducing sample size in Monte Carlo computations. *Journal of the Operations Research Society of America*, 1(5):263–278, 1953. Publisher: INFORMS.
- [222] F. Biscani and D. Izzo. A parallel global multiobjective framework for optimization: pagmo. *Journal of Open Source Software*, 5(53):2338, Sept. 2020. Number: 53.
- [223] D. Izzo, M. Ruciński, and F. Biscani. The Generalized Island Model. In F. Fernández De Vega, J. I. Hidalgo Pérez, and J. Lanchares, editors, *Parallel Architectures and Bioinspired Algorithms*, volume 415, pages 151–169. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. Series Title: Studies in Computational Intelligence.
- [224] Qingfu Zhang and Hui Li. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, Dec. 2007. Number: 6.
- [225] Hui Li and Qingfu Zhang. Multiobjective Optimization Problems With Complicated Pareto Sets, MOEA/D and NSGA-II. *IEEE Transactions on Evolutionary Computation*, 13(2):284–302, Apr. 2009. Number: 2.
- [226] G. Acciarini, D. Izzo, and E. Mooij. MHACO: a Multi-Objective Hypervolume-Based Ant Colony Optimizer for Space Trajectory Optimization. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8, Glasgow, United Kingdom, July 2020. IEEE.
- [227] M. Lambrecht, D. B. Eekers, C. Alapetite, N. G. Burnet, V. Calugaru, I. E. Coremans, P. Fossati, M. Høyer, J. A. Langendijk, A. Méndez Romero, F. Paulsen, A. Perpar, L. Renard, D. De Ruyscher, B. Timmermann, P. Vitek, D. C. Weber, H. L. Van Der Weide, G. A. Whitfield, R. Wiggensraad, E. Roelofs, P. W. Nyström, and E. G. Troost. Radiation dose constraints for organs at risk in neuro-oncology; the European Particle Therapy Network consensus. *Radiotherapy and Oncology*, 128(1):26–36, July 2018. Number: 1.
- [228] N. J. Lomax and S. G. Scheib. Quantifying the degree of conformity in radiosurgery treatment planning. *International Journal of Radiation Oncology*Biophysics*, 55(5):1409–1419, Apr. 2003. Number: 5.

- [229] V. Wase, O. Marthin, A. Fredriksson, and A. Finnson. Optimizing the Traversal Time for Gantry Trajectories for Proton Arc Therapy Treatment Plans. 2023. Publisher: arXiv Version Number: 1.
- [230] M. Uto, T. Mizowaki, K. Ogura, and M. Hiraoka. Non-coplanar volumetric-modulated arc therapy (VMAT) for craniopharyngiomas reduces radiation doses to the bilateral hippocampus: a planning study comparing dynamic conformal arc therapy, coplanar VMAT, and non-coplanar VMAT. *Radiation Oncology*, 11(1):86, Dec. 2016. Number: 1.
- [231] A. Vaniqui, F. Vaassen, D. Di Perri, D. Eekers, I. Compter, I. Rinaldi, W. Van Elmpt, and M. Unipan. Linear Energy Transfer and Relative Biological Effectiveness Investigation of Various Structures for a Cohort of Proton Patients With Brain Tumors. *Advances in Radiation Oncology*, 8(2):101128, Mar. 2023. Number: 2.
- [232] L. Zhao, J. You, G. Liu, S. Wuyckens, X. Lu, and X. Ding. The first direct method of spot sparsity optimization for proton arc therapy. *Acta Oncologica*, 62(1):48–52, Jan. 2023. Number: 1.
- [233] S. Wuyckens, L. Zhao, M. Saint-Guillain, G. Janssens, E. Sterpin, K. Souris, X. Ding, and J. A. Lee. Bi-criteria Pareto optimization to balance irradiation time and dosimetric objectives in proton arc therapy. *Physics in Medicine & Biology*, 67(24):245017, Dec. 2022. Number: 24.
- [234] A. Fu, V. T. Taasti, and M. Zarepisheh. Simultaneous Reduction of Number of Spots and Energy Layers in Intensity Modulated Proton Therapy for Rapid Spot Scanning Delivery. 2023. Publisher: arXiv Version Number: 1.
- [235] N. P. Nguyen, C. Frank, C. C. Moltz, P. Vos, H. J. Smith, U. Karlsson, S. Dutta, A. Midyett, J. Barloon, and S. Sallah. Impact of dysphagia on quality of life after treatment of head-and-neck cancer. *International Journal of Radiation Oncology*Biophysics*, 61(3):772–778, Mar. 2005. Number: 3.
- [236] M. Nascimento, A. Farias, A. Carvalho, R. Albuquerque, L. Ribeiro, J. Leao, and I. Silva. Impact of xerostomia on the quality of life of patients submitted to head and neck radiotherapy. *Medicina Oral Patología Oral y Cirugía Bucal*, pages 0–0, 2019.
- [237] E. Borderias-Villarroel, M. Huet Dastarac, A. M. Barragán-Montero, R. He-lander, M. Holmstrom, X. Geets, and E. Sterpin. Machine learning-based

- automatic proton therapy planning: Impact of post-processing and dose-mimicking in plan robustness. *Medical Physics*, 50(7):4480–4490, July 2023. Number: 7.
- [238] L. Cozzi, A. Bolsi, G. Nicolini, and A. Fogliata. The Simultaneous Integrated Boost with proton beams in head and neck patients. *Zeitschrift für Medizinische Physik*, 14(3):180–188, 2004. Number: 3.
- [239] J. Langendijk, M. Hoogeman, R. Monshouwer, and M. Verheij. Indicatie Protocol Protonen Therapie (versie 2.2) HOOFD-HALSTU- MOREN. Technical report, 2019.
- [240] S. Van De Water, A. C. Kraan, S. Breedveld, W. Schillemans, D. N. Teguh, H. M. Kooy, T. M. Madden, B. J. M. Heijmen, and M. S. Hoogeman. Improved efficiency of multi-criteria IMPT treatment planning using iterative resampling of randomly placed pencil beams. *Physics in Medicine and Biology*, 58(19):6969–6983, Oct. 2013. Number: 19.
- [241] W. Cao, G. Lim, L. Liao, Y. Li, S. Jiang, X. Li, H. Li, K. Suzuki, X. R. Zhu, D. Gomez, and X. Zhang. Proton energy optimization and reduction for intensity-modulated proton therapy. *Physics in Medicine and Biology*, 59(21):6341–6354, Oct. 2014. Number: 21.
- [242] S. Van De Water, M. F. Belosi, F. Albertini, C. Winterhalter, D. C. Weber, and A. J. Lomax. Shortening delivery times for intensity-modulated proton therapy by reducing the number of proton spots: an experimental verification. *Physics in Medicine & Biology*, 65(9):095008, May 2020. Number: 9.
- [243] M. F. Jensen, L. Hoffmann, J. B. B. Petersen, D. S. Møller, and M. Alber. Energy layer optimization strategies for intensity-modulated proton therapy of lung cancer patients. *Medical Physics*, 45(10):4355–4363, Oct. 2018. Number: 10.
- [244] A. Gerbershagen, D. Meer, J. M. Schippers, and M. Seidel. A novel beam optics concept in a particle therapy gantry utilizing the advantages of superconducting magnets. *Zeitschrift für Medizinische Physik*, 26(3):224–237, Sept. 2016. Number: 3.
- [245] A. Gerbershagen, C. Calzolaio, D. Meer, S. Sanfilippo, and M. Schippers. The advantages and challenges of superconducting magnets in particle therapy. *Superconductor Science and Technology*, 29(8):083001, Aug. 2016.

Number: 8.

- [246] J. M. Schippers, D. Meer, and A. Gerbershagen. Particle therapy gantry with an energy degrader and an achromatic final bending system, 2017. EP 3167933 A1.
- [247] G. Fattori, Y. Zhang, D. Meer, D. C. Weber, A. J. Lomax, and S. Safai. The potential of Gantry beamline large momentum acceptance for real time tumour tracking in pencil beam scanning proton therapy. *Scientific Reports*, 10(1):15325, Sept. 2020. Number: 1.
- [248] V. Maradia, I. Colizzi, D. Meer, D. C. Weber, A. J. Lomax, O. Actis, and S. Psoroulas. Universal and dynamic ridge filter for pencil beam scanning particle therapy: a novel concept for ultra-fast treatment delivery. *Physics in Medicine & Biology*, 67(22):225005, Nov. 2022. Number: 22.
- [249] S. Wuyckens. Comparison of SPARC algorithm efficiency using beamlet-based and beamlet-free optimization schemes, 2023. ESTRO 2023.
- [250] R. M. Karp. Reducibility among Combinatorial Problems. In R. E. Miller, J. W. Thatcher, and J. D. Bohlinger, editors, *Complexity of Computer Computations: Proceedings of a symposium on the Complexity of Computer Computations, held March 20–22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, and sponsored by the Office of Naval Research, Mathematics Program, IBM World Trade Corporation, and the IBM Research Mathematical Sciences Department*, The IBM Research Symposia Series, pages 85–103. Springer US, Boston, MA, 1972.
- [251] M. Scholz, A. M. Kellerer, W. Kraft-Weyrather, and G. Kraft. Computation of cell survival in heavy ion beams for therapy. *Radiation and Environmental Biophysics*, 36(1):59–66, Mar. 1997.
- [252] M. Krämer, O. Jäkel, T. Haberer, G. Kraft, D. Schardt, and U. Weber. Treatment planning for heavy-ion radiotherapy: physical beam model and dose optimization. *Physics in Medicine and Biology*, 45(11):3299–3317, Nov. 2000.
- [253] A. S. Azizi Sultan. *Optimization of Beam Orientation in Intensity Modulated Radiation Therapy Planning*. PhD thesis, Technische Universität Kaiserslautern, 2006.
- [254] D. Baatar, M. Ehrgott, H. W. Hamacher, and I. M. Raschendorfer. Mini-

- mizing the number of apertures in multileaf collimator sequencing with field splitting. *Discrete Applied Mathematics*, 250:87–103, Dec. 2018.
- [255] K.-H. Küfer, A. Scherrer, M. Monz, F. Alonso, H. Trinkaus, T. Bortfeld, and C. Thieke. Intensity-modulated radiotherapy - a large scale multi-criteria programming problem. *OR Spectrum*, 25(2):223–249, May 2003.
- [256] R. G. Kierkels, R. Visser, H. P. Bijl, J. A. Langendijk, A. A. Van 'T Veld, R. J. Steenbakkers, and E. W. Korevaar. Multicriteria optimization enables less experienced planners to efficiently produce high quality treatment plans in head and neck cancer radiotherapy. *Radiation Oncology*, 10(1):87, Dec. 2015.
- [257] S. C. Kamran, B. S. Mueller, P. Paetzold, J. Dunlap, A. Niemierko, T. Bortfeld, H. Willers, and D. Craft. Multi-criteria optimization achieves superior normal tissue sparing in a planning study of intensity-modulated radiation therapy for RTOG 1308-eligible non-small cell lung cancer patients. *Radiotherapy and Oncology*, 118(3):515–520, Mar. 2016. Number: 3.
- [258] T. S. Hong, D. L. Craft, F. Carlsson, and T. R. Bortfeld. Multicriteria Optimization in Intensity-Modulated Radiation Therapy Treatment Planning for Locally Advanced Cancer of the Pancreatic Head. *International Journal of Radiation Oncology*Biology*Physics*, 72(4):1208–1214, Nov. 2008.
- [259] D. Craft. Multi-criteria optimization methods in radiation therapy planning: a review of technologies and directions. 2013. Publisher: arXiv Version Number: 1.
- [260] T. R. Mackie. History of tomotherapy. *Physics in Medicine and Biology*, 51(13):R427–R453, July 2006.
- [261] L. Perrier, M. Morelle, P. Pommier, P. Boisselier, B. Coche-Dequeant, O. Gallocher, M. Alfonsi, E. Bardet, M. Rives, V. Calugaru, E. Chajon, G. Noel, H. Mecellem, D. Pérol, S. Dussart, and P. Giraud. Cost Analysis of Complex Radiation Therapy for Patients With Head and Neck Cancer. *International Journal of Radiation Oncology*Biology*Physics*, 95(2):654–662, June 2016.
- [262] D. Pross, S. Wuyckens, S. Deffet, E. Sterpin, J. A. Lee, and K. Souris. Beamlet-free optimization for Monte-Carlo-based treatment planning in proton therapy. *Medical Physics*, page mp.16813, Nov. 2023.

- [263] Y. Nesterov and V. Spokoiny. Random Gradient-Free Minimization of Convex Functions. *Foundations of Computational Mathematics*, 17(2):527–566, Apr. 2017.
- [264] E. Pedroni, R. Bearpark, T. Böhringer, A. Coray, J. Duppich, S. Forss, D. George, M. Grossmann, G. Goitein, C. Hilbes, M. Jermann, S. Lin, A. Lomax, M. Negrazus, M. Schippers, and G. Kotrle. The PSI Gantry 2: a second generation proton scanning gantry. *Zeitschrift für Medizinische Physik*, 14(1):25–34, 2004.
- [265] R. Pidikiti, B. C. Patel, M. R. Maynard, J. P. Dugas, J. Syh, N. Sahoo, H. T. Wu, and L. R. Rosen. Commissioning of the world’s first compact pencil-beam scanning proton therapy system. *Journal of Applied Clinical Medical Physics*, 19(1):94–105, Jan. 2018.
- [266] O. Actis, A. Mayor, D. Meer, U. Rechsteiner, A. Bolsi, A. J. Lomax, and D. C. Weber. A bi-directional beam-line energy ramping for efficient patient treatment with scanned proton therapy. *Physics in Medicine & Biology*, 68(17):175001, Sept. 2023.
- [267] V. Maradia, D. Meer, R. Dölling, D. C. Weber, A. J. Lomax, and S. Psoroulas. Demonstration of momentum cooling to enhance the potential of cancer treatment with proton therapy. *Nature Physics*, July 2023.
- [268] A. C. Giovannelli, V. Maradia, D. Meer, S. Safai, S. Psoroulas, M. Togno, C. Bula, D. C. Weber, A. J. Lomax, and G. Fattori. Beam properties within the momentum acceptance of a clinical gantry beamline for proton therapy. *Medical Physics*, 49(3):1417–1431, Mar. 2022. Number: 3.
- [269] A. Smith, M. Gillin, M. Bues, X. R. Zhu, K. Suzuki, R. Mohan, S. Woo, A. Lee, R. Komaki, J. Cox, K. Hiramoto, H. Akiyama, T. Ishida, T. Sasaki, and K. Matsuda. The M. D. Anderson proton therapy system. *Medical Physics*, 36(9Part1):4068–4083, Sept. 2009.
- [270] J. E. Younkin, M. Bues, T. T. Sio, W. Liu, X. Ding, S. R. Keole, J. B. Stoker, and J. Shen. Multiple energy extraction reduces beam delivery time for a synchrotron-based proton spot-scanning system. *Advances in Radiation Oncology*, 3(3):412–420, July 2018.
- [271] Y. Iwata, T. Kadowaki, H. Uchiyama, T. Fujimoto, E. Takada, T. Shirai, T. Furukawa, K. Mizushima, E. Takeshita, K. Katagiri, S. Sato, Y. Sano, and K. Noda. Multiple-energy operation with extended flattops at HIMAC.

- Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 624(1):33–38, Dec. 2010.
- [272] K. Mizushima, K. Katagiri, Y. Iwata, T. Furukawa, T. Fujimoto, S. Sato, Y. Hara, T. Shirai, and K. Noda. Experimental studies of systematic multiple-energy operation at HIMAC synchrotron. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 331:243–247, July 2014.
- [273] N. Saito, C. Bert, N. Chaudhri, A. Gemmel, D. Schardt, M. Durante, and E. Rietzel. Speed and accuracy of a beam tracking system for treatment of moving targets with scanned ion beams. *Physics in Medicine and Biology*, 54(16):4849–4862, Aug. 2009.
- [274] K. P. Nesteruk, A. Bolsi, A. J. Lomax, D. Meer, S. Van De Water, and J. M. Schippers. A static beam delivery device for fast scanning proton arc-therapy. *Physics in Medicine & Biology*, 66(5):055018, Mar. 2021.
- [275] S. Rahim, J. Korte, N. Hardcastle, S. Hegarty, T. Kron, and S. Everitt. Upright Radiation Therapy—A Historical Reflection and Opportunities for Future Applications. *Frontiers in Oncology*, 10:213, Feb. 2020.
- [276] A. N. Schreuder. *Technological developments allowing for the widespread clinical adoption of proton radiotherapy*. PhD thesis, University College London, London, United Kingdom, 2020.
- [277] M. Durante, J. Debus, and J. S. Loeffler. Physics and biomedical challenges of cancer therapy with accelerated heavy ions. *Nature Reviews Physics*, 3(12):777–790, Sept. 2021.
- [278] O. Marthin, V. Wase, R. Bokrantz, L. Glimelius, B. Andersson, A. Fredriksson, B. A. De Jong, E. W. Korevaar, S. Both, and E. Engwall. Robustness in proton arc treatments for head and neck cancer patients: Impact of gantry angle spacing and number of revolutions, 2022.
- [279] B. A. De Jong, L. Zhao, V. Wase, O. Marthin, X. Ding, E. Engwall, E. W. Korevaar, J. A. Langendijk, and S. Both. Comparison of static and dynamic proton arc treatment planning algorithms in oropharyngeal cancer, 2023.
- [280] G. Divine, H. J. Norton, R. Hunt, and J. Dienemann. A Review of Analysis and Sample Size Calculation Considerations for Wilcoxon Tests. *Anesthesia*

- & *Analgesia*, 117(3):699–710, Sept. 2013.
- [281] F. Albertini, E. B. Hug, and A. J. Lomax. The influence of the optimization starting conditions on the robustness of intensity-modulated proton therapy plans. *Physics in Medicine and Biology*, 55(10):2863–2878, May 2010.
- [282] H. Paganetti, C. Beltran, S. Both, L. Dong, J. Flanz, K. Furutani, C. Grassberger, D. R. Grosshans, A.-C. Knopf, J. A. Langendijk, H. Nystrom, K. Parodi, B. W. Raaymakers, C. Richter, G. O. Sawakuchi, M. Schippers, S. F. Shaitelman, B. K. K. Teo, J. Unkelbach, P. Wohlfahrt, and T. Lomax. Roadmap: proton therapy physics and biology. *Physics in Medicine & Biology*, 66(5):05RM01, Mar. 2021.
- [283] A. L. McNamara, J. Schuemann, and H. Paganetti. A phenomenological relative biological effectiveness (RBE) model for proton therapy based on all published *in vitro* cell survival data. *Physics in Medicine and Biology*, 60(21):8399–8416, Nov. 2015.
- [284] K. R. Tringale, D. L. Casey, G. Niyazov, J. A. Lavery, C. Moskowitz, D. N. Friedman, and S. L. Wolden. Second cancer risk in childhood cancer survivors treated with intensity-modulated radiation therapy: An updated analysis of more than 10 years of follow-up. *Pediatric Blood & Cancer*, 69(5):e29600, May 2022.
- [285] C. H. Stokkevåg, U. Schneider, L. P. Muren, and W. Newhauser. Radiation-induced cancer risk predictions in proton and heavy ion radiotherapy. *Physica Medica*, 42:259–262, Oct. 2017.
- [286] H. Engels and A. Wambersie. Relative Biological Effectiveness of Neutrons for Cancer Induction and Other Late Effects: A Review of Radiobiological Data. In P. M. Schlag, H.-J. Senn, V. Diehl, D. M. Parkin, M. F. Rajewsky, R. Rubens, M. Wannenmacher, R. Engenhart-Cabillic, and A. Wambersie, editors, *Fast Neutrons and High-LET Particles in Cancer Therapy*, volume 150, pages 54–87. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998. Series Title: Recent Results in Cancer Research.
- [287] P. Karlsson, E. Holmberg, A. Samuelsson, K.-A. Johansson, and A. Wallgren. Soft tissue sarcoma after treatment for breast cancer—A Swedish population-based study. *European Journal of Cancer*, 34(13):2068–2075, Dec. 1998.
- [288] E. J. Hall and C.-S. Wu. Radiation-induced second cancers: the im-

- pact of 3D-CRT and IMRT. *International Journal of Radiation Oncology*Biology*Physics*, 56(1):83–88, May 2003.
- [289] F. Nguyen, C. Rubino, S. Guerin, I. Diallo, A. Samand, M. Hawkins, O. Oberlin, D. Lefkopoulos, and F. De Vathaire. Risk of a Second Malignant Neoplasm After Cancer in Childhood Treated With Radiotherapy: Correlation With the Integral Dose Restricted to the Irradiated Fields. *International Journal of Radiation Oncology*Biology*Physics*, 70(3):908–915, Mar. 2008.
- [290] O. A. Trowell. The sensitivity of lymphocytes to ionising radiation. *The Journal of Pathology and Bacteriology*, 64(4):687–704, Oct. 1952.
- [291] X. Wang, P. Wang, Z. Zhao, Q. Mao, J. Yu, and M. Li. A review of radiation-induced lymphopenia in patients with esophageal cancer: an immunological perspective for radiotherapy. *Therapeutic Advances in Medical Oncology*, 12:175883592092682, Jan. 2020.
- [292] S. Rudra, C. Hui, Y. J. Rao, P. Samson, A. J. Lin, X. Chang, C. Tsien, S. Fergus, D. Mullen, D. Yang, D. Thotala, D. Hallahan, J. L. Campian, and J. Huang. Effect of Radiation Treatment Volume Reduction on Lymphopenia in Patients Receiving Chemoradiotherapy for Glioblastoma. *International Journal of Radiation Oncology*Biology*Physics*, 101(1):217–225, May 2018.
- [293] R. Davuluri, W. Jiang, P. Fang, C. Xu, R. Komaki, D. R. Gomez, J. Welsh, J. D. Cox, C. H. Crane, C. C. Hsu, and S. H. Lin. Lymphocyte Nadir and Esophageal Cancer Survival Outcomes After Chemoradiation Therapy. *International Journal of Radiation Oncology*Biology*Physics*, 99(1):128–135, Sept. 2017.
- [294] A. T. Wild, X. Ye, S. G. Ellsworth, J. A. Smith, A. K. Narang, T. Garg, J. Campian, D. A. Laheru, L. Zheng, C. L. Wolfgang, P. T. Tran, S. A. Grossman, and J. M. Herman. The Association Between Chemoradiation-related Lymphopenia and Clinical Outcomes in Patients With Locally Advanced Pancreatic Adenocarcinoma. *American Journal of Clinical Oncology*, 38(3):259–265, June 2015.
- [295] OYEON CHO, MISON CHUN, SUK-JOON CHANG, YOUNG-TAEK OH, and O KYU NOH. Prognostic Value of Severe Lymphopenia During Pelvic Concurrent Chemoradiotherapy in Cervical Cancer. *Anticancer Research*, 36(7):3541, July 2016.

- [296] J. Shin, S. Xing, L. McCullum, A. Hammi, J. Pursley, C. A. Correa, J. Withrow, S. Domal, W. Bolch, H. Paganetti, and C. Grassberger. HEDOS—a computational tool to assess radiation dose to circulating blood cells during external beam radiotherapy based on whole-body blood flow simulations. *Physics in Medicine & Biology*, 66(16):164001, Aug. 2021.
- [297] A. Hammi, H. Paganetti, and C. Grassberger. 4D blood flow model for dose calculation to circulating blood and lymphocytes. *Physics in Medicine & Biology*, 65(5):055008, Mar. 2020.
- [298] O. M. Otterlei, D. J. Indelicato, L. Toussaint, K. S. Ytre-Hauge, S. Pilskog, L. F. Fjaera, E. Rørvik, H. E. S. Pettersen, L. P. Muren, Y. Lassen-Ramshad, M. Di Pinto, and C. H. Stokkevåg. Variation in relative biological effectiveness for cognitive structures in proton therapy of pediatric brain tumors. *Acta Oncologica*, 60(2):267–274, Feb. 2021.
- [299] A. Bertolet, R. Abolfath, D. J. Carlson, R. A. Lustig, C. Hill-Kayser, M. Alonso-Basanta, and A. Carabe. Correlation of LET With MRI Changes in Brain and Potential Implications for Normal Tissue Complication Probability for Patients With Meningioma Treated With Pencil Beam Scanning Proton Therapy. *International Journal of Radiation Oncology*Biophysics*, 112(1):237–246, Jan. 2022.
- [300] J. Unkelbach, P. Botas, D. Giantsoudi, B. L. Gorissen, and H. Paganetti. Reoptimization of Intensity Modulated Proton Therapy Plans Based on Linear Energy Transfer. *International Journal of Radiation Oncology*Biophysics*, 96(5):1097–1106, Dec. 2016.
- [301] W. Cao, A. Khabazian, P. P. Yepes, G. Lim, F. Poenisch, D. R. Grosshans, and R. Mohan. Linear energy transfer incorporated intensity modulated proton therapy optimization. *Physics in Medicine & Biology*, 63(1):015013, Dec. 2017.
- [302] E. Rørvik, L. F. Fjæra, T. J. Dahle, J. E. Dale, G. M. Engeseth, C. H. Stokkevåg, S. Thörnqvist, and K. S. Ytre-Hauge. Exploration and application of phenomenological RBE models for proton therapy. *Physics in Medicine & Biology*, 63(18):185013, Sept. 2018.
- [303] A. Carabe-Fernandez, A. Bertolet-Reina, I. Karagounis, K. Huynh, and R. G. Dale. Is there a role for arcing techniques in proton therapy? *The British Journal of Radiology*, 93(1107):20190469, Mar. 2020.

- [304] Y. R. Lawrence, X. A. Li, I. El Naqa, C. A. Hahn, L. B. Marks, T. E. Merchant, and A. P. Dicker. Radiation Dose–Volume Effects in the Brain. *International Journal of Radiation Oncology*Biology*Physics*, 76(3):S20–S27, Mar. 2010.
- [305] C. E. Mercado, A. L. Holtzman, R. Rotondo, M. S. Rutenberg, and W. M. Mendenhall. Proton therapy for skull base tumors: A review of clinical outcomes for chordomas and chondrosarcomas. *Head & Neck*, 41(2):536–541, Feb. 2019.
- [306] D. C. Weber, R. Malyapa, F. Albertini, A. Bolsi, U. Kliebsch, M. Walser, A. Pica, C. Combescure, A. J. Lomax, and R. Schneider. Long term outcomes of patients with skull-base low-grade chondrosarcoma and chordoma patients treated with pencil beam scanning proton therapy. *Radiotherapy and Oncology*, 120(1):169–174, July 2016.
- [307] D. J. Indelicato, S. Flampouri, R. L. Rotondo, J. A. Bradley, C. G. Morris, P. R. Aldana, E. Sandler, and N. P. Mendenhall. Incidence and dosimetric parameters of pediatric brainstem toxicity following proton therapy. *Acta Oncologica*, 53(10):1298–1304, Oct. 2014.
- [308] M. Tambas, R. J. Steenbakkers, H. P. Van Der Laan, A. M. Wolters, R. G. Kierkels, D. Scandurra, E. W. Korevaar, E. Oldehinkel, T. W. Van Zon-Meijer, S. Both, J. G. Van Den Hoek, and J. A. Langendijk. First experience with model-based selection of head and neck cancer patients for proton therapy. *Radiotherapy and Oncology*, 151:206–213, Oct. 2020.
- [309] C. Bert and M. Durante. Motion in radiotherapy: particle therapy. *Physics in Medicine and Biology*, 56(16):R113–R144, Aug. 2011.
- [310] C. Bert, S. O. Grözinger, and E. Rietzel. Quantification of interplay effects of scanned particle beams and moving targets. *Physics in Medicine and Biology*, 53(9):2253–2265, May 2008.
- [311] H. Li, X. Zhang, P. Park, W. Liu, J. Chang, Z. Liao, S. Frank, Y. Li, F. Poenisch, R. Mohan, M. Gillin, and R. Zhu. Robust optimization in intensity-modulated proton therapy to account for anatomy changes in lung cancer patients. *Radiotherapy and Oncology*, 114(3):367–372, Mar. 2015.
- [312] M. H. Phillips, E. Pedroni, H. Blattmann, T. Boehringer, A. Coray, and S. Scheib. Effects of respiratory motion on dose uniformity with a charged particle scanning method. *Physics in Medicine and Biology*, 37(1):223–233,

Jan. 1992.

- [313] S. M. Zenklusen, E. Pedroni, and D. Meer. A study on repainting strategies for treating moderately moving targets with proton pencil beam scanning at the new Gantry 2 at PSI. *Physics in Medicine and Biology*, 55(17):5103–5121, Sept. 2010.
- [314] J. Seco, D. Robertson, A. Trofimov, and H. Paganetti. Breathing interplay effects during proton beam scanning: simulation and statistical analysis. *Physics in Medicine and Biology*, 54(14):N283–N294, July 2009.
- [315] E. Engwall, L. Glimelius, and E. Hynning. Effectiveness of different rescanning techniques for scanned proton radiotherapy in lung cancer patients. *Physics in Medicine & Biology*, 63(9):095006, May 2018.
- [316] T. Bortfeld. The number of beams in IMRT—theoretical investigations and implications for single-arc IMRT. *Physics in Medicine and Biology*, 55(1):83–97, Jan. 2010.
- [317] G. Liu, L. Zhao, D. Yan, R. Deraniyagala, C. Stevens, X. Li, and X. Ding. The First Modeling of the Spot-Scanning Proton Arc (SPArc) Delivery Sequence and Investigating its Efficiency Improvement in the Clinical Proton Treatment Workflow. *International Journal of Radiation Oncology*Biophysics*, 111:e524, Nov. 2021.
- [318] K. Suzuki, M. T. Gillin, N. Sahoo, X. R. Zhu, A. K. Lee, and D. Lippy. Quantitative analysis of beam delivery parameters and treatment process time for proton beam therapy. *Medical Physics*, 38(7):4329–4337, July 2011.