

**CONTROLLING BEAM COMPLEXITY IN
INTENSITY MODULATED RADIATION THERAPY**

by

Martha Marie Matuszak

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Nuclear Engineering and Radiological Sciences)
in the University of Michigan
2007**

Doctoral Committee:

**Professor Edward W. Larsen, Co-Chair
Professor Benedick A. Fraass, Co-Chair
Professor William R. Martin
Professor Alex F. Bielajew
Associate Professor Marc L. Kessler**

OCKHAM'S RAZOR

“Entia non sunt multiplicanda praeter necessitatem”

“Entities should not be multiplied unnecessarily”

© Martha Marie Matuszak

All rights reserved

2007

ACKNOWLEDGEMENTS

There are a number of people that have made this dissertation possible through their academic, professional and personal contributions and support. This dissertation is not simply a compilation of my research, but represents many fruitful collaborations with colleagues from the Departments of Radiation Oncology and Nuclear Engineering & Radiological Sciences at the University of Michigan. I would like to take this time to sincerely thank the people who have contributed to this work.

I have been extremely lucky to have two very intelligent and dedicated thesis advisors. I would like to extend the warmest of thanks and admiration for Dr. Benedick Fraass of Radiation Oncology and Dr. Edward Larsen of Nuclear Engineering & Radiological Sciences. They have been a great source of knowledge throughout the last five years, and I owe much of my success to their guidance. I am honored to have worked under their direction for my doctoral degree. They allowed me to work independently to develop my research skills, but were always available when I needed support. I have had the pleasure of working with Dr. Fraass for over seven years, and he has been and will remain a strong guiding force in my research and in my career. I could not have asked for a more exceptional leader and mentor to learn from. His work ethic and integrity are two things to be truly admired. Dr. Larsen was always very enthusiastic and interested in my work, and for that I am very grateful. He took an active role in my research in Radiation Oncology, and his expertise was invaluable in formulating many of the ideas in this work. He spent many hours of careful and prompt editing of abstracts, manuscripts, and thesis chapters. Dr. Larsen is an exceptional teacher and advisor. He was always very approachable when I had questions, and his answers were always clear and concise.

I would like to thank my committee members, Drs. William Martin and Alex Bielajew of Nuclear Engineering & Radiological Sciences, and Dr. Marc Kessler of Radiation Oncology. I have taken classes from Drs. Martin and Bielajew, and I admire them a great deal. Both have a variety of commitments, and it really meant a lot to me

that they were able to serve on my committee. Dr. Kessler has been a great source of help in many ways over the past seven years. From advising me in a first year project in image registration, to serving as my thesis proposal chair, to serving on my dissertation committee, Dr. Kessler has taught me a lot about organizing scientific studies, presentations, and manuscripts. I would also like to thank him for answering my many programming questions with a smile, and for his enthusiasm.

All of the co-authors that I have worked with on abstracts and papers also deserve my gratitude. They have contributed to this dissertation and in my development as a scientist. Karen Vineberg supervised my first project in Radiation Oncology as an undergraduate. I thank her not only for her friendship, but for helping to fuel my interest in the field, and teaching me about treatment planning and optimization. I also would like to thank Dr. James Balter for guiding me through my first AAPM abstract and paper. He is an excellent mentor and advisor, and I would have been lucky to continue working with him on my Ph.D. I'd also like to thank Dr. Randy Ten Haken for working with me on several projects, and also for being a constant presence and source of feedback for presentations. Dr. Daniel McShan has been patiently answering all of my optimization and programming questions since I began working in Radiation Oncology in 2000. I certainly would not have gotten very far into the optimization code without his advice. A big thank you also to Dr. Kyung-Wook (Ken) Jee for helping me understand the different optimization methods. His work on our optimization system has made it possible for my research to have a much bigger impact than it would have otherwise.

I would like to express my appreciation to Robin Marsh for being a wonderful friend and source of knowledge during my time in Radiation Oncology. She has answered all of my treatment planning questions over the years and was the first to try out basis function optimization and the smoothing costlets. She has been a huge asset to me in so many ways over the past several years.

Dr. Jean Moran has been a close friend and mentor since I began research in Radiation Oncology as an undergraduate. She has always gone out of her way to advocate for the students, and I would like to recognize and thank her for that. Her support and thoughtful advice on presentations, papers, and career paths have meant a great deal to me. I've been lucky enough to have worked with her on several different projects—most which are included in this dissertation. I thank her very much for her contributions to this work, and for her guidance and friendship.

It has truly been a pleasure to work with the entire faculty and staff in Radiation Oncology. I have enjoyed working with many fellow graduate students, post-docs, and residents, and I thank them for being there for support and camaraderie throughout graduate school. Thank you to Mihaela Rosu, Neelam Tyagi, and Rojano Kashani especially for being great friends throughout my time at the University of Michigan.

I would like to thank all of my friends and family for being supportive and understanding throughout my years at the University of Michigan. I owe a great deal of thanks to my husband, Dr. Sean Matuszak, for his love and support during graduate school, and especially during the last few months. His passion for medicine and his work ethic have been inspirational to me. I know that it wasn't always easy to put up with my computer and programming struggles and my late-night hours finishing up abstracts and presentations. Whenever I started to doubt myself, he was always there to support me and help me realize my contributions.

I would not be writing this if it weren't for my parents, Susan and Christopher Coselmon. They have always supported me and shown me how important education is. I'm extremely proud to be their daughter, and I thank them for all of the sacrifices they made to support me while I've been in college. I know I've missed out on a lot of things during graduate school, and I thank them for being understanding about all of those times. I'd also like to thank the rest of my family, especially my brother and his family (Adam, Rachel and Sophia Coselmon), my grandparents (Beatrice and Duane Coselmon and Stella Sizemore (Rest in Peace Grandpa Sizemore)), and my mother, father, and brother-in-law (Cynthia, Gerald and David Matuszak) for always supporting me.

Finally, I would like to acknowledge the many scientists that have come before me and made advancements much greater than my own. I would also like to thank God for all of the many blessings in my life. I know that He has gently steered me to this point, and I look forward to the many opportunities and challenges that lie ahead.

I would like to acknowledge financial support from the University of Michigan Rackham Graduate School, the Departments of Nuclear Engineering & Radiological Sciences and Radiation Oncology, and the National Institute of Health.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF FIGURES	vii
LIST OF TABLES	xvii
ABSTRACT	xix
CHAPTER I. INTRODUCTION	1
I.A. Progression of Radiation Therapy: The Path to Intensity-Modulated Radiation Therapy (IMRT)	2
I.B. Inverse-Planned IMRT	6
I.C. The Fundamental Problem in IMRT	15
I.D. Review of IMRT Complexity Reduction Research	18
I.E. Dissertation Overview	20
CHAPTER II. A REVIEW OF THE SOURCES AND CONSEQUENCES OF IMRT BEAM COMPLEXITY	29
II.A. Overview of Treatment Planning Details and Methods	30
II.B. Consequences of Highly Modulated Fields	31
II.C. Sources of IMRT Beam Complexity	37
CHAPTER III. REDUCING THE PROBLEM SCALE: OPTIMIZATION OF MATHEMATICAL SURFACES VERSUS INDIVIDUAL BEAMLET INTENSITIES	67
III.A. Motivation	67
III.B. Methods	68
III.C. Results	81
III.D. Discussion	93

CHAPTER IV. RESTRICTING OR PENALIZING BEAM MODULATION DURING INVERSE PLANNING FOR IMRT	103
IV.A. Motivation	103
IV.B. Methods	105
IV.C. Results	115
IV.D. Discussion	135
CHAPTER V. ADAPTIVE DIFFUSION SMOOTHING	140
V.A. Motivation	140
V.B. Methods	141
V.C. Results	161
V.D. Discussion	168
V.E. Conclusion	170
CHAPTER VI. APPLICATIONS OF ADAPTIVE DIFFUSION SMOOTHING	174
VI.A. ADS applied within Lexicographic Ordering	174
VI.B. Minimizing Beam Modulation with ADS: A Retrospective Study in Partial Breast IMRT	190
VI.C. Use of ADS in Radiobiological Targeting	194
VI.D. Future Applications of ADS	205
CHAPTER VII. POTENTIAL CLINICAL IMPACT OF CONTROLLING IMRT BEAM MODULATION	210
VII.A. Treatment Planning Robustness and Efficiency	210
VII.B. Delivery Accuracy and Efficiency	222
VII.C. Sensitivity to Geometric Uncertainty	230
VII.D. Potential Impact on Adaptive Radiotherapy	237
VII.E. Conclusions	242
CHAPTER VIII. SUMMARY	245

LIST OF FIGURES

Figure I-1. An illustration of the components inside a typical linear accelerator head.....	3
Figure I-2. Left: The Varian Trilogy™ linear accelerator shown with an on-board imager. Right: A close-up of the treatment head and the multi-leaf collimator. Images courtesy of Varian Medical Systems, Inc.....	4
Figure I-3. An example of the progression from conventional to 3D conformal to intensity modulated radiation therapy.....	5
Figure I-4. An example of some of the steps involved in inverse-planned IMRT. Upper-left: Contouring on CT, Upper-right: placing points in regions of interest, Lower-left: setting up beam directions, and Lower right: defining the beamlet grid.....	8
Figure I-5. A 3D visualization of an optimized plan for a solid water phantom geometry test plan.....	11
Figure I-6. (a) Isodose contours and (b) a colorwash dose display for the final IMRT dose calculation with the colorscale given in Gy.....	13
Figure I-7. Clockwise from top left: a conformal “4 field box” prostate treatment; a 7 beam inverse planned IMRT treatment with an inset showing the modulation in beam 1; the dose colorwash for the IMRT plan; the dose colorwash for the 3D conformal prostate plan. The prostate, bladder, rectum, and femora are represented by the surfaces and contours in the colors given, and the isodose color scale is also given.....	16
Figure I-8. An example of the complicated geometries defined for (a) head and neck cancer and (b) pancreatic cancer treatment, and the highly complex beamlet intensity patterns that are created by inverse IMRT planning.....	17
Figure II-1. An illustration of the tongue and groove effect that occurs in IMRT beam delivery. (a) A film measurement of an IMRT beam, (b) a segment of the delivery with leaves going across the width of the segment, and (c) a dose difference display of the calculations minus the film that shows underdose in the “tongue” region of the MLC.....	34
Figure II-2. An illustration of the decreasing delivery efficiency of IMRT fields as they become more complex. The MU required for step-and-shoot delivery and a conservative approximate time in seconds for each delivery are shown.....	35

Figure II-3. The modified IMRT benchmark phantom designed for IMRT validation and testing.....	38
Figure II-4. The optimal IMRT beams as fewer points are used in the optimization. Results are shown with the full sampling and when the number of points is decrease by factors of 2-5.	40
Figure II-5. The plan intensity map variation, monitor units, cost and average correlation coefficient for plans with decreasing point densities relative to the most dense point sampling (1).	41
Figure II-6. (a) The optimal DVHs for plans using increasingly fewer points in the optimization, and (b) the true DVHs for each of the subsampled plans when the intensity patterns are evaluated using the full sample of points.....	42
Figure II-7. Dose-volume histograms for the phantom when using 5 different random instances of point sampling in the structures (A-E).....	43
Figure II-8. Optimized beamlet patterns for 5 different point samplings. Beams 1-5 are shown and each different point sampling is labeled A-E.	44
Figure II-9. Total monitor units, plan intensity map variation, and number of iterations to full convergence for plans that are identical except for different random point samplings of the regions of interest.....	45
Figure II-10. Dose-volume histograms for the phantom when the optimized beams from all point samplings are evaluated with the point sampling of A.....	46
Figure II-11. A tangential left breast IMRT treatment. The left breast, heart, and left lung are shown in relation to the two beams.	47
Figure II-12. (a) Dose-volume histograms for three different optimized breast plans starting at different random beamlet intensities, and (b) the corresponding beamlet distributions. The violet circles and arrows represent points in the field that roughly cover the same geometry in both beams.....	48
Figure II-13. Dose-volume histograms and intensity modulated beams for plans optimized with quasi-Newton and simulated annealing optimization algorithms.....	53
Figure II-14. The PIMV, MU and cost for plans optimized with the given the quasi-Newton vs. simulated annealing optimization methods.....	54
Figure II-15. Relative PIMV, MU and Cost for plans optimized with the given (a) Toler stopping criteria and starting at uniform beamlet intensities of 0 (dotted lines) or random beamlet intensities (solid lines), and (b) Factr stopping criteria starting at starting uniform beamlet intensities of zero. All values are relative to the fully converged plans with starting intensities of zero.....	54
Figure II-16. DVHs and beams for the series of plans optimized with increasing (a and b) Toler stopping criteria and starting at uniform beamlet intensities of 0 (dotted lines) or (c and d) random beamlet	

intensities, and (e and f) Factor stopping criteria starting at starting uniform beamlet intensities of zero.....	55
Figure II-17. Dose-volume histograms for the phantom when the optimized beams from 5 different random starting points	56
Figure II-18. IMRT beams for cases optimized to full convergence from different starting points. The uniform start is shown as a control for the 5 random starts	57
Figure II-19. Beamlet different displays between the uniform start and each of the random starts. The colors are cut off between -10 and 10 although the actual differences could be higher.....	58
Figure II-20. The results from each step of the brain optimization outlined in Table 3. On the top left are DVHs for the first three steps of the optimization process and on the bottom left are DVHs from the final three steps. Steps 5 and 6 compare the use of Mean vs. Threshold costlets to minimize normal tissue dose.....	61
Figure II-21. The relative monitor units and plan intensity map variation after each step of the brain optimization outlined in Table II-3. Steps 5 and 6 compare the use of Mean vs. Threshold costlets to minimize normal tissue dose	62
Figure II-22. DVHs and beams for the conventional cost function and the equivalent uniform dose (EUD1) cost function and the stricter EUD cost function (EUD2)	62
Figure III-1. An example of a 2D 2 nd order polynomial that could be used to calculate the beamlet intensities. Here $\mathbf{c} = [16.5 \ 0.5 \ 0 \ 0.25 \ -0.15 \ 0.05]$ in Equation III.7	72
Figure III-2. (a) A single Gaussian radial basis function centered at (0,0) with $a = 0.25$ and amplitude 10, and (b) A grid of radial basis functions identical to (a), but centered at various (x,y) positions in the field.....	73
Figure III-3. An example of a beam that could be made optimizing the amplitudes of a 3 x 3 grid of Gaussian radial basis functions	74
Figure III-4. Flow chart representing how basis function optimization fits into the conventional beamlet optimization infrastructure	75
Figure III-5. Basis function optimization user interface that has been implemented into UMOpt.....	76
Figure III-6. Geometry and beam arrangement for the phantom case used in the basis function optimization characterization studies.....	78
Figure III-7. Dose-volume histograms and the corresponding intensity modulated beams for plans optimized with various order polynomial functions. Also shown are the baseline IMRT results.....	82
Figure III-8. Dose-volume histograms and the corresponding intensity modulated beams for plans optimized with various sized grids of radial basis functions. Also shown are the baseline IMRT results	83

Figure III-9. The fitted dilation (a) for each radial basis function set plotted as a function of the number of RBFs in the set. The data were successfully fitted to a power function (shown in inset).....	84
Figure III-10. The relative number of MU and optimization variables for each of the optimized cases as a function of the cost. The absolute number of total MU and variables in the baseline plan were 309, and 19,399, respectively	85
Figure III-11. Dose metrics for all optimization trials for the PTV, OAR1, OAR2, and normal tissue	86
Figure III-12. Dose-volume histograms for the baseline beamlet plan in the brain in addition to plan optimized with 2 nd and 4 th order polynomial functions and 4 x 4 and 7 x 7 radial basis function grids.....	87
Figure III-13. The standard deviation in several relevant dose metrics between all optimization trials including baseline, 2 nd , 4 th , 6 th , and 8 th order polynomials, and 3x3, 4x4, 5x5, and 7x7 gaussian radial basis function grids. More details on the metrics with standard deviations equal to or greater than 1 Gy are shown in Figure III-14.....	87
Figure III-14. Dose metrics for all optimization trials for the structures with large dose variations between methods	88
Figure III-15. Intensity modulated beams for the brain example. Baseline, 4 th and 8 th order polynomials, and 4x4 and 7x7 radial basis functions are shown	88
Figure III-16. The relative number of MU and optimization variables for each of the optimized brain plans. The absolute number of total MU and variables in the baseline plan were 354 and 26740, respectively	89
Figure III-17. Dose-volume histograms for the baseline beamlet plan in the prostate example in addition to plan optimized with 4 th and 8 th order polynomial functions and 4 x 4 and 7 x 7 radial basis function grids.....	90
Figure III-18. Intensity modulated beams for the prostate example. Baseline, 4 th and 8 th order polynomials, and 4x4 and 7x7 radial basis functions are shown along with the relative MU requirements	90
Figure III-19. The relative number of MU and optimization variables for each of the optimized prostate plans. The absolute number of total MU and variables in the baseline plan were 837, and 41,714, respectively	91
Figure III-20. Dose-volume histograms for the baseline beamlet plan in the head/neck example in addition to plan optimized with 8 th order polynomial functions and a 7 x 7 radial basis function grid	92
Figure III-21. Intensity modulated beams for the head/neck example. Baseline, 8 th order polynomials, and 7x7 radial basis functions are shown along with the relative MU requirements.....	93
Figure IV-1. An illustration of a smoothing window, D centered at beamlet, b_i , from a beamlet intensity matrix, B . The elements within the smoothing window are denoted by a vector of intensities, d	108

Figure IV-2. An illustration of the SG1D and SG2D smoothing windows, D centered at beamlet, b_j , from a beamlet intensity pattern. The elements within the smoothing window are inside the black rectangle. Note the direction of MLC motion is shown to be parallel to the SG1D smoothing window.....	111
Figure IV-3. Illustration of the CT phantom geometry and beam placement. The PTV is the violet center sphere, “Small” is the yellow off-center sphere, and “Cylinder” is the cylindrical gray structure. The blue external contours were not used in the cost function to save calculation time.....	113
Figure IV-4. The relative monitor units, mean correlation coefficient, and relative plan intensity map variation as a function of total cost for the CT phantom when optimized with different maximum intensity limits on the beamlets. The maximum intensity ratio for each plan is shown below the lines. All results are relative to baseline plan which had a maximum intensity ratio of 4. The correlation coefficient value was averaged over all beams	116
Figure IV-5. DVHs and beams for plans at various maximum intensity ratios (MIR) in the CT phantom	117
Figure IV-6. The relative MU, mean correlation coefficient, and relative PIMV as a function of total cost for the CT phantom when optimized with the SG1D or SG2D penalty in the objective function. All results are relative to baseline plan. The correlation coefficient value was averaged over all beams	118
Figure IV-7. DVHs and beams for plans at various SG1D penalty weights in the CT phantom. Weights in increasing order are given by i-iv	119
Figure IV-8. DVHs and beams for plans at various SG2D penalty weights in the CT phantom. Weights in increasing order are given by i-iv	119
Figure IV-9. The relative MU, mean correlation coefficient, and relative PIMV as a function of total cost for the CT phantom when optimized with the PIMVq penalty in the objective function. All results are relative to baseline plan and the correlation coefficient value was averaged over all beams	121
Figure IV-10. DVHs and beams for plans at various PIMVq penalty weights in the CT phantom. Weights in increasing order are given by i-iv	122
Figure IV-11. Plots illustrating relative MU versus net cost tradeoffs in the brain cases when using each of the modulation penalties compared to a standard cost function and maximum intensity limits. All values are relative to the standard plan	123
Figure IV-12. DVHs and corresponding beams shown optimized with each of the beam modulation reduction methods for (a,b) Brain 1, (c,d) Brain 2, and (e,f) Brain 3. SMLC MU reductions are in parentheses.....	124
Figure IV-13. Plots illustrating relative MU versus net cost tradeoffs in the prostate cases when using each of the modulation penalties	

compared to a baseline cost function and maximum intensity limits. All values are relative to the standard plan	126
Figure IV-14. DVHs and corresponding beams 1-4 out of 9 shown optimized with each of the beam modulation reduction methods for (a,b) Prostate 1, (c,d) Prostate 2, and (e,f) Prostate 3. SMLC MU reductions are in parentheses	127
Figure IV-15. Doses in each prostate case plotted as a function of SMLC MU when using each of the optimization techniques. Trends in the PTV minimum and mean dose are shown along with normal tissue mean doses	128
Figure IV-16. <i>Prostate1</i> DVHs shown optimized with each of the beam modulation reduction methods and maximum intensity limits when including mean dose costlets for the normal tissues equal to the mean doses achieved by the baseline plan	129
Figure IV-17. Plots illustrating relative MU versus net cost tradeoffs in the head/neck cases when using each of the modulation penalties compared to a standard cost function and maximum intensity limits. All values are relative to the standard plan	130
Figure IV-18. DVHs and corresponding beams 1-4 out of 9 shown optimized with each of the beam modulation reduction methods for (a,b) HN 1, (c,d) HN 2, and (e,f) HN 3. SMLC MU reductions are in parentheses	132
Figure V-1. Illustration of the diffusion smoothing procedure and some of the notation used in this work. We start with an intensity modulation beam, I_0 , and a grid of corresponding diffusion coefficients, D , that dictate the degree of the smoothing allowed or applied over each beamlet and then solve our diffusion equation analogy to get the diffusion smoothed beam, I_s	142
Figure V-2. An example of the increase in smoothing that occurs as the diffusion coefficient value increases from the bottom to the top of the field	151
Figure V-3. The diffusion coefficient as a function of the cost function gradients with respect to the beamlets at different c and n values according to Equation V.26	154
Figure V-4. (a) An optimized brain IMRT beam, (b) the diffusion smoothed beams when using uniform coefficients, (c) absolute value of the cost function gradients with respect to the beamlets, $ \partial OF / \partial I_{ij} $, (d) adaptive diffusion smoothing coefficients that are defined as a function of (c), and (e) diffusion smoothed beams when using the gradient-based diffusion coefficients in (d). Beams are shown in the beam's eye view in relation to the PTV, brainstem, optic nerves, and optic chiasm. Blue represents low intensity beamlets (gradients, or diffusion coefficients) while red represents high intensity	155
Figure V-5(a) The geometry and beam arrangement, (b) relative MU (to the baseline plan) as a function of plan degradation units, (c) dose- volume histograms for the circled plans in (b), and (d) the	

corresponding intensity modulated beams in the beam's eye view for the phantom case 1 for standard IMRT and the three different modulation penalties. The outline of the two OARs is shown by the dotted lines. Red corresponds to high intensity beamlets and blue to low intensity beamlets	162
Figure V-6(a) MU (relative to the baseline plan) as a function of plan degradation units, (b) dose-volume histograms for the circled plans in (a), and (c) the corresponding intensity modulated beams in the beam's eye view for the prostate case for baseline IMRT and the three different modulation penalties. From left to right is Baseline, PIMVq, ADS-Uniform, and ADS-Gradient. The bladder, rectum, and penile bulb are shown in the beam's eye view	164
Figure V-7(a) Relative SMLC monitor units are shown as a function of the plan degradation baseline cost for the head/neck plan. (b) DVHs and (c) a typical intensity modulated beam are shown for the circled plan in (A) for each of the optimization methods. From left to right is Baseline, ADS-Gradient, ADS-Uniform, and PIMVq. Several normal structures are outlined in the beam's eye view	166
Figure V-8 (a) Head/neck geometry and intensity modulated beams for the baseline plan and the last (most smoothed) ADS-Gradient plan shown in Figure 5. The ADS-gradient plan can be delivered with 57 % fewer SMLC monitor units with minimal effect on the baseline plan quality, as shown in the DVHs in (b)	167
Figure VI-1. The modified IMRT benchmark geometry and beam arrangement used to study the initial characterization of adaptive diffusion smoothing and lexicographic ordering	180
Figure VI-2. The percent reduction in monitor units possible in each of the tradeoff scenarios (corresponding to Table VI-1) relative to the baseline geometry for the phantom case. The DVHs and beams for the tradeoff scenarios denoted in black are shown in Figure VI-3	184
Figure VI-3. The optimized dose-volume histograms and beams for the baseline plan and tradeoff scenarios 0, B, F and H for the phantom case shown in Figure VI-1	184
Figure VI-4. The percent reduction in monitor units possible in each of the tradeoff scenarios (corresponding to Table VI-1) relative to the baseline geometry for the prostate case. Tradeoff scenario G was not applicable because the rectum dose-volume constraint was considerably below the objective. The DVHs and beams for the tradeoff scenarios denoted in black are shown in Figures VI-5 and VI-6	185
Figure VI-5. The optimized dose-volume histograms for the baseline plan and tradeoff scenarios 0, D, and J. The DVHs are shown in two panels for viewing purposes.....	185
Figure VI-6. The optimized beams for the baseline plan and tradeoff scenarios 0, D, and J. The corresponding DVHs are shown in Figure VI-5.....	186

Figure VI-7. The percent reduction in monitor units possible in each of the tradeoff scenarios (corresponding to Table VI-3) relative to the baseline geometry for the head/neck case. Tradeoff scenario D was not applicable because the oral cavity mean dose was already past its intended limit. The DVHs and beams for the tradeoff scenarios denoted in black are shown in Figures VI-8 and VI-9.....	187
Figure VI-8. The optimized head/neck dose-volume histograms for the baseline plan and tradeoff scenarios 0, F, and J. Not all structures are shown for easier viewing although those not pictured showed similar trends.....	188
Figure VI-9. The optimized head/neck beams for the baseline plan and tradeoff scenarios 0, F, and J. The corresponding DVHs are shown in Figure VI-8.....	188
Figure VI-10. Required MU/Fraction for static delivery for each of optimization methods for each of the 10 breast patients.....	192
Figure VI-11. (a) A typical intensity modulated beam in the beam's eye view for the Baseline, ADS-Equivalent, ADS-Relax 3% and ADS-Relax 5% cases with the breasts (brown and green), lung (white), heart (red), and LAD (blue) shown. The beam is covering the CTV and the PTV (wireframe). (b) the corresponding DVHs or each of the plans.....	193
Figure VI-12. Example of how the $ a $ values are distributed in the beamlets to be used in the diffusion coefficient calculations. The higher $ a $ structures take priority	196
Figure VI-13. The test case geometry and beams. The PTV is gray ($a=-5$), and the boost targets are red ($a=-15$) and yellow ($a=-20$).....	198
Figure VI-14. Dose-volume histograms for the targets and normal tissues for the baseline and ADS IMRT plans. ADS 1, 2, and 3 denote the relative ADS penalty weight	201
Figure VI-15. A typical beam in the test case as the ADS penalty is increased. The two boost targets are shown. Blue denotes cold beamlets and red denotes hot beamlets	201
Figure VI-16. The gEUD for PTV50, and the two boost targets (70 and 90) before and after 3 mm isocenter shifts in the x (lateral), y (posterior), and z (superior) directions	202
Figure VI-17. Pancreas case DVHs for baseline and ADS IMRT plans	203
Figure VI-18. 3-D view of baseline vs. ADS IMRT plans in the pancreas	204
Figure VII-1. The modified IMRT benchmark phantom designed for IMRT validation and testing (reproduced from Chapter II).....	211
Figure VII-2. The group of (a) Baseline, (b) PIMV, (c) ADS-U, and (d) ADS-G DVHs optimized for each different point sampling A-E	213
Figure VII-3. Beam 3 as optimized with each different technique in each of the different random point sampling A-E	214

Figure VII-4. The mean correlation coefficient (and σ) between plans optimized on 5 different point sampling instances using 4 different techniques.....	214
Figure VII-5. DVHs for the PIMV plans from all point samplings evaluated on point sampling A.....	215
Figure VII-6. The mean cost (and σ) between plans optimized using 4 different techniques on random point sampling instances B-E and then evaluated on the point sampling of A. This would represent the cost of B-E if A were the true geometry.....	216
Figure VII-7. The DVHs for the phantom case when optimized with the full point sampling for all smoothing techniques compared to the baseline plan. This is shown to verify that the choice of the smoothing costlet weights for this comparison was conservative and resulted in equivalent smoothing for each of the cases.....	217
Figure VII-8. The average correlation coefficient between the optimal beams from each of the subsampled plans relative to fully sampled plan for each of the 4 techniques. The standard deviation of the correlation coefficient for each of the 5 beams is also shown.....	218
Figure VII-9. The difference in the true and perceived optimal cost function values as a function of subsampling for each of the techniques. Exponential fits to the data are shown in the lower right.....	219
Figure VII-10. The time and number of iterations required for full convergence in both optimization techniques as a function of the total cost function value (not including the smoothing costlet penalties) for the (a) brain, (b) prostate, and (c) head/neck studies. The leftmost datapoint in all cases represents the baseline plan, or a smoothing penalty weight of zero. Note the two scales.....	221
Figure VII-11. DVHs (for all of the techniques in the prostate cases) demonstrate the clinical acceptability of the weighting factors chosen. The different trials (PIMV1 and 2, ADS-G1 and 2) are relative indicators of the weighting factors.....	222
Figure VII-12. (a) Geometry, (b) DVHs, and (c) intensity distributions for Standard vs. ADS-Equivalent and ADS-Relaxed IMRT plans in the 3-field phantom case.....	224
Figure VII-14. Prostate example: Top: DVHs (Standard IMRT solid, ADS-Equivalent dashed, ADS-Relaxed dotted) and Bottom: IMRT intensity maps and dose difference displays for Standard IMRT (left), ADS-Equivalent (middle), and ADS-Relaxed (right). Differences are calculations – film and isodose contours for both are shown as overlays in cGy along with example absolute dose differences (arrows).....	226
Figure VII-13. Dose difference displays (calculations – film) for the 3 field phantom case for standard IMRT (top row), ADS-Equivalent (middle row), and ADS-Relaxed (bottom row) plans. Absolute dose difference examples are given by the arrows.....	227

Figure VII-15. Head/neck example: Top: DVHs (Standard IMRT solid, ADS-Equivalent dashed) and Bottom: Planned IMRT intensity maps and dose difference displays (calculations – film) for Standard IMRT (left) and ADS-Equivalent (right). Examples of absolute dose differences are shown by the arrows.....	228
Figure VII-16. Relative MU and time for all cases for both ADS-Equivalent and ADS-Relaxed plans relative to standard IMRT	229
Figure VII-17. The sampled residual setup errors (σ) in each direction for 10 fractions to be simulated with the phantom case	231
Figure VII-18. The DVHs and beams for each of the techniques used to study the impact of smoothing costlets on the robustness of the plan to geometric residual setup errors.....	233
Figure VII-19. PTV point dose differences in the Baseline plan (blue) and the ADS-Gradient (red) (a) Equivalent plan (b) Weighted1 plan and (c) Weighted2 plan	235
Figure VII-20. Histograms of cumulative point dose differences for the (a) PTV, (b) OAR1, and (c) OAR2 are shown in for baseline plan (blue) and the ADS-Gradient, Weighted2 plan (red)	236
Figure VII-21. Generalized equivalent uniform dose for all structures and techniques as evaluated in each simulated fraction. The horizontal lines approximate the gEUD values in the original plan.....	237
Figure VII-22. The simulated three stage dose escalation adaptive plan for a head/neck case	239
Figure VII-23. Final composite dose-volume histograms for the Baseline and ADS dose escalation treatment courses	240
Figure VII-24. The MU required per fraction for each stage of the adaptive dose escalation simulation for the Baseline and ADS treatment courses.....	240
Figure VII-25. Optimal IMRT beams as each stage in the adaptive dose escalation example for the Baseline and ADS treatment courses	241

LIST OF TABLES

Table II-1. Optimization trials for IMRT benchmark phantom	52
Table II-2. Mean and s of converged cost, MU, PIMV, and CC for 5 random starting beamlet intensities.....	57
Table II-3. Stepwise brain cost function	59
Table III-1. Basis function optimization parameters to be used in UMOpt.....	75
Table III-2. Cost functions used in the phantom and clinical cases	80
Table IV-1. Inverse planning objectives for the CT phantom and clinical cases.....	114
Table IV-2. Maximum MU reductions possible in the brain using each optimization technique while (i) meeting high priority plan objectives, and (ii) maintaining mean normal tissue doses from the baseline plan.....	125
Table IV-3. Maximum MU reductions possible in the prostate using each optimization technique while (i) meeting high priority plan objectives, and (ii) maintaining mean normal tissue doses from the baseline plan.....	128
Table IV-4. Maximum MU reductions possible in the head/neck using each optimization technique while (i) meeting high priority plan objectives, and (ii) maintaining mean normal tissue doses from the baseline plan	131
Table V-1. Test case plan objectives and plan degradation unit (PDU) scale for comparisons to baseline plan	157
Table V-2. Prostate plan objectives and plan degradation unit (PDU) scale for comparisons to baseline plan	158
Table V-3. Head/neck plan objectives and plan degradation unit (PDU) scale for comparisons to baseline plan	159
Table VI-1. Prioritized objectives and clinical tradeoff scenarios (allowed relaxations to promote beam smoothing and delivery efficiency) for the phantom case.....	182
Table VI-2. Prioritized objectives and clinical tradeoff scenarios (allowed relaxations to promote beam smoothing and delivery efficiency) for the prostate case.....	182
Table VI-3. Prioritized objectives and clinical tradeoff scenarios (allowed relaxations to promote beam smoothing and delivery efficiency) for the head/neck case	183

Table VI-4. Accelerated partial breast IMRT plan objectives	191
Table VI-5. Relative mean monitor unit requirements and target dose objectives for accelerated partial breast IMRT with and without an adaptive diffusion smoothing penalty used during optimization	192
Table VI-6. Test case objectives.....	199
Table VI-7. Pancreas case objectives	200
Table VI-8. Pancreas surgical boost gEUD [Gy] before and after 3 mm isocenter shifts	204
Table VII-1. The cost function applied to modified IMRT benchmark phantom	212
Table VII-2. Beam area outside +/- 5 cGy agreement between calculations and film measurements	224
Table VII-3. Percentage of points with > 2 Gy dose differences.....	234

ABSTRACT

CONTROLLING BEAM COMPLEXITY IN INTENSITY MODULATED RADIATION THERAPY

by

Martha Marie Matuszak

Co-Chairs: Edward W. Larsen and Benedick A. Fraass

External beam intensity modulated radiation therapy (IMRT) is a technique in which the spatial intensity of radiation from each beam direction can be modulated to provide superior conformality of dose to a tumor volume while sparing important normal tissues. A fundamental and potentially limiting feature of IMRT is the highly complex fields that can be created through inverse plan optimization. Highly modulated treatments are a large departure from conventional radiotherapy methods, are difficult to deliver accurately and efficiently, and can result in an undesirable increase in leakage dose being delivered to the patient. Longer deliveries may also increase the chance for patient motion during treatment and could potentially reduce the probability of controlling some tumors. The large intensity fluctuations observed in IMRT beams are often a result of the degeneracy of the optimization problem, and the types of optimization method and cost function used. This work demonstrates that beam complexity is a result of these two issues, and is dependent on the placement of dose evaluation points in the target and normal tissues. This research shows that (i) optimizing surfaces

instead of discrete beamlet intensities to represent the beam can reduce the degrees of freedom in IMRT and results in much smoother beams at the expense of a slight increase in normal tissues, (ii) maximum beamlet intensity restrictions are useful for improved delivery efficiency, but may restrict the optimizer at low limits, and (iii) modulation penalties can be incorporated into the cost function to promote plan smoothness without sacrificing plan quality. Penalizing the overall plan modulation is an effective way to reduce modulation, but it falsely penalizes the desirable beam modulation as well as the undesirable modulation. To address this problem, diffusion principles are used to develop a spatially adaptive smoothing method that only penalizes the unnecessary beam modulation and can be used without degrading plan quality. This method is customizable to a variety of treatment scenarios. The clinical impact of reducing beam complexity is significant, as it can result in an improvement in delivery accuracy and efficiency, quicker optimization times, and increased robustness to point sampling and geometric uncertainty.

CHAPTER I.

INTRODUCTION

Cancer continues to be one of the leading causes of death in the world, with as many as 10 million new cases diagnosed annually. The American Cancer Society predicts that in the United States in 2007, nearly 560,000 cancer deaths will occur and over 1.4 million new cases of cancer will be diagnosed¹. The many different types and presentations of cancer make it a difficult disease to combat and control. Research in one area may or may not cross over to another, and treatments that are successful in one area may fail in another. A common characteristic between different types of cancer cells is that they usually grow and divide more rapidly than healthy cells. Growing and dividing cells are generally very sensitive to radiation damage, making radiation therapy a common treatment among many different types of cancer. In fact, over half of all cancer patients receive some form of radiation treatment, and the majority of those patients will receive external beam radiation therapy.

The idea that radiation could be used therapeutically came to light over a century ago². Since then, the knowledge and technology behind radiation therapy has grown exponentially. The first radiation treatments were performed with little knowledge of the mechanism of tissue damage and repair and hence were full of uncertainty. Today's treatments are backed by years of research, development, and understanding, but there is still much to learn and much that can be improved upon.

External beam therapy delivers high energy electrons or photons (x-rays) from an external source—usually a linear accelerator—to cancerous tissue. The ultimate goal of external beam radiation therapy is to control the malignant cells while sparing as much healthy tissue as possible. Therefore, the planning of radiation therapy treatment is a process that must make tradeoffs between these two competing goals. One radiobiological aspect that aids in the planning process is the fact that healthy tissues,

while still at risk from radiation damage, can usually repair themselves more effectively than cancer cells. In addition, radiation treatments are generally divided into a series of treatments, called fractions, in order to allow the normal tissues ample time for damage repair and to take advantage of some of the radiobiological properties of cancer cells that may make them more sensitive to fractionated therapy. New findings in radiobiology, imaging, and physics continue to improve and alter the techniques used to deliver external radiation to patients.

I.A. Progression of Radiation Therapy: The Path to Intensity-Modulated Radiation Therapy (IMRT)

I.A.1. Conventional Radiation Therapy

What is now termed *conventional radiation therapy* refers to a two-dimensional radiation therapy planning technique with only a small number of beam directions (usually 1-4). Conventional treatment originated before three-dimensional imaging techniques, such as computed tomography, were available for volumetric imaging of the patient and remains the most common treatment in most of the world. This technique consists of simulating the treatment position of the patient and designing a plan that uses several usually well-established beam directions that will focus on the treatment area. This is performed on an *conventional radiotherapy simulator*, and the patient is usually placed in a supine position on the simulator platform. Two-dimensional imaging may be used to help visualize the tumor and surrounding normal tissue. In most situations, an x-ray is be used to determine the location of bony landmarks that are used to set the beam borders. To ensure coverage of the tumor, the beams have wide margins and intersect a large volume of normal tissue. The use of cerrobend blocks may be used to shield critical organs, such as the spinal cord, from the radiation beams.

The limitation of conventional radiation therapy is that low radiation doses must be used to not cause toxicity to the large amounts of healthy tissue that are irradiated. For moderate-to-aggressive tumors, the probability of controlling the disease for a long period of time may be low. In addition, if large areas of the beam are blocked to shield organs, the risk of underdosing the tumor volume can be high. Over the past several decades, the combination of increasing knowledge about radiation therapy and technological advances in imaging and linear accelerator hardware has allowed for more targeted and conformal treatments.

I.A.2. Conformal Radiation Therapy

Conformal radiation therapy, as its name implies, aims to conform the dose more closely around the tumor, in an effort to escalate dose to the target and spare more normal tissue. Conformal therapy is possible, in part, due to (i) the advancement of imaging technologies that provide an accurate 3D representation of the patient's internal anatomy and disease location, (ii) the development of three-dimensional radiation treatment planning software³⁻⁵, and (iii) advancements in the collimation capabilities of the treatment machine. In 3D conformal radiation therapy (3D-CRT), the conventional treatment simulator is replaced by a *CT treatment simulator*. The CT simulator differs from a diagnostic CT scanner in several respects; the simulator is designed with a larger bore that allows patients to be scanned in the treatment position and is equipped with localization devices to tie the position of the treatment planning CT to the coordinates at the treatment machine. The radiation oncologist can then define the tumor and normal structures by outlining them on each slice of the CT scan. This can be done at the CT simulator workstation or after the treatment planning CT is transferred to a 3D treatment

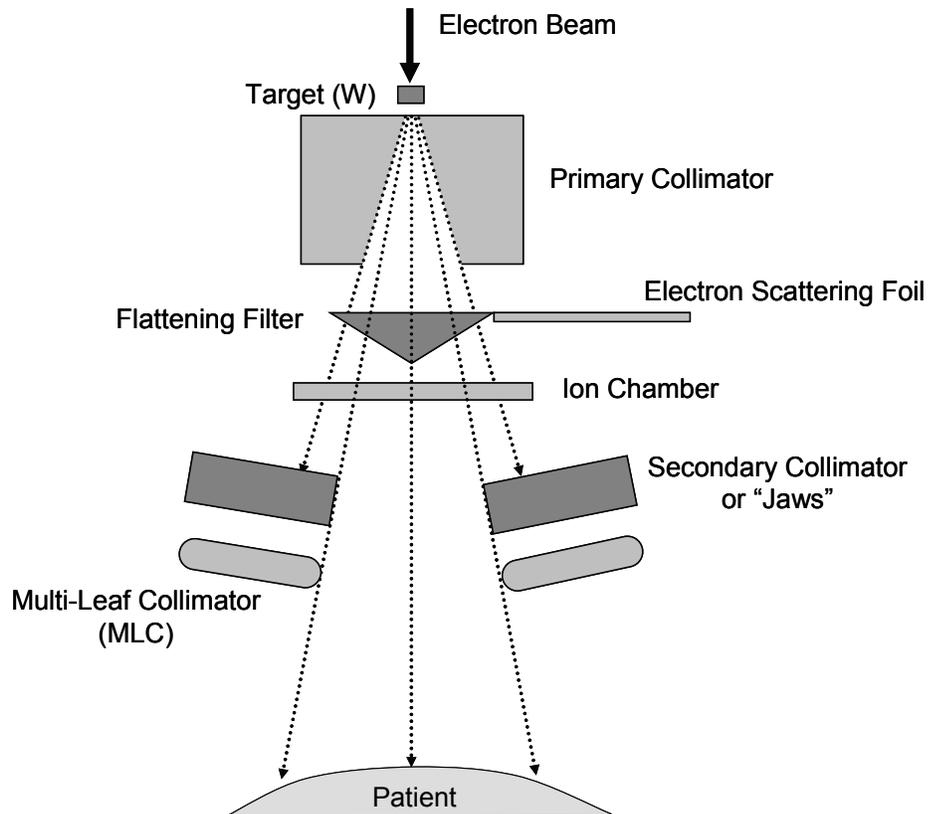


Figure I-1. An illustration of the components inside a typical linear accelerator head.

planning system. This ability to target the actual delineated tumor and internal organs allows a much higher conformity of treatment to the tumor as opposed to using pre-defined bony landmarks to set the treatment borders. It also individualizes the treatment to the specific anatomy of each patient. Treatment beams can then easily be defined by projecting the outline of the tumor and normal structure onto the same plane from the direction of the beam. This *beam's eye view* (BEV)^{6,7} is an essential tool for beam shaping in conformal radiation therapy.

The degree to which the beams can conform to the tumor depends on the collimation and beam shaping capabilities of the treatment machine. An illustration of the most common means for delivering external beam therapy—the linear accelerator—is shown in Figure I-1. Shown is the head of the accelerator which houses several components that are used to produce, shape, and filter the radiation beam. For photon therapy, a megavoltage electron beam is directed at a high atomic number target (usually tungsten) to produce bremsstrahlung radiation. This beam then is collimated by the *primary collimator* before passing through a *flattening filter*. The purpose of this conical filter is to make the beam intensity uniform across the field. The beam then passes through an ion chamber, which is used to monitor the output of the machine. Next, the beam is shaped by two additional collimation devices. The *secondary collimator* can shape the radiation beam into a customized rectangle with “jaws” that close in around the radiation beam, and more conformal shapes can be made using a

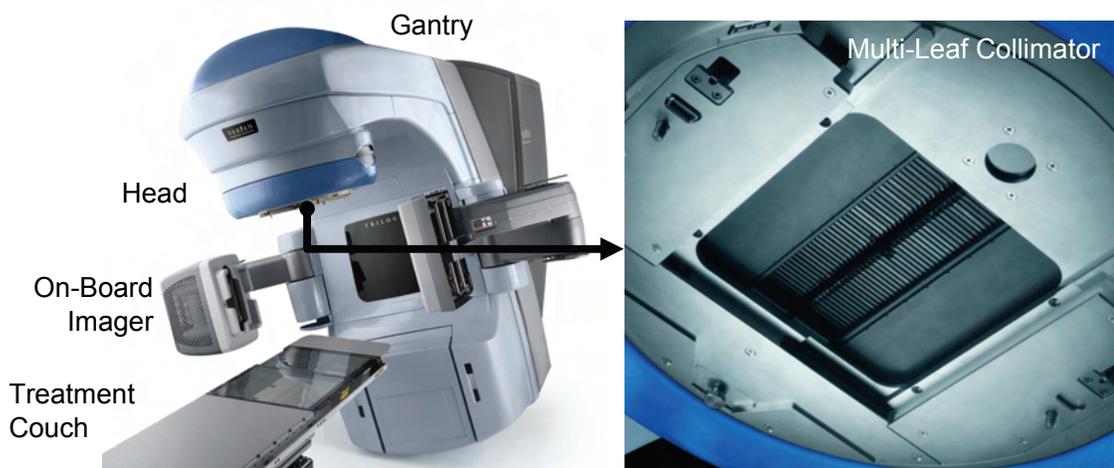


Figure I-2. Left: The Varian Trilogy™ linear accelerator shown with an on-board imager. Right: A close-up of the treatment head and the multi-leaf collimator. Images courtesy of Varian Medical Systems, Inc.

device called a *multi-leaf collimator* (MLC). A standard MLC consists of many pairs of “leaves” that can move in and out of the radiation field independently, to create an approximation to many different tumor shapes and sizes. Figure I-2 shows a modern linear accelerator and the attached MLC device. Linear accelerator designs vary among manufacturers, and in some cases, the MLC is placed above the secondary collimator⁸. However, the general beam shaping capabilities are similar.

Three-dimensional conformal therapy (3D-CRT)³⁻⁵ arose as a leading technology in radiation therapy in the mid-1980’s and increased the number of patients that were eligible for and treated successfully with radiation therapy. Improvements in local control and normal tissue sparing compared to conventional therapy have been documented in many treatment sites⁹⁻¹⁵. The use of volumetric imaging and dose calculations to more accurately locate the tumor position and calculate dose led to a gradual reduction in the margins around the tumor volume, which spared more normal tissue from radiation damage. This presented an opportunity to further escalate the dose to the tumor. More beam directions were used so that large volumes of normal tissue were not treated to very high doses. Instead, several beam directions were used to spread out the dose to the normal tissue, while focusing on the tumor. As 3D-CRT techniques continued to evolve and improve, the fast pace of the technological advancement in the field also continued. A notable advance was the idea that the beams weights could be chosen automatically by a computer optimization procedure called *inverse planning*. As opposed to *forward planning*, where a dosimetrist would iterate through beam weights and analyze the plan, inverse planning gave the treatment goals to a computer program which was used to choose the optimal beam weights^{16,17}. In some instances, several

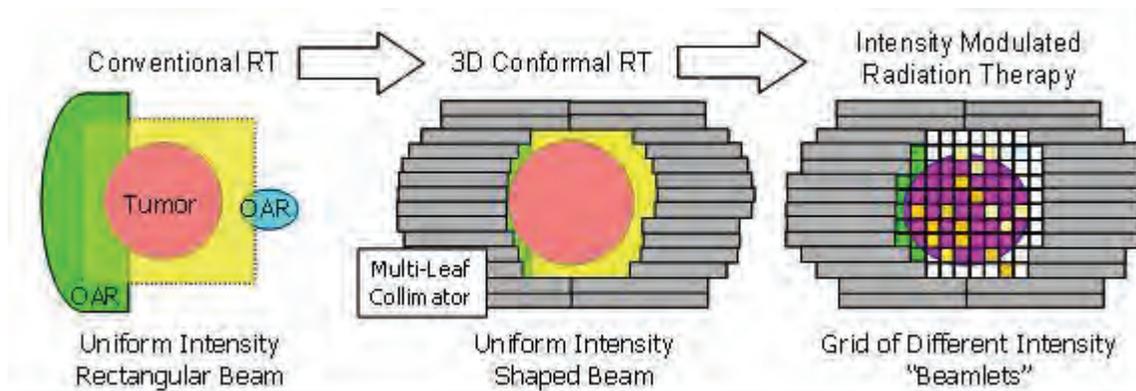


Figure I-3. An example of the progression from conventional to 3D conformal to intensity modulated radiation therapy.

beams from the same direction were used to modulate the radiation intensity and take advantage of the geometrical relations of the targets and normal tissues. This led to the birth of modern intensity-modulated radiation therapy (IMRT)^{16,18-22}. IMRT is a form of conformal radiation therapy in which the spatial intensity of each beam can be varied to further conform to the tumor and spare normal tissues. Figure I-3 illustrates the progression from conventional therapy to the most advanced form of inverse-planned IMRT from the beam's eye point of view for an arbitrary geometry.

I.A.3. Intensity-Modulated Radiation Therapy

There are many forms of IMRT in use today in radiation treatment centers²³. As mentioned in the previous section, forward-planned IMRT consists of placing many multiple shaped fields from different directions to ultimately conform to the tumor shape and spare any critical normal tissues. This level of IMRT overlaps with the planning of 3D-CRT and requires a very skilled planner. The natural progression of radiotherapy planning has led to an increase in the number of target volumes and normal structures that are delineated on the treatment planning CT with the goal to deliver varying prescription doses to targets or to limit the dose to minimize potential toxicity. As the treatment planning problem became more complex, the planning methods must be able to generate equally complex treatment plans. Therefore, it seemed natural to apply inverse planning techniques to the planning of intensity modulated radiation therapy^{16,18,21,22}. In inverse-planned IMRT, treatment planning optimization software determines the optimal intensity profiles from each beam direction that best achieve the treatment planning goals.

I.B. Inverse-Planned IMRT

In most treatment centers, IMRT is planned by dividing each radiation beam into a grid of beamlets (see Figure I-3) and then optimizing the desired weight or intensity of each beamlet based on various treatment goals specified by the radiation oncologist. This problem is a much larger-scale optimization problem than optimizing single beam weights in conformal radiation therapy and usually requires the use of sophisticated optimization algorithms, which will be discussed later in this section. However, there are many integral steps involved in reaching this optimization stage that can affect the quality of the final treatment plan. The common IMRT planning steps include:

1. Acquisition of the treatment planning CT and other multi-modal images to aid in tissue delineation.

Once a patient is recommended for radiation therapy, the treatment planning process commences by the acquisition of the treatment planning CT. In many treatment sites, the soft tissue contrast of the treatment planning CT is not sufficient to outline the tumor or normal structures. Therefore, other imaging studies, such as positron emission tomography (PET) and magnetic resonance imaging (MRI) may also be performed to allow better visualization of the patient's anatomy. Treatment planning dose calculations are performed on the CT so it is essential to be able to geometrically connect data from other imaging modalities to the CT scan. This is usually done through an image registration procedure, in which the data in one image is transformed to the coordinate system of the other image. There are a variety of different registration techniques that range from simple affine matching to fully deformable techniques. A review of the current issues and techniques in image registration for radiotherapy was recently published by Balter and Kessler²⁴.

2. Contouring target volumes, organs, and other clinical regions of interest on the treatment planning CT scan.

Once the imaging data is acquired, all of the important aspects of the patient's anatomy that can be distinguished are outlined or contoured by a radiation oncologist on the axial CT slices. Alternatively, contours can also be drawn on other imaging data such as PET or MR and then transferred to the CT, using the coordinate transformation obtained during the image registration process. This combination of the data from different modalities into one common set of data is generally referred to as data *fusion*. Because of our reliance on the drawn contours to delineate the target and organ boundaries for therapy, care must be taken to properly and carefully outline the anatomy so that delineation errors do not adversely affect patient outcome. In addition, errors in the registration of two imaging datasets can propagate errors through the contouring and treatment planning process.

An example of contouring is shown on the CT of a solid water phantom in Figure I-4 (upper left).

3. Creating the 3D representation of the patient that will be used for treatment planning.

Once the anatomy has been contoured in the multiple 2D planes of the CT scan, the collection of 2D information is converted in a 3D representation of the anatomy (i.e. a surface or voxel-based representation of the anatomy). This 3D model of the patient aids in the remainder of the treatment planning process. In inverse planning, each of regions of interest must be represented by a collection of voxels or data points.

At the University of Michigan, this representation consists of a set of discrete points that are distributed in each of the volumes. These points act as the representation of each structure in the optimization process. Because the optimization process uses a large amount of memory, it is not practical to use an unlimited number of points in each structure. Therefore, point placement must be appropriate so that important regions are not undersampled. An example of

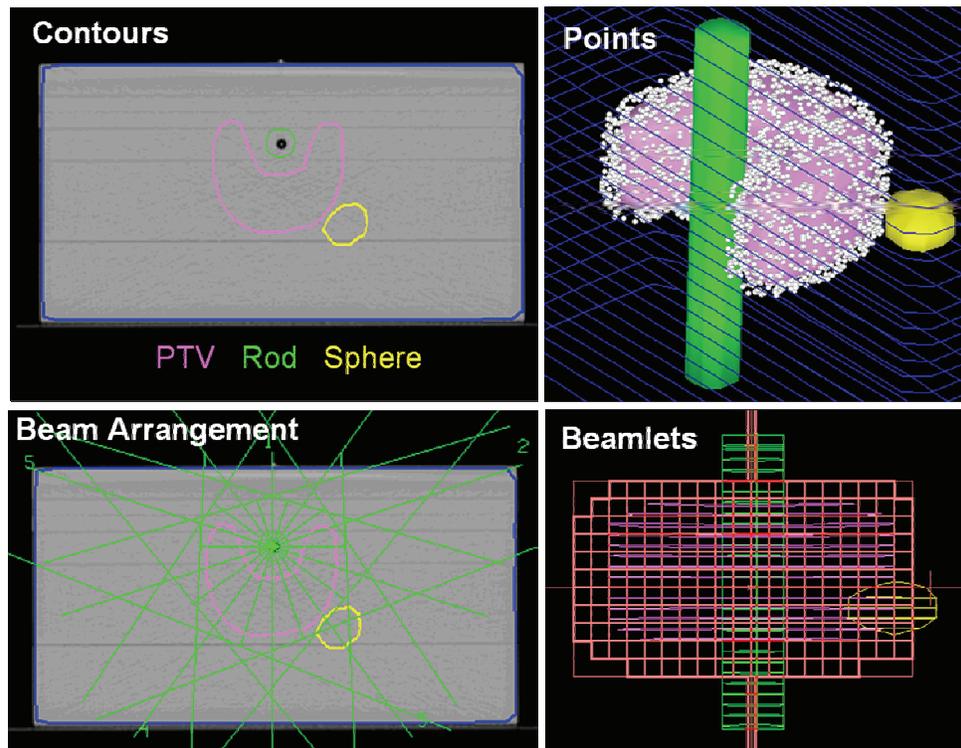


Figure I-4. An example of some of the steps involved in inverse-planned IMRT. Upper-left: Contouring on CT, Upper-right: placing points in regions of interest, Lower-left: setting up beam directions, and Lower right: defining the beamlet grid.

this point placement step is shown in the upper-right of Figure I-4. The regions of interest used in inverse planned IMRT must be comprehensive. The optimization system will only be aware of the regions of interest that are present in the data and will not use “common sense” to place the dose outside of the important regions unless they are specified. Because of this, accurate contouring and 3D representation is critical to achieve a high quality treatment plan.

4. *Choosing the number and orientation of the radiation beams and dividing them into theoretical beamlet grids.*

After defining the patient geometrical representation, the radiation beams must be placed around the patient. While beam orientation optimization is a current field of research²⁵⁻³⁰, most centers rely on experienced dosimetrists to choose the number and orientation of beams to be used. In many IMRT cases, 7 to 9 axial beams are placed around the patient. These numbers of beams are often sufficient to provide the necessary degrees of freedom to make beam orientation optimization less essential, and therefore beams can be equally spaced around the patient. Once the beam directions are chosen, each beam is divided into theoretical beamlet grids. Usually, the beamlet grids coincide with the width of the MLC leaves. The most common beamlet sizes are 1 cm x 1 cm and 0.5 cm x 0.5 cm, although larger beam segments can also be defined. The lower half of Figure I-4 shows the beam placement and definition of 0.5 cm x 0.5 cm beamlets for the solid water CT phantom example.

5. *Calculating the dose from each of the unit intensity beamlets to each of the points of interest.*

Once the beamlets are defined, a dose calculation is performed to determine the dose contribution from each unit intensity beamlet to each of the points defined in the regions of interest. This creates a sizeable matrix, which is used during the optimization process to calculate the dose and help optimize the beamlet intensities. The University of Michigan utilizes a convolution superposition algorithm derived from the work of Mackie *et al.*³¹ to perform beamlet dose calculations.

6. *Creating an objective function, or mathematically stating the treatment goals in penalty form based on their importance to a successful treatment.*

The objective function, or cost function, is an integral part of any optimization process. It puts the treatment goals from the radiation oncologist into mathematical form. In most cases, a penalty or cost is assigned to violating specific dosimetric goals. The more important the goal, such as minimum dose to the target, the higher the penalty. Each penalty is a component of the cost function, or a costlet. Adding the penalties and weighting them according to importance results in a weighted sum cost function. This is the most common form of cost function in use in IMRT optimization. However, recent research has also led to the emergence of multi-criteria optimization methods, such as lexicographic ordering (LO), which sorts the costlets according to priority levels^{32,33}. In LO, the most important goals are met first and constrained, and then the less important goals are attempted, and so on. This way, the most important goals are not sacrificed for less important ones, and the optimization system does not waste time making tradeoffs between objectives that are not at the same priority level.

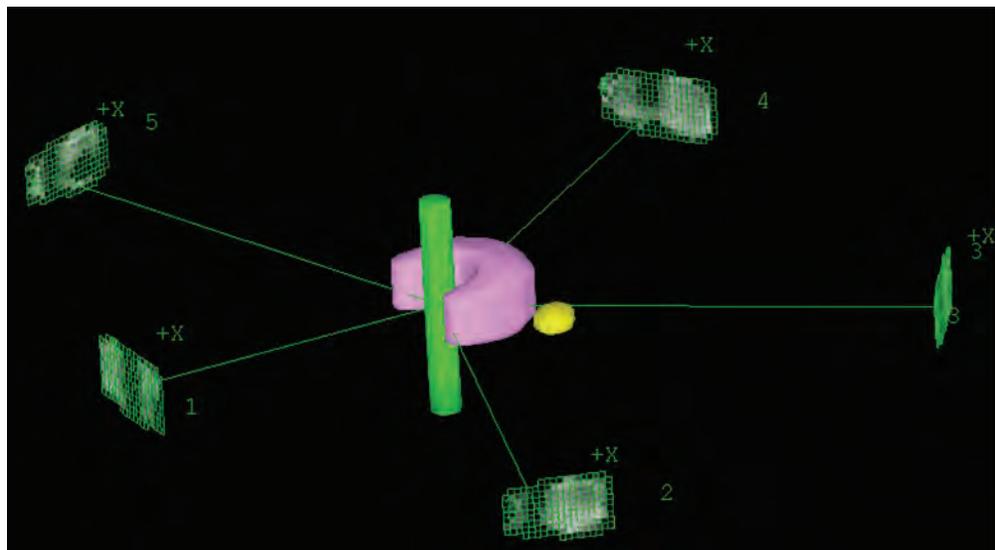
A summary of different cost function components available in the University of Michigan's optimization software is given by Kessler *et al.*³⁴ Treatment goals can be specified by dose-based metrics such as minimum, maximum, and mean dose as well as volume based metrics that allow only certain volumes of tissue to receive greater than or less than a certain dose. In addition, different radiobiological models, such as the tumor control probability (TCP)³⁵ and normal tissue complication probability (NTCP)³⁶ can also be used as treatment goals.

7. *Optimizing the beamlet intensities to best meet the goals stated in the objective function.*

After the treatment goals are laid out in the objective function, the beamlet intensities that best achieve the treatment goals must be chosen. This is done through a process called optimization. In the case of a weighted sum cost function, the beamlets are changed until the total cost function penalty, or simply cost, is minimized. Many optimization algorithms are available to solve these types of problems, with the most common being a type of gradient descent

algorithm, such as the quasi-Newton method, which is used at the University of Michigan. Care must be taken when using a gradient descent algorithm because it may not be able to find the global solution of the cost function. Depending on the definition of the cost function, multiple local minima may be present in the solution space. A gradient descent algorithm will find the closest local minima to its starting point and therefore may not achieve the global minimum. In these cases, a slower, but more reliable global search algorithm, such as simulated annealing, can be used. This is a stochastic search algorithm that can escape local minima to find the global solution. A more detailed explanation of the optimization algorithms at the University of Michigan is given in Chapter II.

Regardless of the optimization method chosen, the output of the optimization process should be the set of beamlet intensities that best meet the goals laid out by the cost function. The 3D visualization of an optimized plan is shown in Figure I-5.



8. Sequencing the leaves of the MLC in order to deliver the desired intensity-modulated fields.

Once the optimal beamlet intensity patterns are found, a technique for delivering them by the multi-leaf collimator must be devised. This process is referred to as *leaf sequencing*. The treatment machine can deliver intensity modulated beams

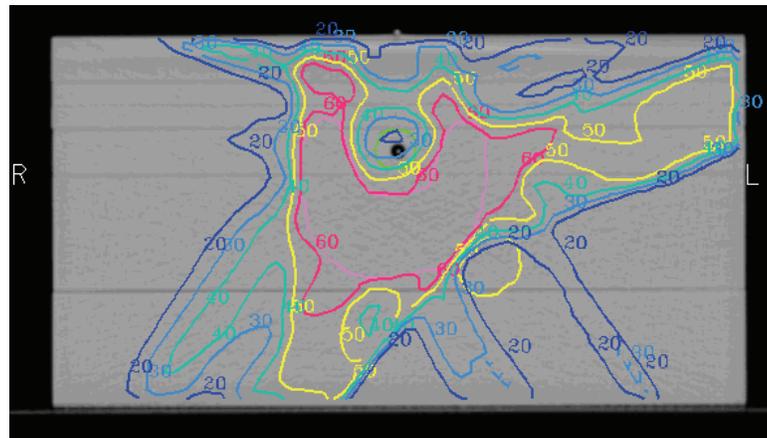
in dynamic or static mode (or both), depending on the manufacturer³⁷⁻⁴⁰. In dynamic mode, the leaf sequencing process generates a series of trajectories for each of the MLC leaves to travel in order to deliver the desired intensity pattern while the radiation beam is on. In static mode, the leaf sequencing process generates a series of shapes or segments that can be made and delivered. In this mode, the beam is turned off while the leaves move to each new segment position.

9. *Performing a final 3D calculation of the dose delivered to the patient for final evaluation by the radiation oncologist.*

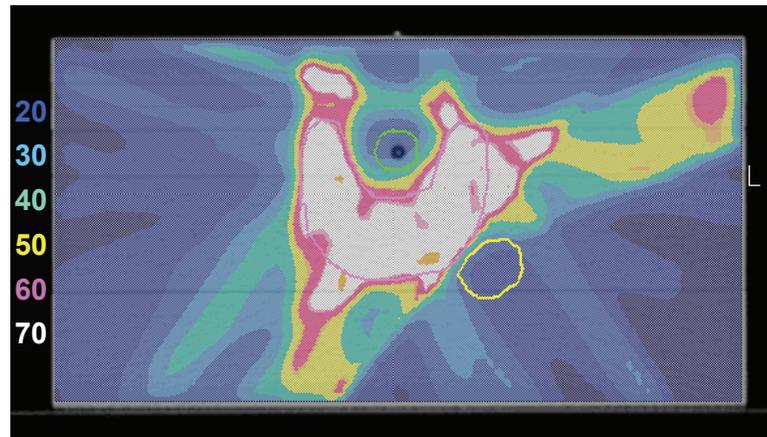
In most IMRT treatment planning systems, there are some differences between the way that dose calculations are performed for beamlet optimization to create the dose matrix versus 3D conformal planning. Depending on the optimization software, these differences include: (i) approximations, such as ignoring scatter dose, are made in the beamlet dose calculations to improve speed, (ii) the dose evaluation points are not located in the same voxel positions that are used in the final 3D dose calculations due to sampling points randomly or sparsely, and (iii) the initial dose calculation cannot incorporate delivery effects based on the position of the machine MLC. To assess the impact of these differences on the final treatment plan, some institutions perform a final 3D dose calculation on the patient CT scan and evaluate the dose using the full dose calculation and the full complement of voxel data at an adequate resolution of all structures. It is also possible to include the leaf sequence that will be used in the delivery to calculate any perturbations of the dose that occur due to the treatment delivery parameters, such as MLC positions.

This final dose calculation is another layer of quality assurance in the inverse planning process because it calculates dose to all of the voxels in the patient scan and therefore is not subject to the undersampling that can occur in the dose-to-points matrix. It also allows the planner and physician to visualize the full 3D dose distribution on the patient data to determine whether or not any dose or point sampling approximations or an omission of an important region of interest resulted in undesirable dose deposition. The dose distributions can be evaluated by using isodose lines and colorwash displays, which are overlaid on the patient's CT scan along with the treatment planning contours. These types of

displays are also available in many optimization systems and can be compared with the data from the final dose calculation. In many centers, the final approval of the plan by the radiation oncologist is performed on this final dose calculation by analyzing the dose distributions, dose metrics to relevant anatomy, and the cumulative dose-volume histograms calculated from the 3D calculation data.



(a)



(b)

Figure I-6. (a) Isodose contours and (b) a colorwash dose display for the final IMRT dose calculation with the colorscale given in Gy

An isodose and colorwash display on the solid water phantom is shown in Figure I-6. In this case, there is a “hotspot” of dose outside the target volume near the surface of the phantom. Features such as this can appear in the final calculation if the normal tissue is not properly sampled or the minimization of dose to the normal tissue is not specified in the cost function.

10. Performing quality assurance on the radiation plan to make sure it conforms to certain dosimetric standards.

Because IMRT is a recent technology, several quality assurance procedures are usually performed on each new IMRT plan. While this varies for different treatment centers, patient-specific IMRT QA usually consists of an overall verification of the dose to be delivered by a composite ion chamber measurement as well as an individual verification of each of the treatment beams through film or other 2D measurement devices. This level of patient-specific QA was originally performed because of the potential for errors to be introduced at different stages in the multi-stage IMRT planning process, and it continues to be necessary according to published reports of IMRT QA failures and delivery errors^{41,42}.

The goal of patient-specific IMRT QA is to verify that dose differences between the planned and delivered treatments agree within a certain tolerance. Some sources of discrepancy between calculated and measured plans are (i) errors in the delivery sequence by the linear accelerator or MLC, (ii) artifacts of the delivery due to the dosimetric characteristics of the MLC that are not modeled by the dose calculation, (iii) limitations of the dose calculation algorithm or leaf sequencing process, and (iv) experimental errors during measurements or imaging processing⁴³. It is generally difficult to identify a single source of error between measurements and calculations unless the error is large and obvious. IMRT QA aims to identify these large errors before they are propagated to the patient and attempts to disqualify plans in which the composite errors caused by the combination of the above factors is outside of the set tolerance level. If the IMRT passes QA, it moves into the treatment delivery queue. Otherwise, the discrepancies are analyzed and a decision is made to either alter the IMRT plan and re-perform QA or to deliver an alternate 3D-CRT plan.

11. Delivering the radiation treatment in fractions to the patient.

Once the patient begins IMRT treatment, care must be given to ensure that the very conformal treatment plan is delivered accurately. This requires careful patient setup and immobilization, to minimize any random or systematic setup errors. In some cases, the patient's anatomical or physiological properties may change during treatment, making the IMRT plan less optimal and requiring an

adaptation of the plan to the new input data. This process is becoming a popular research area, called *adaptive radiation therapy*⁴⁴⁻⁴⁸.

The above steps demonstrate the complex IMRT treatment planning and delivery process. This process can lead to very conformal radiation treatments that improve patient tolerance and outcome of radiation therapy. The advantage of IMRT over conventional and conformal radiation therapy has been the subject of many publications, although limited data on any overall survival benefit of IMRT over conformal therapy is available at this time. The patient groups that have been the most positively impacted by IMRT are head and neck and prostate cancer patients. Many studies have shown improved conformality and salivary function with IMRT in the head/neck, which has resulted in improved quality of life for these patients⁴⁹⁻⁵². Similarly, IMRT has improved normal structure sparing and allowed for safe dose escalation to the prostate⁵³⁻⁵⁶. Other studies have suggested that IMRT may also provide an advantage in the treatment of brain, lung, and other cancers⁵⁷⁻⁵⁹. This revolutionary technique is beginning to play a leading role in the targeting of biologically defined and possibly dynamic target volumes⁶⁰⁻⁶⁴. The use of IMRT to “paint” dose according to the specific properties of the local tumor and normal tissue may enable better targeting of tumor biology and improve tumor control.

I.C. The Fundamental Problem in IMRT

As described in the previous section, solving an inverse IMRT problem requires the use of hundreds to thousands of independent variables and a potentially large set of objectives or planning goals. The discrete nature of the beamlet solution potentially allows for the intensity modulated beams to take on nearly any shape or distribution. To the optimization engine, the shape of the optimized beam does not matter if the dosimetric objectives are achieved. However, compared to conventional and conformal beams, the inverse planned IMRT beams can be very complex and unintuitive. Figure I-7 shows a comparison between a conformal prostate treatment and a prostate IMRT plan. On the upper left is a common “4 field box” treatment for prostate cancer. The fields are shaped around the prostate with the multi-leaf collimator, and the beams are uniform across each field. The resulting dose distribution calculated in the UPlan treatment planning system is shown below in the lower left. On the upper right is a 7-field IMRT plan for the same patient. One of the intensity-modulated fields is shown in

the inset along with the corresponding dose distribution. While the IMRT plan is superior in terms of sparing the bladder and rectum from high doses of radiation, the beams are much more complex and the dose distribution is much more heterogeneous than that of the conformal plan. Most of the current prescription doses and normal tissue dose limits

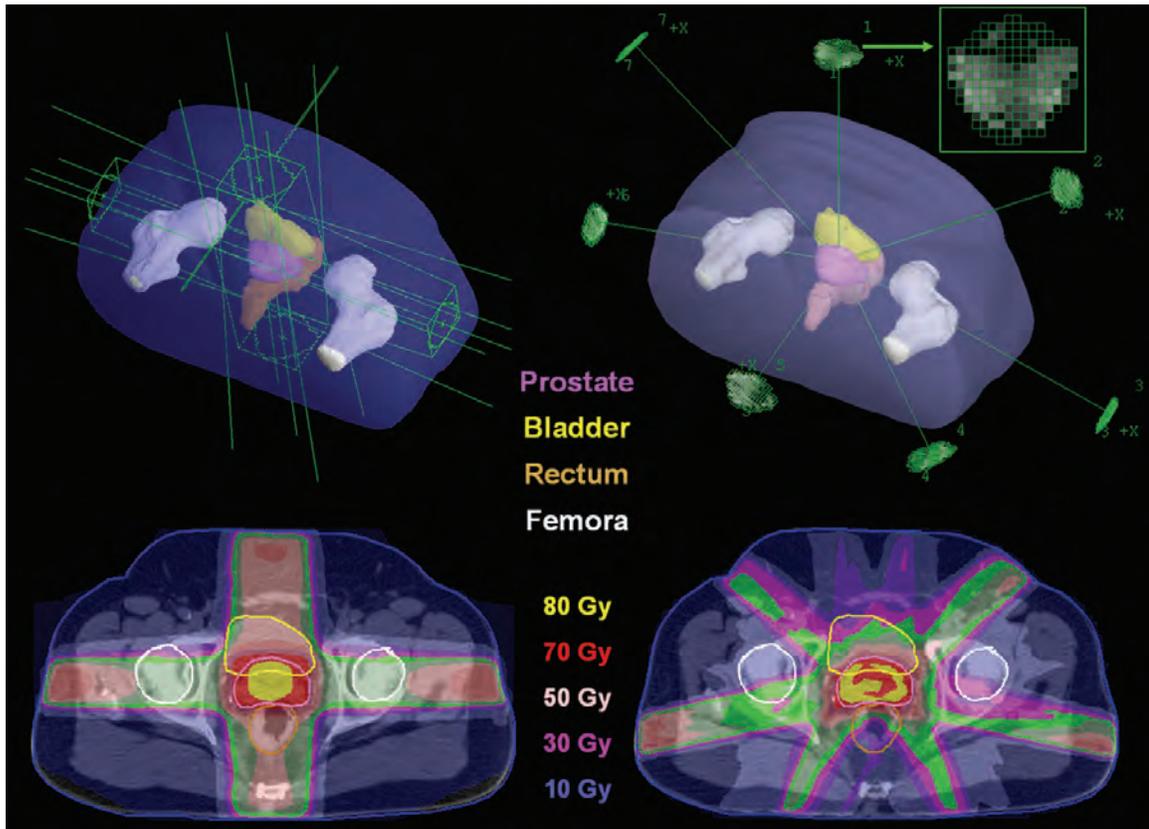
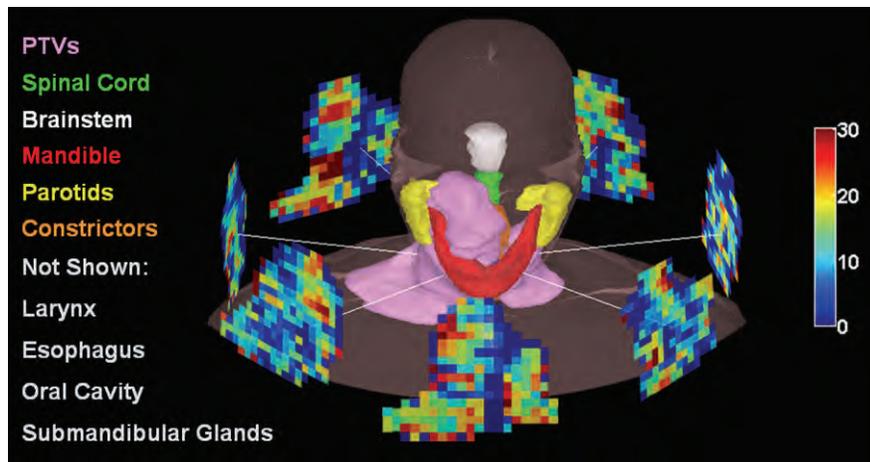


Figure I-7. Clockwise from top left: a conformal “4 field box” prostate treatment; a 7 beam inverse planned IMRT treatment with an inset showing the modulation in beam 1; the dose colorwash for the IMRT plan; the dose colorwash for the 3D conformal prostate plan. The prostate, bladder, rectum, and femora are represented by the surfaces and contours in the colors given, and the isodose color scale is also given.

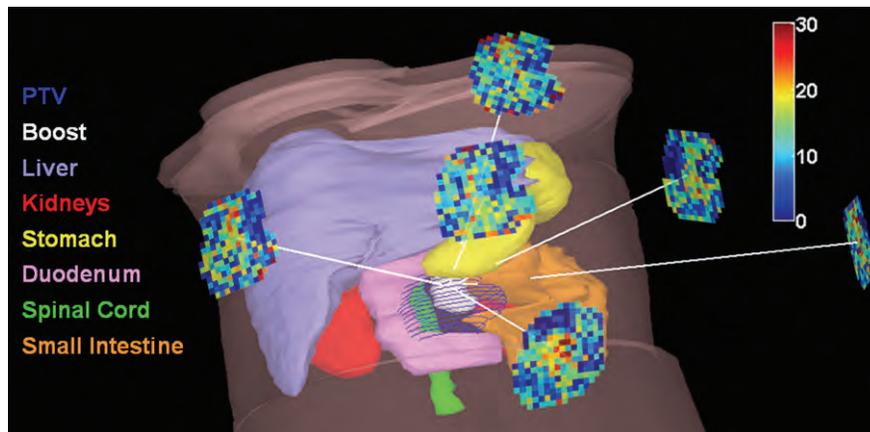
used in radiation therapy planning are based on doses that were calculated and delivered with conformal radiation therapy beams. Therefore, the significance of the difference between conformal beams and IMRT beams needs to be studied to determine the effect that the differences may have on the way patients are treated. The considerable increase in the complexity of the beam intensity distribution being delivered to the patient is a by-product of the IMRT treatment planning problem. IMRT cases commonly require the definition of 10-20 regions of interests and thousands of beamlets. Each of the regions of interest may then have multiple treatment goals associated with it.

This comprises a complicated cost function with many competing goals and a large number of independent variables to optimize. Figure I-8 shows a 3D representation of the complicated geometries of the head and neck and pancreas as well as the equally complicated beamlet patterns produced by inverse IMRT planning. These beam intensities are a drastic contrast to the uniform intensity beams that would be used in conformal therapy, making it difficult to judge their efficacy for treatment planning.

Treating a patient with very complex beams causes several concerns. First, high degrees of complexity, or modulation, can make it difficult to sequence the multi-leaf collimator without approximations, leading to degradation in plan quality. If leaf



(a)



(b)

Figure I-8. An example of the complicated geometries defined for (a) head and neck cancer and (b) pancreatic cancer treatment, and the highly complex beamlet intensity patterns that are created by inverse IMRT planning.

sequencing is accurate, then the machine sequence can be very complicated, possibly leading to errors in the delivery itself. Beyond these potential errors, a complicated sequence takes the machine much longer to deliver, due to the fact the majority of beam area is blocked by the MLC at any one time during the delivery⁶⁵. This delivery inefficiency reduces the efficiency of the clinic and causes undesirable wear-and-tear on the treatment machine. More importantly, it delivers extra dose to the patient due to the increased transmission and leakage of radiation from the head of machine and through the MLC during the lengthy delivery time. This extra dose can increase the risk of inducing secondary malignancies⁶⁶⁻⁶⁸. In addition, prolonging the time it takes to deliver a fraction of treatment may have a negative impact on tumor control by decreasing the effective dose rate⁶⁹⁻⁷¹.

I.D. Review of IMRT Complexity Reduction Research

The above issues have led to a variety of possible improvements to the way IMRT is planned, sequenced and eventually delivered. The IMRT complexity problem has been addressed at points both during and after the optimization process. Most early research was aimed at developing leaf sequencing algorithms that improved the agreement with optimized plans and maximized delivery efficiency. In many cases these have been competing goals, and much work in the area has been focused on a compromise between the two⁷²⁻⁷⁶. Because of their use of discrete intensity levels and improved efficiency, newer leaf sequencing algorithms can produce accurate deliveries with many fewer MU than the original algorithms. However, if there is unwanted modulation or noise present in the planned fields, leaf sequencing is not designed to remove or correct for these issues. The presence of high intensity peaks and large beamlet to beamlet fluctuations in the field, while deliverable, still require long times and many MU to deliver. Filtering and smoothing of the optimized intensity maps has also been applied during the sequencing process to help improve delivery efficiency. However, the utility of post-optimization smoothing is limited because it can quickly degrade plan quality. Iterative smoothing is usually necessary to minimize the loss of necessary modulation.^{65,77} Sun *et al.* have introduced a smoothing algorithm based on the location of the regions of interest so that only beamlets away from target and normal tissue interfaces and overlap regions are smoothed⁷⁸. Coupled with re-optimization of segment weights, this method has shown an ability to reduce MU significantly in brain and prostate cases, although its utility is limited in more complicated geometries.

Since post-optimization interventions have limitations due to their potential negative effects on plan quality, several other methods have been developed to decrease IMRT complexity during the optimization process. There have been two distinct approaches to incorporating interventions inside the optimization loop. The first is incorporation directly into the objective function, and the other is applying the interventions after each cycle or iteration. Systems that smooth at each iteration without taking into account the effect of the smoothing procedure on the objective function value are generally incapable of distinguishing between desirable and undesirable gradients, thus likely degrade the plans in areas where gradients are necessary^{65,79}. Webb *et al.* have shown that median filtering (replacing the center beamlet value in the smoothing window with the median beamlet value in that window) during the optimization process could improve delivery efficiency without significant degradation to the plan cost value or DVHs in simple geometries, but the method resulted in unacceptable plan degradation in more complex geometries⁸⁰.

An alternate method that is now widely used is the direct optimization of the shapes and weights of the actual segments that will be delivered^{81,82}. This technique, called direct aperture or direct segment optimization, can be implemented in several different ways with varying segments per beam. The advantages are fewer segments and improved delivery efficiency. However, irregular segments are possible and can still be subject to delivery errors and artifacts. In addition, there have not been any fair comparison studies published between the direct aperture methods and full intensity map modulation. Some preliminary studies suggest that direct aperture methods may not be able to satisfy the tight planning constraints of complex cases to same dosimetric level as full intensity modulated plans.

Applications that can reduce modulation as a part of the cost function can be effective at improving delivery efficiency without adversely affecting the plan quality. Spirou *et al.* have presented a comparison of smoothing inside and outside of the cost function. They found that smoothing as part of the cost function is superior in terms of producing sharper dose gradients, better dose homogeneity, and better critical organ sparing, especially for more complex cases⁷⁹. However, the filtering operation was only applied in the direction of MLC travel, which can cause undesirable fluctuations in the direction perpendicular to the MLC travel. A disadvantage of incorporating smoothing into the cost function is a difficulty in assessing the tradeoffs between the smoothness criteria and the target and normal tissue objectives. For example, despite inclusion into

the cost function so that “the filtering operation is a relatively small perturbation to the parameter estimation process,” Llacer *et al.* shows that filtering can deteriorate the PTV coverage⁸³.

The IMRT complexity problem lacks a solution that adequately and simultaneously addresses the problems of planning efficiency, delivery accuracy and efficiency, and geometric robustness, while maintaining the high dosimetric quality and benefits of full intensity modulation. Since the optimization engine does not care about the features of the intensity distribution beyond meeting the objectives, it is unknown whether or not there are simpler fluence patterns that could achieve similar dosimetric results. Optimization problems can be nearly degenerate, and in the case of IMRT, that means that there may be many beamlet patterns that can achieve almost the same minimum cost value. This dissertation is focused on (i) investigating how much beam complexity is necessary, and aims to (ii) develop methods to control and minimize beam complexity, while still achieving the superior dosimetric performance possible with intensity modulation.

I.E. Dissertation Overview

This chapter has introduced radiation therapy and more specifically, inverse-planned intensity modulated radiation therapy and some of its potential limitations. Although IMRT is an important treatment modality in radiation oncology, the fundamental feature of IMRT—intensity modulation—can lead to several negative characteristics. Before trying to remedy these issues, we review some of the known sources of beam complexity and try to identify the extent to which factors such as point sampling, the optimization method, and other parameters affect beam complexity. Chapter II discusses in more detail the consequences of beam complexity, including its effect on delivery efficiency and accuracy, geometric sensitivity, and radiation biology. We also include a survey of some of the known sources of excessive beam modulation and identify and study several suspected causes of modulation.

Chapter III presents a new method to reduce the scale and complexity of the IMRT problem by optimizing mathematical surfaces instead of individual beamlet weights to represent the beam. Our implementation of this surface or basis function-based optimization for 3-D inverse IMRT planning is presented, and several examples are used to characterize basis function optimization and show its clinical potential—especially in cases with simple-to-moderate levels of geometric complexity.

Chapter IV returns to the beamlet paradigm to study the effect of placing new maximum intensity restrictions on beamlets or using modulation penalties during optimization. Placing maximum intensity restrictions on beamlet values during optimization prevents the optimization engine from allowing single beamlets to reach very high intensity levels (which decreases the delivery efficiency). Since beamlet restrictions are not necessarily a comprehensive solution to the IMRT complexity problem, we have developed several other cost function components, or costlets, that can be used to penalize modulation during optimization. Costlets that filter the IMRT beam and penalize the difference between the original and filtered beam and a costlet that penalizes an overall measure of modulation are studied. Chapter IV includes our design, implementation, and use of each of these modulation penalty costlets.

Chapters V and VI focus on the theory, implementation, and application of a new diffusion-based smoothing method called adaptive diffusion smoothing (ADS). Using an analogy to the diffusion equation, this method is unique in radiation therapy because it allows for preferential smoothing of different areas of the beam based on diffusion coefficients that are automatically defined for each beamlet. These coefficients can be customized based on the characteristics of different IMRT problems, and thus the applications of ADS are widespread. Chapter V discusses our development of the adaptive diffusion smoothing method and our implementation of this method into the optimization system infrastructure. We then present a characterization of the method and demonstrate its potential for improving IMRT delivery efficiency with minimal dosimetric impact. Chapter VI goes a step further to demonstrate several applications of ADS, including its use in accelerated partial breast irradiation, its use in promoting smoothness over key radiobiological targets where dose heterogeneity can be detrimental under geometric uncertainty, and its ability to help facilitate controlled decisions between smoothness and dosimetric goals with help of lexicographic ordering.

Chapter VII discusses the clinical impact of reducing IMRT beam complexity, focusing on the effect that smoothing procedures have on treatment planning, delivery, geometric sensitivity, and adaptive therapy. We study the improvement in treatment planning efficiency when using modulation penalties in the cost function, and the robustness of smoother plans to point sampling effects and geometric setup errors. Finally, we consider the potential impact of plan smoothness on the ability to correct for treatment course errors caused by geometric uncertainty. Specifically, we determine if it

is easier to re-plan treatments based on various forms of patient feedback when modulation penalties are used during optimization.

Finally, Chapter VIII summarizes our work in reducing and controlling IMRT beam complexity and discusses some of the work that can be done to make further research advances in this area.

References

- ¹ (American Cancer Society, www.cancer.gov, 2007).
- ² K. Leszczynski and S. Boyko, "On the controversies surrounding the origins of radiation therapy," *Radiother Oncol* **42**, 213-217 (1997).
- ³ B. A. Fraass and D. L. McShan, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F.H. van der Giessen, H.J. van Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 273-276.
- ⁴ R. Mohan, G. Barest, L. J. Brewster, C. S. Chui, G. J. Kutcher, J. S. Laughlin, and Z. Fuks, "A comprehensive three-dimensional radiation treatment planning system," *Int J Radiat Oncol Biol Phys* **15**, 481-495 (1988).
- ⁵ J. A. Purdy, J. W. Wong, and W. B. Harms, "Three dimensional radiation treatment planning system," Proceedings of IVth International Conference of the Use of Computers in Radiotherapy, Schevenigen, The Netherlands, 1987.
- ⁶ D. L. McShan, A. Silverman, D. M. Lanza, L. E. Reinstein, and A. S. Glicksman, "A computerized three-dimensional treatment planning system utilizing interactive colour graphics," *Br J Radiol* **52**, 478-481 (1979).
- ⁷ L. E. Reinstein, D. McShan, B. M. Webber, and A. S. Glicksman, "A computer-assisted three-dimensional treatment planning system," *Radiology* **127**, 259-264 (1978).
- ⁸ M. S. Huq, I. J. Das, T. Steinberg, and J. M. Galvin, "A dosimetric comparison of various multileaf collimators," *Phys Med Biol* **47**, N159-170 (2002).
- ⁹ L. C. Fang, R. Komaki, P. Allen, T. Guerrero, R. Mohan, and J. D. Cox, "Comparison of outcomes for patients with medically inoperable Stage I non-small-cell lung cancer treated with two-dimensional vs. three-dimensional radiotherapy," *Int J Radiat Oncol Biol Phys* **66**, 108-116 (2006).
- ¹⁰ D. E. Morris, J. D. Bourland, J. G. Rosenman, and E. G. Shaw, "Three-dimensional conformal radiation treatment planning and delivery for low- and intermediate-grade gliomas," *Semin Radiat Oncol* **11**, 124-137 (2001).
- ¹¹ B. Emami, "Three-Dimensional Conformal Radiation Therapy In Bronchogenic Carcinoma," *Semin Radiat Oncol* **6**, 92-97 (1996).
- ¹² J. Waldron, M. M. Tin, A. Keller, C. Lum, B. Japp, S. Sellmann, M. van Prooijen, L. Gitterman, R. Blend, D. Payne, F. F. Liu, P. Warde, B. Cummings, M. Pintilie, and B. O'Sullivan, "Limitation of conventional two dimensional radiation therapy planning in nasopharyngeal carcinoma," *Radiother Oncol* **68**, 153-161 (2003).
- ¹³ J. S. Belderbos, W. D. Heemsbergen, K. De Jaeger, P. Baas, and J. V. Lebesque, "Final results of a Phase I/II dose escalation trial in non-small-cell lung cancer using three-dimensional conformal radiotherapy," *Int J Radiat Oncol Biol Phys* **66**, 126-134 (2006).
- ¹⁴ J. Armstrong and C. McGibney, "The impact of three-dimensional radiation on the treatment of non-small cell lung cancer," *Radiother Oncol* **56**, 157-167 (2000).
- ¹⁵ D. E. Morris, B. Emami, P. M. Mauch, A. A. Konski, M. L. Tao, A. K. Ng, E. A. Klein, N. Mohideen, M. D. Hurwitz, B. A. Fraas, M. Roach, 3rd, E. M. Gore, and J. E. Tepper, "Evidence-based review of three-dimensional conformal radiotherapy for

- localized prostate cancer: an ASTRO outcomes initiative," *Int J Radiat Oncol Biol Phys* **62**, 3-19 (2005).
- 16 A. Brahme, "Optimization of conformation and general moving beam radiation therapy techniques," *Proceedings of IVth International Conference on the use of Computers in Radiotherapy, Holland, The Netherlands, 1987.*
 - 17 S. Webb, "Optimization by simulated annealing of three-dimensional conformal treatment planning for radiation fields defined by a multileaf collimator," *Phys Med Biol* **36**, 1201-1226 (1991).
 - 18 S. Webb, "Optimisation of conformal radiotherapy dose distributions by simulated annealing," *Phys Med Biol* **34**, 1349-1370 (1989).
 - 19 W. D. Powlis, M. D. Altschuler, Y. Censor, and E. L. Buhle, Jr., "Semi-automated radiotherapy treatment planning with a mathematical model to satisfy treatment goals," *Int J Radiat Oncol Biol Phys* **16**, 271-276 (1989).
 - 20 T. Holmes, T. R. Mackie, D. Simpkin, and P. Reckwerdt, "A unified approach to the optimization of brachytherapy and external beam dosimetry," *Int J Radiat Oncol Biol Phys* **20**, 859-873 (1991).
 - 21 Y. Censor, M. Altschuler, and W. Powlis, "A computational solution of the inverse problem in radiation-therapy treatment planning," *Appl Math Comput* **25**, 57-87 (1988).
 - 22 T. Bortfeld, J. Burkelbach, R. Boesecke, and W. Schlegel, "Methods of image reconstruction from projections applied to conformation radiotherapy," *Phys Med Biol* **35**, 1423-1434 (1990).
 - 23 "Intensity-modulated radiotherapy: current status and issues of interest," *Int J Radiat Oncol Biol Phys* **51**, 880-914 (2001).
 - 24 J. M. Balter and M. L. Kessler, "Imaging and alignment for image-guided radiation therapy," *J Clin Oncol* **25**, 931-937 (2007).
 - 25 D. Craft, "Local beam angle optimization with linear programming and gradient search," *Phys Med Biol* **52**, N127-135 (2007).
 - 26 X. Wang, X. Zhang, L. Dong, H. Liu, Q. Wu, and R. Mohan, "Development of methods for beam angle optimization for IMRT using an accelerated exhaustive search strategy," *Int J Radiat Oncol Biol Phys* **60**, 1325-1337 (2004).
 - 27 Y. Li, J. Yao, and D. Yao, "Automatic beam angle selection in IMRT planning using genetic algorithm," *Phys Med Biol* **49**, 1915-1932 (2004).
 - 28 A. H. Baydush, L. B. Marks, and S. K. Das, "Penalized likelihood fluence optimization with evolutionary components for intensity modulated radiation therapy treatment planning," *Med Phys* **31**, 2335-2343 (2004).
 - 29 D. Djajaputra, Q. Wu, Y. Wu, and R. Mohan, "Algorithm and performance of a clinical IMRT beam-angle optimization system," *Phys Med Biol* **48**, 3191-3212 (2003).
 - 30 A. Pugachev and L. Xing, "Computer-assisted selection of coplanar beam orientations in intensity-modulated radiation therapy," *Phys Med Biol* **46**, 2467-2476 (2001).
 - 31 T. R. Mackie, J. W. Scrimger, and J. J. Battista, "A convolution method of calculating dose for 15-MV x rays," *Med. Phys.* **12**, 188-196 (1985).

- ³² K.-W. Jee, D. L. McShan, and B. A. Fraass, "Lexicographic Ordering: Intuitive Multicriteria Optimization for IMRT," *Phys Med Biol* **52**, 1845-1861 (2007).
- ³³ J. J. Wilkens, J. R. Alaly, K. Zakarian, W. L. Thorstad, and J. O. Deasy, "IMRT treatment planning based on prioritizing prescription goals," *Phys Med Biol* **52**, 1675-1692 (2007).
- ³⁴ M. L. Kessler, D. L. McShan, M. Epelman, K. A. Vineberg, A. Eisbruch, T. S. Lawrence, and B. A. Fraass, "Costlets: a generalized approach to cost functions for automated optimization of IMRT treatment plans," *Optimization and Engineering* **6**, 421-448 (2005).
- ³⁵ A. Brahme and A. K. Agren, "Optimal dose distribution for eradication of heterogeneous tumours," *Acta Oncol* **26**, 377-385 (1987).
- ³⁶ J. T. Lyman, "Complication probability as assessed from dose-volume histograms," *Radiat Res Suppl* **8**, S13-19 (1985).
- ³⁷ S. V. Spirou and C. S. Chui, "Generation of arbitrary intensity profiles by dynamic jaws or multileaf collimators," *Med Phys* **21**, 1031-1041 (1994).
- ³⁸ M. L. Dirx, B. J. Heijmen, and J. P. van Santvoort, "Leaf trajectory calculation for dynamic multileaf collimation to realize optimized fluence profiles," *Phys Med Biol* **43**, 1171-1184 (1998).
- ³⁹ D. J. Convery and S. Webb, "Generation of discrete beam-intensity modulation by dynamic multileaf collimation under minimum leaf separation constraints," *Phys Med Biol* **43**, 2521-2538 (1998).
- ⁴⁰ T. R. Bortfeld, D. L. Kahler, T. J. Waldron, and A. L. Boyer, "X-ray field compensation with multileaf collimators," *Int J Radiat Oncol Biol Phys* **28**, 723-730 (1994).
- ⁴¹ N. L. Childress, R. A. White, C. Bloch, M. Salehpour, L. Dong, and Rosen, II, "Retrospective analysis of 2D patient-specific IMRT verifications," *Med Phys* **32**, 838-850 (2005).
- ⁴² A. M. Stell, J. G. Li, O. A. Zeidan, and J. F. Dempsey, "An extensive log-file analysis of step-and-shoot intensity modulated radiation therapy segment delivery errors," *Med Phys* **31**, 1593-1602 (2004).
- ⁴³ D. A. Low, "Quality assurance of intensity-modulated radiotherapy," *Semin Radiat Oncol* **12**, 219-228 (2002).
- ⁴⁴ J. Lof, B. K. Lind, and A. Brahme, "An adaptive control algorithm for optimization of intensity modulated radiotherapy considering uncertainties in beam profiles, patient set-up and internal organ motion," *Phys Med Biol* **43**, 1605-1628 (1998).
- ⁴⁵ C. Wu, R. Jeraj, G. H. Olivera, and T. R. Mackie, "Re-optimization in adaptive radiotherapy," *Phys Med Biol* **47**, 3181-3195 (2002).
- ⁴⁶ K. L. Lam, R. K. Ten Haken, D. Litzenberg, J. M. Balter, and S. M. Pollock, "An application of Bayesian statistical methods to adaptive radiotherapy," *Phys Med Biol* **50**, 3849-3858 (2005).
- ⁴⁷ H. Rehbinder, C. Forsgren, and J. Lof, "Adaptive radiation therapy for compensation of errors in patient setup and treatment delivery," *Med Phys* **31**, 3363-3371 (2004).
- ⁴⁸ D. Yan, F. Vicini, J. Wong, and A. Martinez, "Adaptive radiation therapy," *Phys. Med. Biol.* **42**, 123-132 (1997).

- ⁴⁹ R. K. Ten Haken and T. S. Lawrence, "The clinical application of intensity-modulated radiation therapy," *Semin Radiat Oncol* **16**, 224-231 (2006).
- ⁵⁰ P. M. Braam, C. H. Terhaard, J. M. Roesink, and C. P. Raaijmakers, "Intensity-modulated radiotherapy significantly reduces xerostomia compared with conventional radiotherapy," *Int J Radiat Oncol Biol Phys* **66**, 975-980 (2006).
- ⁵¹ P. Graff, M. Lapeyre, E. Desandes, C. Ortholan, R. J. Bensadoun, M. Alfonsi, P. Maingon, P. Giraud, J. Bourhis, V. Marchesi, A. Mege, and D. Peiffert, "Impact of intensity-modulated radiotherapy on health-related quality of life for head and neck cancer patients: matched-pair comparison with conventional radiotherapy," *Int J Radiat Oncol Biol Phys* **67**, 1309-1317 (2007).
- ⁵² E. H. Pow, D. L. Kwong, A. S. McMillan, M. C. Wong, J. S. Sham, L. H. Leung, and W. K. Leung, "Xerostomia and quality of life after intensity-modulated radiotherapy vs. conventional radiotherapy for early-stage nasopharyngeal carcinoma: Initial report on a randomized controlled clinical trial," *Int J Radiat Oncol Biol Phys* **66**, 981-991 (2006).
- ⁵³ B. S. Teh, C. M. Amosson, W. Y. Mai, J. McGary, W. H. Grant, 3rd, and E. B. Butler, "Intensity modulated radiation therapy (IMRT) in the management of prostate cancer," *Cancer Invest* **22**, 913-924 (2004).
- ⁵⁴ G. Luxton, S. L. Hancock, and A. L. Boyer, "Dosimetry and radiobiologic model comparison of IMRT and 3D conformal radiotherapy in treatment of carcinoma of the prostate," *Int J Radiat Oncol Biol Phys* **59**, 267-284 (2004).
- ⁵⁵ X. A. Li, J. Z. Wang, P. A. Jursinic, C. A. Lawton, and D. Wang, "Dosimetric advantages of IMRT simultaneous integrated boost for high-risk prostate cancer," *Int J Radiat Oncol Biol Phys* **61**, 1251-1257 (2005).
- ⁵⁶ A. Wang-Chesebro, P. Xia, J. Coleman, C. Akazawa, and M. Roach, 3rd, "Intensity-modulated radiotherapy improves lymph node coverage and dose to critical structures compared with three-dimensional conformal radiation therapy in clinically localized prostate cancer," *Int J Radiat Oncol Biol Phys* **66**, 654-662 (2006).
- ⁵⁷ U. Hermanto, E. K. Frija, M. J. Lii, E. L. Chang, A. Mahajan, and S. Y. Woo, "Intensity-modulated radiotherapy (IMRT) and conventional three-dimensional conformal radiotherapy for high-grade gliomas: does IMRT increase the integral dose to normal brain?," *Int J Radiat Oncol Biol Phys* **67**, 1135-1144 (2007).
- ⁵⁸ J. A. Christian, J. L. Bedford, S. Webb, and M. Brada, "Comparison of inverse-planned three-dimensional conformal radiotherapy and intensity-modulated radiotherapy for non-small-cell lung cancer," *Int J Radiat Oncol Biol Phys* **67**, 735-741 (2007).
- ⁵⁹ T. Girinsky, C. Pichenot, A. Beaudre, M. Ghalibafian, and D. Lefkopoulos, "Is intensity-modulated radiotherapy better than conventional radiation treatment and three-dimensional conformal radiotherapy for mediastinal masses in patients with Hodgkin's disease, and is there a role for beam orientation optimization and dose constraints assigned to virtual volumes?," *Int J Radiat Oncol Biol Phys* **64**, 218-226 (2006).
- ⁶⁰ X. Geets, M. Tomsej, J. A. Lee, T. Duprez, E. Coche, G. Cosnard, M. Lonneux, and V. Gregoire, "Adaptive biological image-guided IMRT with anatomic and functional imaging in pharyngo-laryngeal tumors: Impact on target volume delineation and dose distribution using helical tomotherapy," *Radiother Oncol* (2007).

- ⁶¹ A. Narayana, J. Chang, S. Thakur, W. Huang, S. Karimi, B. Hou, A. Kowalski, G. Perera, A. Holodny, and P. H. Gutin, "Use of MR spectroscopy and functional imaging in the treatment planning of gliomas," *Br J Radiol* (2006).
- ⁶² S. M. Bentzen, "Theragnostic imaging for radiation oncology: dose-painting by numbers," *Lancet Oncol* **6**, 112-117 (2005).
- ⁶³ J. M. Galvin and W. De Neve, "Intensity modulating and other radiation therapy devices for dose painting," *J Clin Oncol* **25**, 924-930 (2007).
- ⁶⁴ D. Thorwarth, S. M. Eschmann, F. Paulsen, and M. Alber, "Hypoxia dose painting by numbers: a planning study," *Int J Radiat Oncol Biol Phys* **68**, 291-300 (2007).
- ⁶⁵ R. Mohan, M. Arnfield, S. Tong, Q. Wu, and J. Siebers, "The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy," *Med. Phys.* **27**, 1226-1237 (2000).
- ⁶⁶ E. J. Hall and C. Wu, "Radiation induced second cancers: The impact of 3D-CRT and IMRT," *Int J Radiat Oncol Biol Phys* **56**, 83-88 (2003).
- ⁶⁷ E. J. Hall, "Intensity-modulated radiation therapy, protons, and the risk of second cancers," *Int J Radiat Oncol Biol Phys* **65**, 1-7 (2006).
- ⁶⁸ S. F. Kry, M. Salehpour, D. S. Followill, M. Stovall, D. A. Kuban, R. A. White, and Rosen, II, "The calculated risk of fatal secondary malignancies from intensity-modulated radiation therapy," *Int J Radiat Oncol Biol Phys* **62**, 1195-1203 (2005).
- ⁶⁹ J. Z. Wang, X. A. Li, W. D. D'Souza, and R. D. Stewart, "Impact of prolonged fraction delivery times on tumor control: A note of caution for intensity-modulated radiation therapy (IMRT)," *Int J Radiat Oncol Biol Phys* **57**, 543-552 (2003).
- ⁷⁰ H. Paganetti, "Changes in tumor cell response due to prolonged dose delivery times in fractionated radiation therapy," *International Journal of Radiation Oncology, Biology, Physics* **63**, 892-900 (2005).
- ⁷¹ J. F. Fowler, J. S. Welsh, and S. P. Howard, "Loss of biological effect in prolonged fraction delivery," *International Journal of Radiation Oncology, Biology, Physics* **59**, 242-249 (2004).
- ⁷² Y. Chen, Q. Hou, and J. M. Galvin, "A graph-searching method for MLC leaf sequencing under constraints.," *Med. Phys.* **31**, 1504-1511 (2004).
- ⁷³ S. M. Crooks, L. F. McAven, D. F. Robinson, and L. Xing, "Minimizing delivery time and monitor units in static IMRT by leaf-sequencing," *Phys. Med. Biol.* **47**, 3105-3116 (2003).
- ⁷⁴ M. Langer, V. Thai, and L. Papiez, "Improved leaf sequencing reduces segments or monitor units needed to delivery IMRT using multileaf collimators," *Med. Phys.* **28**, 2450-2458 (2001).
- ⁷⁵ E. Nioutsikou, J. L. Bedford, J. A. Christian, M. Brada, and S. Webb, "Segmentation of IMRT plans for radical lung radiotherapy delivery with the step-and-shoot technique," *Med. Phys.* **31**, 892-901 (2004).
- ⁷⁶ Y. Yang and L. Xing, "Incorporating leaf transmission and head scatter corrections into step-and-shoot leaf sequences for IMRT," *Int J Radiat Oncol Biol Phys* **55**, 1121-1134 (2003).

- ⁷⁷ L. Ma, "Smoothing intensity-modulated treatment delivery under hardware constraints," *Med. Phys.* **29**, 2937-2295 (2002).
- ⁷⁸ X. Sun and P. Xia, "A new smoothing procedure to reduce delivery segments for static MLC-based IMRT planning," *Med. Phys.* **43**, 2785-2794 (2004).
- ⁷⁹ S. Spirou, N. Fournier-Bidoz, J. Yang, C. Chui, and C. Ling, "Smoothing intensity-modulated beam profiles to improve the efficiency of delivery," *Med. Phys.* **28**, 2105-2112 (2001).
- ⁸⁰ S. Webb, D. J. Convery, and P. M. Evers, "Inverse planning with constraints to generate smoothed intensity-modulated beams," *Phys. Med. Biol.* **43**, 2785-2794 (1998).
- ⁸¹ B. A. Fraass and D. L. McShan, "Optimization of Conformal and IMRT Plans with Direct Segment Optimization," *Proceedings of XVth International Conference on the Use of Computers in Radiation Therapy*, Toronto, ON, Canada, 2007.
- ⁸² D. M. Shepard, M. A. Earl, X. A. Li, S. Naqvi, and C. Yu, "Direct aperture optimization: A turnkey solution for step-and-shoot IMRT," *Med. Phys.* **29**, 1007-1018 (2002).
- ⁸³ J. Llacer, N. Agazaryan, T. D. Solberg, and C. Promberger, "Degeneracy, frequency response and filtering in IMRT optimization," *Phys. Med. Biol.* **49**, 2853-2880 (2004).

CHAPTER II.

A REVIEW OF THE SOURCES AND CONSEQUENCES OF IMRT BEAM COMPLEXITY

Intensity modulated radiation therapy has been quoted as “*one of the most important technological advances in radiation therapy since the advent of the medical linear accelerator*”¹. This excitement is a by-product of the potential to create plans with inverse IMRT planning techniques that would not be realizable with conventional or conformal planning techniques. The number of degrees of freedom available in IMRT optimization has made it possible to partially automate the creation of complex plans that previously were not possible or required a great deal of time, skill and effort to plan with conformal techniques. Such treatments and improvements include (i) simultaneous boost treatments to increase dose to high-risk sub-volumes of the tumor without additional treatment fractions, (ii) sparing of critical normal tissues that are in close proximity to the tumor, and (iii) safe dose escalation of target volumes while maintaining acceptable doses to surrounding normal tissues. While IMRT shows dosimetric advantages in the planning of many of these treatments, there is a disproportionate increase in the complexity of the beams created with inverse planned IMRT versus those that may be forward or inverse-planned with conformal therapy. This large departure from accepted and understood conformal techniques has raised concerns about the possible consequences of IMRT beam complexity and questions about the sources of this complexity and the non-intuitive distribution of intensity that is observed in many IMRT beams.

This chapter first introduces the relevant details of treatment planning optimization system (UMOpt) used at the University of Michigan and throughout this dissertation. Then, we review the clinical implications and concerns encountered when treating patients with highly modulated IMRT plans and investigate several sources of beam complexity to help determine whether the high degrees of modulation are truly necessary or could be an unnecessary artifact of the involved IMRT treatment planning

process. The contents of this chapter are largely a review of previously published work regarding the potential disadvantages of highly modulated IMRT beams, and the reasons that IMRT beams can become complex. We include examples of many of these ideas from UMOpt and also investigate in greater detail several potential sources of IMRT beam complexity that have not been previously studied.

II.A. Overview of Treatment Planning Details and Methods

All of the IMRT treatment plan designed in this dissertation are planned for a 6 MV linear accelerator (Varian Medical Systems, 21EX) with 120 leaf multi-leaf collimator (MLC) equipped with 0.5 cm width central leaves and 1 cm width outer leaves. Treatment planning setup and execution is performed in the in-house developed treatment planning and optimization systems at the University of Michigan. These software packages, UMPlan and UMOpt²⁻⁸, give us a unique opportunity to study beam complexity in IMRT because they are not the usual commercially designed “black box” systems that are used for treatment planning and optimization in most treatment centers.

All imaging datasets and registration coordinate transformations are managed by UMPlan. Contouring and surface generation of the important regions of interest are performed in UMPlan along with beam placement and the definition of the beamlet grids. The placement of points is done with another in-house program called Points. This program is used to manage the discrete point placement in the regions of interest that has been contoured in UMPlan. Boolean operations can also be used to combine the UMPlan structures to make regions of interest in Points. Once the beamlet grids and points are defined, the large dose-to-points matrix is calculated. Currently, a convolution/superposition dose calculation algorithm derived from the work of Mackie *et al.*⁹ and subsequently optimized in house⁹ for beamlet calculations is used to calculate the dose to each point from each unit intensity beamlet. This dose data provides the input to the UMOpt optimization software. Inside UMOpt, the cost or objective function is designed and the beamlet weight optimization is performed. As mentioned in Chapter I, a summary of the different cost function components available in UMOpt is given by Kessler *et al.*⁷ UMOpt supports the specification of dose-based and volume-based metrics as well as different radiobiological models. Several different optimization algorithms have been implemented in UMOpt to minimize the weighted sum cost functions designed in UMOpt, including simulated annealing (a stochastic algorithm) and the quasi-Newton method (a gradient descent algorithm). These two methods are

discussed in more detail at the end of this chapter. UMOpt also employs a sequential quadratic programming algorithm to plan cases according to set priority levels instead of using a weighted-sum cost function. This capability is discussed in Chapter VI.

Leaf sequencing of clinical IMRT plans is done for the static or segmental method of delivery. Our static multi-leaf collimator (SMLC) leaf sequencer is an in-house application based on the method reported by Bortfeld *et al.*¹⁰. Delivery sequences allow up to 250 segments per beam, with the goal of achieving a correspondence between planned and delivered intensities of 1 %. Our leaf sequencer does not aim to minimize the number of segments or delivery time in an effort to improve deliver efficiency. This is partly due to the fact that Varian linear accelerators can efficiently deliver a large number of segments. Treatment machines developed by other manufacturers can be much more limited by the number of segments they can deliver in a reasonable amount of time. A final 3D dose calculation of the optimized plan can be performed in UMPlan using the optimized intensity maps or the data from the sequenced segments using the same convolution/superposition algorithm using from the point dose calculations.

II.B. Consequences of Highly Modulated Fields

The full potential of combining inverse planning and intensity modulation was far from well understood when the technology started to be explored. First, the careful placement of beams in 3D conformal therapy to spare important organs was no longer a necessity because the optimization program was expected to simply “turn off” the beamlets that would intersect the normal tissue. Therefore, beams could be aimed at the target from a variety of new directions. In order to take full advantage of this new capability and also to see what the true potential of intensity modulation could be, users planned IMRT cases with multiple beams equally spaced around the patient. In most cases, planners started with 9 beams—even in geometries that were normally treated with 3 or 4 beams. Right away, this increased the potential complexity of the plans compared to conformal therapy. There was concern about spreading so much extra dose around the patient instead of from only a few directions. Then, when planners analyzed the optimized beamlet patterns, they found complex patterns of intensity that were non-intuitive compared to conformal therapy. Beam complexity was present in the form of both high intensity spikes and seemingly random fluctuations in the beam intensity patterns. The complexity of some of these intensity patterns was called into

question by physicians and physicists alike. There was and is concern that treating patients with highly modulated intensity patterns may have negative consequences, such as inefficient and inaccurate treatment deliveries, increased sensitivity to geometric uncertainties, and possible biological risks arising from increased leakage dose and prolonged fraction times.¹¹⁻¹³ Much research effort has been spent analyzing the implications of highly complex IMRT fields, and it is unnecessary to reproduce that work here. In fact, previous research in this area serves as motivation for controlling the complexity of IMRT fields, which is the overall aim of this dissertation. Therefore, a brief survey of published work on the consequences of highly modulated fields will be summarized below, along with some practical examples generated with our radiotherapy IMRT optimization system, UMOpt.

II.B.1. Delivery Accuracy

One of the potential disadvantages of highly complex IMRT fields is the possibility for the delivered fields to vary significantly from the planned fields. Possible scenarios that could lead to this kind of problem include (i) approximations made during the leaf sequencing process that converts the beamlet intensities into a series of segments that can be delivered by the treatment machine and multi-leaf collimator (see Chapter I), (ii) delivery errors which occur due to complex delivery sequences, and (iii) delivery artifacts that occur but are not modeled by the beamlet dose calculations.

As we discussed in Chapter I, leaf sequencers cannot always perfectly translate the planned intensity patterns into sequences that can be delivered by the machine. In addition, some leaf sequencing algorithms incorporate filtering or smoothing procedures that can degrade the quality of the IMRT solutions. There are leaf sequencing algorithms that can accurately sequence even very complex intensity patterns, and plan degradation can be lessened by using sequencers that do not allow excessive approximations or smoothing during the sequencing process^{14,15}. Unfortunately, these more accurate sequencing algorithms can further reduce the delivery efficiency of IMRT fields and lead to potential complications that will be discussed in Section II.B and II.C. Leaf sequencing algorithms that simultaneously minimize (i) the difference between the planned and sequenced dose distribution and (ii) the monitor units and time required for delivery are available¹⁵⁻¹⁷. These algorithms may approach the current limit of what can be done solely inside the leaf sequencing process to improve delivery accuracy and efficiency without affecting the dosimetric quality of the plan.

Errors that can occur during the delivery of an IMRT field vary depending on the delivery method chosen. In step-and-shoot delivery, delivery errors include the “overshoot” effect¹⁸, missed segments, leaf position errors, redistribution of the monitor units over different segments than intended, and leaf motion during delivery¹⁹. The overshoot effect, which results in an over- and under-dose in the first and last segments of the delivery, respectively, is not caused by excessive modulation, but by a delay in the MLC controller, and can be corrected to a certain degree²⁰. However, the less predictable errors, such as leaf position inaccuracies during the segment sequence and missed segments are exacerbated by the use of the high dose rate (600 MU/min), high segment numbers, and low dose/segment values that are generally required to treat complex IMRT plans. Stell *et al.* have evaluated the delivery log-files for 91 SMLC IMRT plans and found that (i) MU errors increased as a function of dose rate, and (ii) the size of MU errors increased with the number of segments and dose rate²¹. In addition, for plans delivered at 600 MU/min, they reported that between 5 % and 23 % of the plan MUs were delivered during leaf motion that exceeded a 1 mm position tolerance. These types of errors tend to be difficult, if not impossible, to account for before treatment.

In dynamic IMRT delivery, most errors stem from the finite capabilities of the MLC controller and motors. There are complex interactions that deal with the leaf acceleration and velocity and corrections to leaf positions executed by the MLC controller. These positional errors lead to fluence errors that tend to be larger for higher MU treatments and higher intensity gradients²². In addition, dynamic sequences generally require approximately 20 % more monitor units but less time to deliver than static sequences^{23,24}. The dangers of low delivery efficiency in radiotherapy are discussed in sections II.A.2 and II.A.3.

An example of a delivery artifact that can occur, but is not modeled during the beamlet dose calculation is a partial underdosing under the “tongue” region of the MLC. This occurs due to the tongue and groove interlocking mechanism between parallel leaves on certain type of MLCs. When the leaf with the “tongue” is extended into the field without the leaf with the “groove” region, the area of the field under the tongue is partially blocked. Figure II-1 shows an example of this so-called “tongue and groove effect”. Figure II-1(a) shows a film measurement of an IMRT beam used to deliver prostate radiation therapy, (b) shows one delivery segment used in the field delivery, and (c) shows the dose difference display between convolution calculations of the field and the film measurement (calculations – film). In (c), several apparent areas of underdosing

can be seen in the same area where single leaves are extended across the field to make a complex segment. This is a common occurrence in IMRT delivery due to the complex segments that must be used to deliver the highly modulated beams. In IMRT, where strict localization of the patient is necessary, these regions of underdosing over the tumor volume could be a cause for concern if they are consistently in the same region of the tumor over a course of treatment.

The dosimetric effect of delivery artifacts, such as the tongue-and-groove effect, can be analyzed by recalculating the IMRT plan using data about the actual delivery sequence to be used in the plan. However, correcting for the dose discrepancies caused by the delivery artifacts is challenging. For example, if reoptimization of the beamlet intensities is required, the delivery sequences can change and introduce new

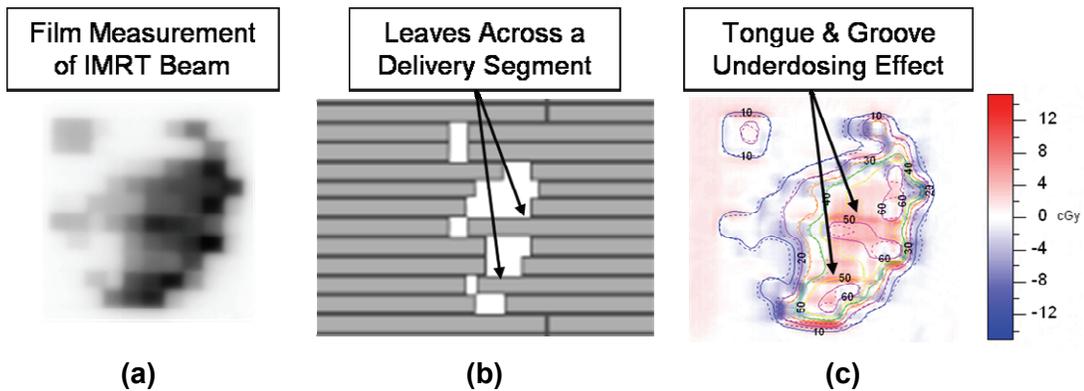


Figure II-1. An illustration of the tongue and groove effect that occurs in IMRT beam delivery. (a) A film measurement of an IMRT beam, (b) a segment of the delivery with leaves going across the width of the segment, and (c) a dose difference display of the calculations minus the film that shows underdose in the “tongue” region of the MLC.

artifacts. To some extent, these artifacts can be minimized by better leaf sequencing algorithms, but this may require an decrease in delivery efficiency²⁵⁻²⁷. Luan *et al.* have presented an algorithm to reduce tongue-and-groove errors, but this algorithm can increase the number of segments and delivery time by up to 30 % and introduce other unwanted approximations²⁷. Kamath *et al.* and Que *et al.* have also shown that the removal of tongue and groove error through leaf sequencing can significantly increase the number of segments required to deliver each IMRT field²⁵⁻²⁷.

II.B.2. Treatment Delivery Efficiency

As mentioned in the previous section, complex intensity patterns require complex leaf delivery sequences. These complex sequences or trajectories are comprised of numerous segments ranging in size and shape or numerous leaf starts, stops, and accelerations. Due to the high amount of radiation that is blocked by small and irregular segments, the linear accelerator must be in operation for a relatively long time to deliver the prescription dose compared to delivering the same dose with an open field. This means that IMRT treatments require more monitor units to deliver the fields as well as longer delivery times. As an example, Figure II-2 displays a variety of beamlet patterns for a 10 x 10 grid of 0.5 x 0.5 cm² beamlets. The beam on the left is a uniform beam of

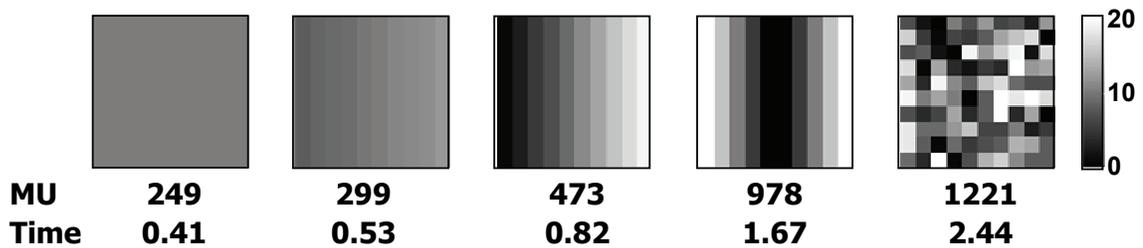


Figure II-2. An illustration of the decreasing delivery efficiency of IMRT fields as they become more complex. The MU required for step-and-shoot delivery and a conservative approximate time in seconds for each delivery are shown.

intensity 10, and the remaining beams have mean intensities of 10. The required MU and time for static MLC delivery are shown below each beamlet pattern. The times shown are in minutes and are conservative, but relatively accurate approximations of the delivery time from the leaf sequencing process. There is nearly a 1000 MU increase from a uniform beam to a beam of random intensity. These results agree with work done previously by Mohan *et. al.*,¹² which suggests the need to filter IMRT fluence patterns to improve delivery efficiency, but demonstrates the possible degradation that can occur in the dosimetric quality of the plan due to filtering the intensity profile.

The consequences of inefficient treatment delivery go far beyond just the simple increase in time required to treat each patient and the extra wear on the treatment machine and multi-leaf collimator. More importantly, the whole body dose delivered to the patient due to leakage and scattered radiation increases with MU, and the prolonged time required for each fraction delivery can both have radiobiological implications, which are discussed below.

II.B.3. Radiobiological Implications

Two distinct potential risks that are discussed when it comes to the complexity of IMRT beams. The first is the possibility of an increased risk of inducing a secondary cancer due to the increased whole body dose delivered to IMRT patients compared to conventional therapy^{28,29}. As mentioned previously, this increased dose is a result of increased leakage and scattered radiation that arises from the additional monitor unit requirements of complex IMRT treatments. Hall has reported that the incidence of secondary cancers in survivors could double with IMRT treatment compared to conventional treatment²⁸. He also points out that such a doubling would be unacceptable in pediatric patients, due to the already high incidence of secondary cancers observed in that population. Kry *et al.* have shown that the secondary cancer risk is highly dependent on the treatment energy used in IMRT as well as the treatment machine manufacturer and technique. However, the relative risk was always higher with an IMRT technique in the prostate compared to a conventional treatment because of the large increase in MU requirements²⁹.

Additional concerns arise from the uncertainty of the radiobiological effect of prolonging the delivery of single fraction. For tumors that repair sub-lethal damage quickly, the tumor control probability (TCP) could decrease as the effective dose rate decreases. Wang *et al.* have shown that the time required to deliver a single fraction of IMRT may significantly decrease cell killing in tumors with small α/β ratios or short tumor repair half-lives³⁰. They found that a 20 minute increase in delivery time for a prostate cancer case could decrease the TCP from 95 % to 73 %. Paganetti reached similar conclusions based on *in vitro* irradiation of a variety of different cancer cell lines, demonstrating the need to either reduce IMRT delivery time or compensate for the loss in tumor control³¹. Dose corrections to account for loss in biologically equivalent dose due to prolonged IMRT fractions were also suggested by Fowler *et al.* after *in situ* review of tumor repair half-lives in humans and animals³².

II.B.4. Geometric Uncertainty

In addition to requiring more time and MU to deliver, highly complex IMRT patterns may also suffer drawbacks relating to geometric sensitivity. Because of their conformal nature and high spatial modulation, IMRT plans tend to be much more sensitive to small geometric setup errors and internal patient motion than their 3D conformal counterparts.

Plans that include high gradient fluence areas may be even more sensitive to geometric uncertainties³³. Another concern is that the motion of the MLC combined with internal motion could result in parts of the tumor being entirely missed in some fractions³⁴. Seco *et al.* have demonstrated that segments that deliver only a few MU, which are common in complex IMRT treatments, are at an increased risk of this so-called interplay effect³⁵. Duan *et al.* have demonstrated that the dosimetric error observed in IMRT can approach 50 %, and that the dosimetric error increases more rapidly in beams with high intensity gradients than those dominated by more low intensity gradients across the field.

Most studies conclude that dosimetric errors lessen as beam complexity decreases, and the number of beams and fractions increase. However, significant concerns about underdosing aggressive tumor regions in individual fractions still remain. Tomé and Fowler have shown that underdosing of greater than 10 % of the prescribed dose in tumor volumes as small as 1 % can degrade the probability of tumor control³⁶. In addition, as more research is done on hypofractionation protocols, the geometric sensitivity of complex IMRT fields becomes even more significant. This leads to further concerns relating to radiobiology, including a possible loss in tumor control probability if parts of the tumor receive much less than their intended dose. Alternately, nearby organs-at-risk could also receive more dose than planned. While the impact of geometric sensitivity on normal tissues is complicated and depends on the treatment site and organ motion characteristics, the use of less complex fields should make it easier to predict, model, and account for these effects.

II.C. Sources of IMRT Beam Complexity

While the consequences of complex intensity patterns have been addressed in many studies, the underlying sources of IMRT beam complexity have been largely overlooked. A potential source of IMRT beam complexity that has been assumed, but not studied is the point sampling density of the regions of interest. Two other potential sources that have been suggested previously are the degeneracy of the optimization problem, and the optimization algorithm and objective function structure. In this section, we will (i) investigate the effect of discrete point sampling on IMRT beam complexity, (ii) review the literature and demonstrate examples of the degeneracy of the optimization problem, and (iii) review the literature and further study the impact of the optimization algorithm, optimization parameters and optimization cost function on beam complexity.

II.C.1. Point Sampling

High intensity and high frequency fluctuations may be a result of the discrete IMRT problem definition. In other words, the use of beamlets and point or voxel representations of anatomical structures (and the objectives) creates specific instances in which certain beamlet values can be exploited to cause improvements in the single objective function value for that instance of the problem. However, when the problem is changed slightly due to geometric uncertainty (or even point resampling), the complex plan may not be very robust. The following sections analyze the effects of both undersampling and random sampling on the complexity of IMRT fields.

II.C.1.1. Point Sub-Sampling

An investigation was performed on a test case (Figure II-3) to analyze the impact of point density on the complexity of the IMRT beam patterns. The test case geometry closely resembles the IMRT benchmark phantom, although an additional organ-at-risk (OAR2) has been simulated. The non site-specific IMRT benchmark has been accepted by all National Cancer Institute (NCI) funded cooperative groups and quality assurance offices as a minimum standard for an institution to be credentialed for use of IMRT in clinical trials. For this study, points were distributed in the phantom regions of interest at an

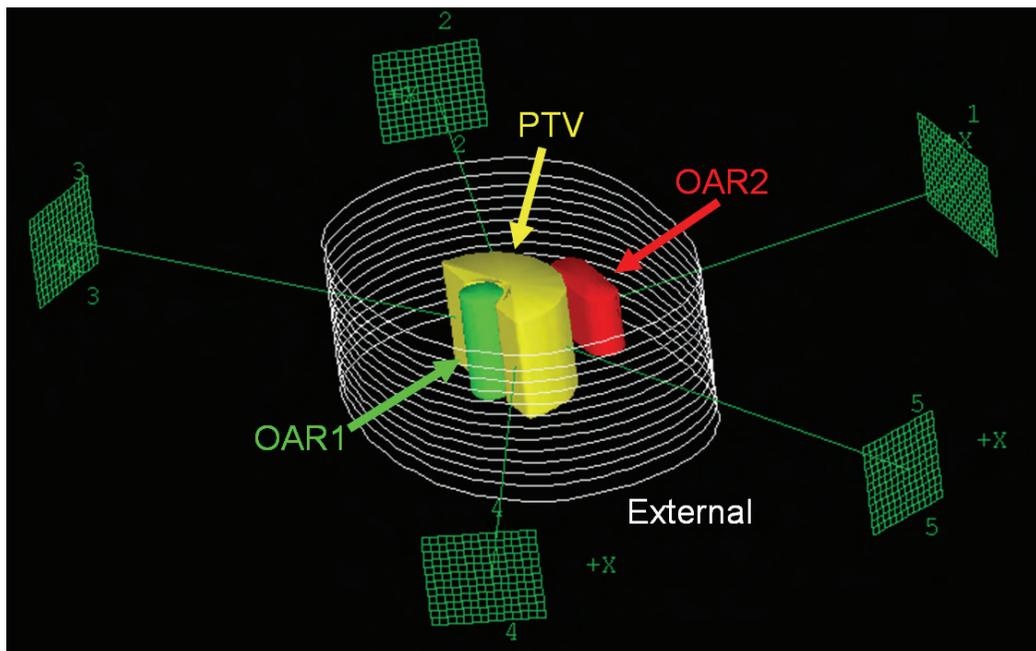


Figure II-3. The modified IMRT benchmark phantom designed for IMRT validation and testing.

average voxel size of .015 cm³ using a quasi-random point distribution algorithm. Then, for a series of optimization runs, the points were downsampled by factors of 2, 3, 4, and 5.

The results of the optimized cases were analyzed according to dosimetric quality (dose-volume histograms or dose metrics), delivery efficiency (monitor units), and measures of beam modulation and beam similarity. To give a measure of the beam modulation that is not dependent on the monitor units or leaf sequencing algorithm, we have developed a metric called the plan intensity map variation (PIMV), which is a measure of the overall modulation of the plan:

$$\text{PIMV} = \sum_{n=1}^{N_b} \left(\sum_{j=1}^{J-1} \sum_{k=1}^{K-1} \left[|I_{jk} - I_{j,k+1}| + |I_{jk} - I_{j+1,k}| + \dots \right] \right). \quad (\text{II.1})$$

Here, N_b is the number of beams in the plan, J is the maximum number of beamlets in the direction parallel to the motion of the MLC, K is the maximum number of beamlets in the direction perpendicular to the motion of the MLC, and I_{jk} is the intensity of the beamlet at the (j,k) grid position.

In addition to measuring the overall field modulation, it was also important to determine the extent to which the different point sampling parameters affect the optimized beamlet intensity pattern relative to the fully sampled beams. For a measure of the similarity between two intensity maps, the correlation coefficient for each intensity grid with respect to another optimized intensity grid can be calculated. Thus, for one beam, given an optimal intensity map A , and another optimal intensity map B , we can define the correlation coefficient as,

$$C_{I_A/I_B} = \frac{\sum_j \sum_k (I_{A,jk} - \bar{I}_A)(I_{B,jk} - \bar{I}_B)}{\sqrt{\left(\sum_j \sum_k (I_{A,jk} - \bar{I}_A)^2 \right) \left(\sum_j \sum_k (I_{B,jk} - \bar{I}_B)^2 \right)}}, \quad (\text{II.2})$$

where j and k are the dimensions of the intensity map, $I_{X,jk}$ is the intensity of grid element (j,k) for plan X , and \bar{I}_X is the mean intensity of grid I_X . The correlation coefficient may vary from -1.0 to 1.0. A value of 1.0 means that the two patterns are perfectly linearly

and positively correlated, while a value of -1.0 means that the two patterns are perfectly linearly and oppositely correlated. A high absolute number means there is a high level of correlation, while a small absolute number represents a weak correlation.

The resulting beamlet patterns from the optimization of each of the sub-sampled cases are shown in Figure II-4. Examination the beam intensity patterns reveals that the randomly appearing noise and modulation of the patterns increases as fewer points are

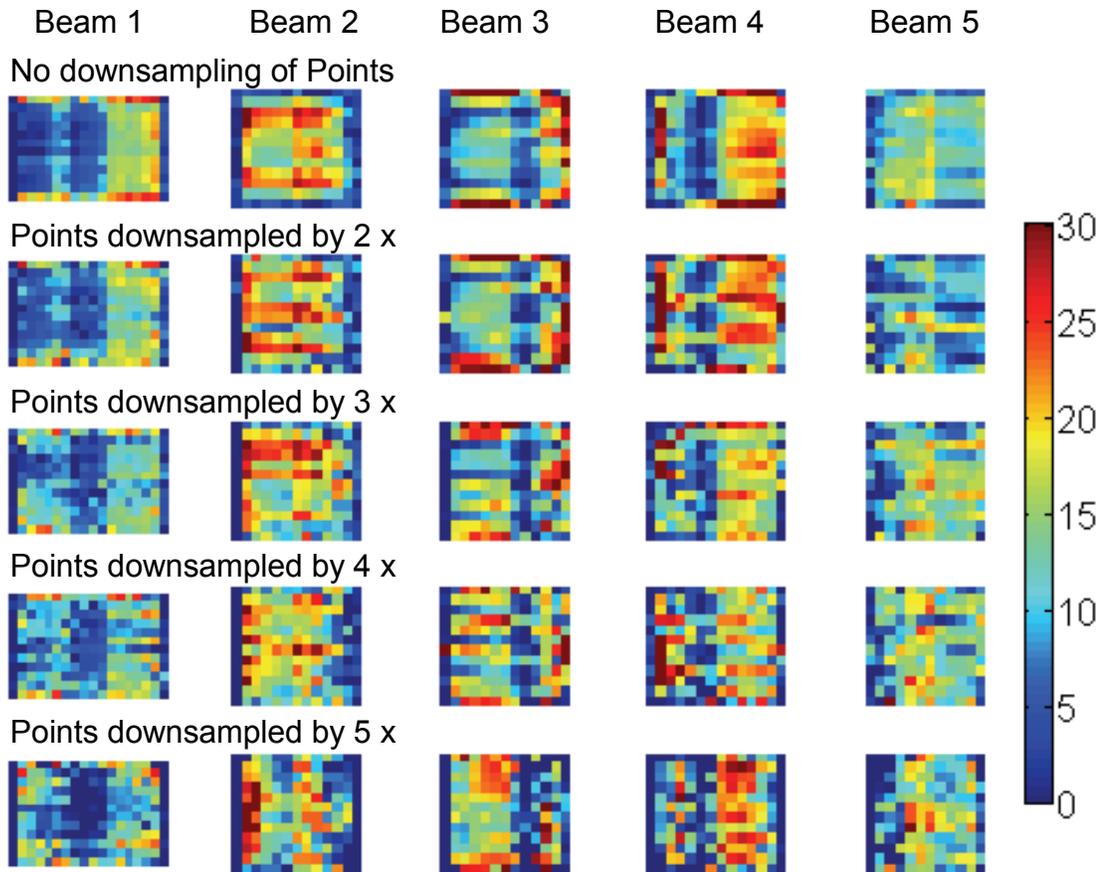


Figure II-4. The optimal IMRT beams as fewer points are used in the optimization. Results are shown with the full sampling and when the number of points is decrease by factors of 2-5.

used in the optimization. There are also fewer isolated high intensity peaks as the point density is increased, although regions of very high intensity are still present. Therefore, it is possible that some of the isolated high intensity peaks seen in optimized IMRT plans are the result of inadequate point sampling.

In Figure II-5, the relative plan intensity map variation, monitor units, cost, and average correlation coefficient of the beamlet patterns relative to the fully sampled plan are shown. Here, the correlation coefficient approaches one (almost linearly, $R^2 = 0.96$) as the sub-sampling is reduced to 1. The PIMV and MU both exhibit a trend of reduced complexity as the sub-sampling goes to 1. However, a decrease in both from subsampling 2 to 3 shows that a strict trend in the complexity reduction is not seen. The optimal cost value, as computed with the points in each case decreases very quickly, showing that a “better” plan can be created with fewer points. This is a result of the plan

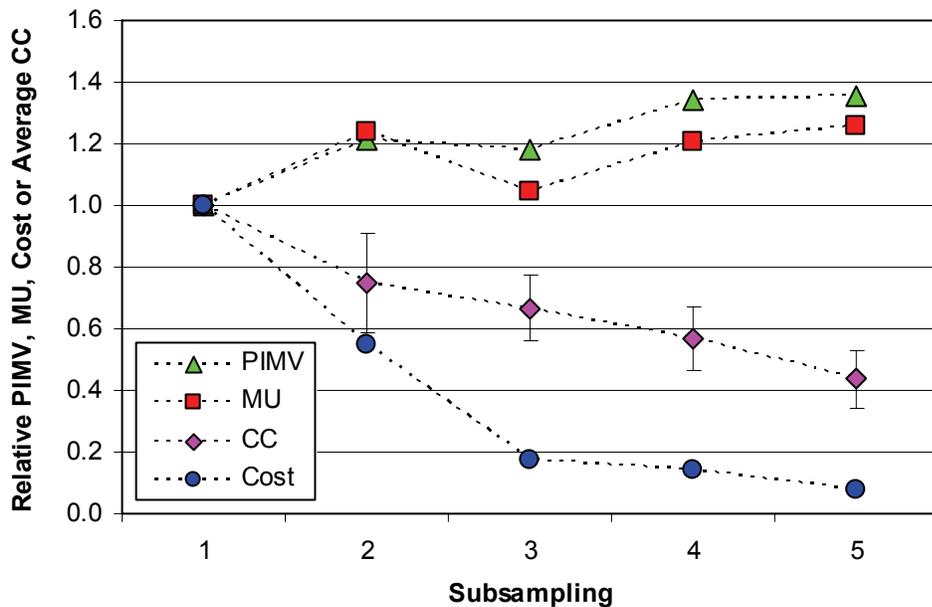


Figure II-5. The plan intensity map variation, monitor units, cost and average correlation coefficient for plans with decreasing point densities relative to the most dense point sampling (1).

geometry being undersampled and the optimizer being able to exploit that and minimize the cost through large beamlet fluctuations. This is also evidenced by the perceived improvements in the dose-volume histograms for each plan in Figure II-6(a). The DVHs show a clear increase in dose to the normal tissues when all points are taken into account. This data shows that care must be taken to properly sample the critical regions of interest so that overdosing does not occur in the normal tissues. To demonstrate this, Figure II-6(b) shows what the “true” DVHs would be for each of the subsampled cases. The optimal beamlet intensities from SS5 through SS1 are applied to the case with no subsampling (SS1), and the DVHs are recomputed using the full point sampling. This

shows the degradation in plan quality that would occur if an insufficient number of points was used in the optimization and treated. The PTV and OAR1 degradation would be critical, causing underdosing and loss of homogeneity in the target and a violation of the maximum dose objective in OAR1. However, the changes seen in the remainder of the normal tissues are much less significant and those DVHs do not differ significantly from those shown in Figure II-6(a).

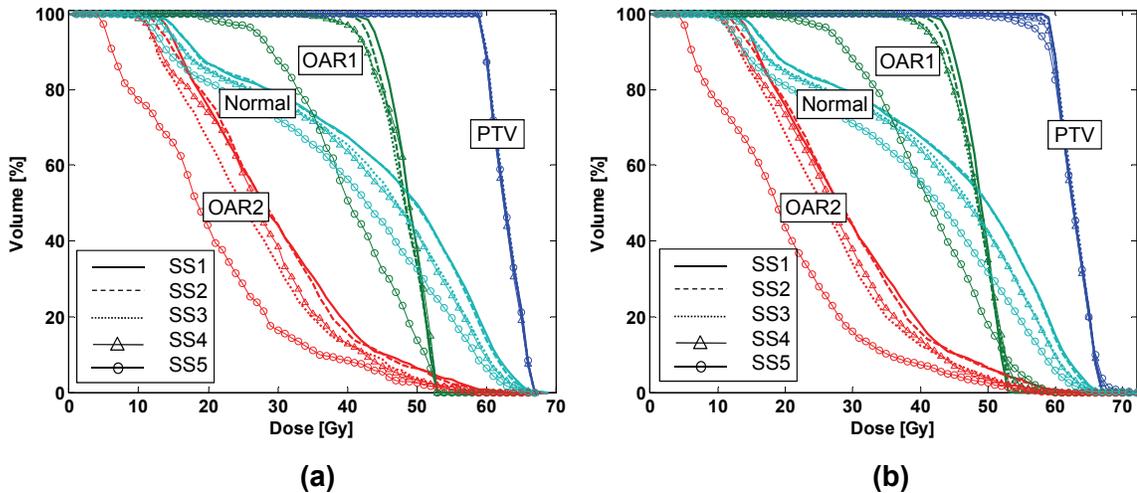


Figure II-6. (a) The optimal DVHs for plans using increasingly fewer points in the optimization, and (b) the true DVHs for each of the subsampled plans when the intensity patterns are evaluated using the full sample of points.

It is clear that undersampling the points give the optimizer a false and simplified representation of the geometry, which makes it easier to achieve the objectives. The result is an increase in modulation as the optimizer exploits the reduced point density by allowing extreme fluctuations from beamlet to beamlet. In systems where the point density cannot be controlled, serious dosimetric discrepancies are possible, and the plans may also be more sensitive to geometric changes due to the high intensity peaks and randomness that appear in the intensity patterns. These features of the optimized beams stem from the fact that the system can exploit the sparse point placement in reduced point densities to decrease the objective function value. Thus, if point sampling can be controlled, importance should be placed on sampling the PTV and other structures within the fields with the greatest point density and so on until the least important structure. It is also very important to realize that the “best” plan from the

optimizer may not be the true optimal plan because of undersampling. Unfortunately, undersampling is common in cases with large fields, such as head/neck patients.

In the remaining work in this dissertation, the density of point sampling was not sacrificed to improve the optimization results or speed. The pointed density was kept as high as possible within the constraints of the system to ensure the integrity of the results.

II.C.1.2. Random Point Sampling

An investigation was performed on the phantom shown in Figure II-3 to determine how much the optimized intensity patterns and dosimetric results are affected by different random point samplings. To do this, five identical cases were created that differed only in point selection. All points were defined using a quasi-random point sampling algorithm and our clinical recommendation of average voxel size (0.027 cm^3) in each region of interest. After beamlet calculations were performed, each of the plans was optimized with an identical quadratic cost function using the quasi-Newton optimization algorithm. No convergence criteria were used to ensure that the optimization process did not terminate prematurely (this is discussed more in Section II.B.3). The optimal dose-volume histograms and beamlet patterns for each instance of point sampling are

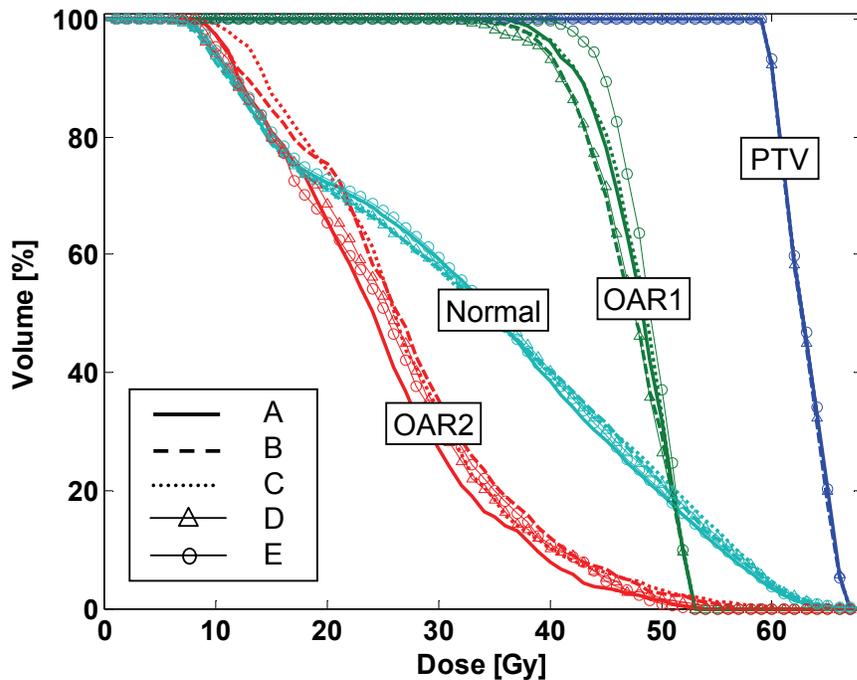


Figure II-7. Dose-volume histograms for the phantom when using 5 different random instances of point sampling in the structures (A-E).

shown in Figures II-7 and II-8, respectively. The number of total monitor units required to deliver the treatment course, the plan intensity map variation, and the number of iterations required for full convergence are given in Figure II-9.

An inspection of the DVHs reveals that the results in each of the optimized cases are very similar. Slight differences in the DVHs for the two OARs are observed for the difference instances of point sampling. These differences are due in part to the small volume of OAR2, and the close proximity of OAR1 to the PTV. Differences in point placement are not as likely to be masked by a large volume in DVHs, and point placement near the high dose region of the PTV could have an effect on the optimal doses. It is also important to note that two plans could have the same cost while having very different DVHs. For example, the OAR1 maximum dose and OAR2 mean dose are the highest priority objectives in those structures. Therefore, slight changes in the “elbow region” of the OAR or in the general shape of the OAR2 DVH will not be important in the cost function. The idea that plans can have a variety of different DVHs and intensity maps while maintaining the same cost level is an example of the degeneracy of the IMRT problem, which is discussed later in this chapter.

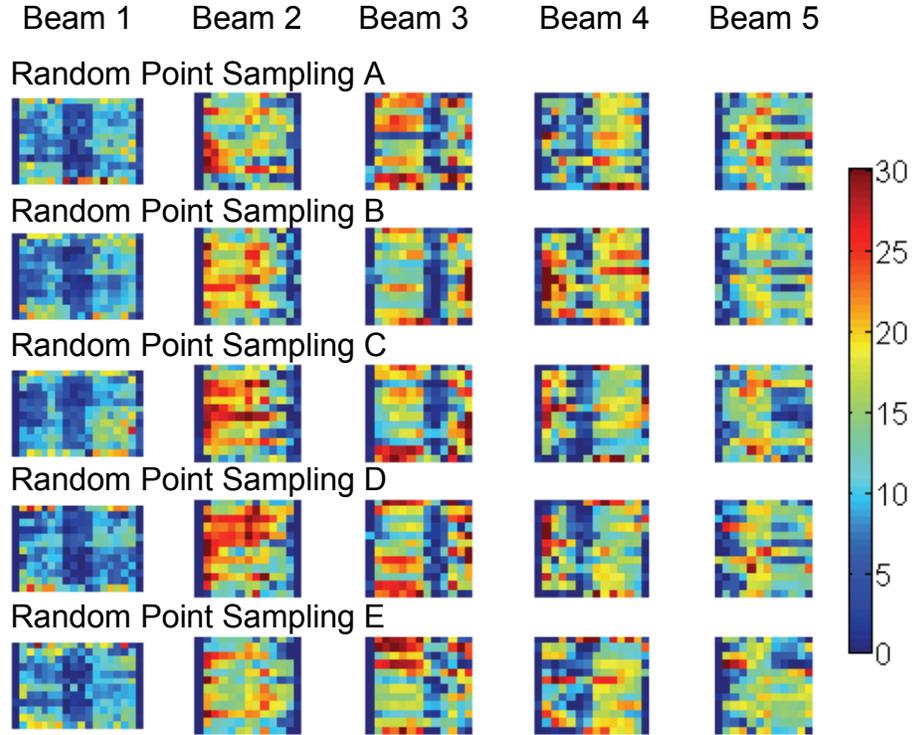


Figure II-8. Optimized beamlet patterns for 5 different point samplings. Beams 1-5 are shown and each different point sampling is labeled A-E.

Figure II-8 clearly demonstrates the effect that different instances of the point sampling can have on the beamlet intensity distribution. While the overall contributions from each beam are similar for all five plans, the individual beamlet intensities are quite varied. While all plans are highly modulated, the MU required for each plan can differ significantly. This is shown in Figure II-9. The MU required for these plans was 35480 +/- 2570 MU. The standard deviation is larger than the average fraction requirement of 1010 MU, demonstrating the large effect that point sampling can have on the optimal beamlet distribution. The PIMV was 16900 +/- 830, which is more consistent. This demonstrates that even though plans can have a similar amount of modulation, the MU required can vary by a large amount. The number of iterations required for convergence did not appear to have any effect on the modulation or MU.

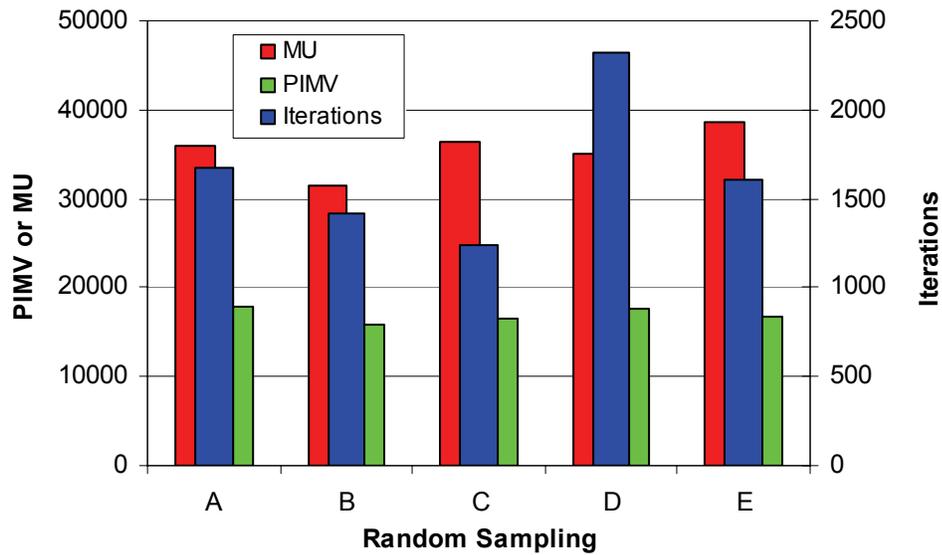


Figure II-9. Total monitor units, plan intensity map variation, and number of iterations to full convergence for plans that are identical except for different random point samplings of the regions of interest.

In summary, the specific point sampling can be responsible for many of the small random fluctuations seen in the optimized beams, and can cause significantly different optimal beamlet patterns, as evidenced Figure II-8. The average correlation coefficient between each of the different plans was 0.62 +/- 0.40, which is a relatively poor correlation based on previous experience. The fluctuations shown are evidence that smoothing beamlet patterns in an intelligent way may reduce the complexity and possibly make IMRT solutions more robust to the placement of the points in the regions

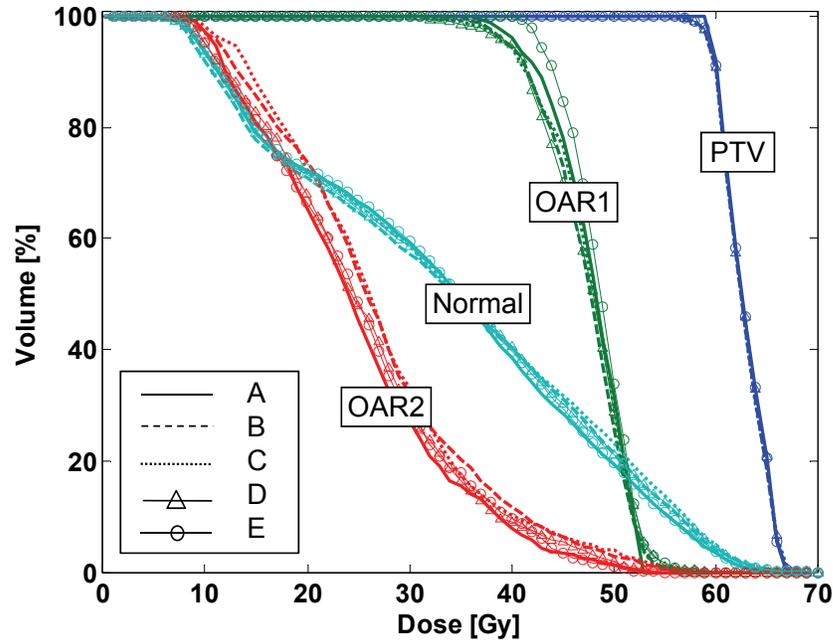


Figure II-10. Dose-volume histograms for the phantom when the optimized beams from all point samplings are evaluated with the point sampling of A.

of interest. In support of this, Figure II-10 shows the DVHs for all of the separately optimized IMBs evaluated using the random points defined in A. The DVHs are almost identical, showing that the varieties of beamlet patterns shown in Figure II-8 produce almost identical results for the specific point sampling in A. This demonstrates the extent to which the optimization system can exploit the specific point placement to make very small changes in the objective function value with no real clinical significance.

II.C.2. Degeneracy

Several authors have contended that IMRT plans can be highly degenerate, leading to inefficiency in planning and a reduction in confidence that one has the “optimal” solution^{13,37}. Degeneracy in IMRT can present itself in several forms. The first, which was discussed in the previous section, is when two different discretizations of the point samplings have different optimal beamlet intensities, but can achieve the same dosimetric result or cost function value. The second, which is the more common definition of degeneracy occurs when there are several different solutions, or different intensity patterns for the same problem that can achieve the equivalent dosimetric results or cost function values. An excellent discussion of this topic appears in Alber *et al.*¹³. In summary, they conclude that most radiotherapy optimization problems are

determined by the interaction of a small number of eigenvectors, meaning that in most areas of the beam, individual beamlet values do not have a large impact on the ability to minimize the objective function. Thus, they conclude that IMRT problems with a large number of high resolution beamlets can be highly degenerate, resulting in very noisy IMRT profiles if nothing is done to prevent the degeneracy. In addition, they note that the degeneracy problem may be more dependent on the objective function definition than the beamlets or optimization method.

Another common source of IMRT degeneracy is the use of parallel-opposed beams. This simply means that the beam central axes overlap, but the beams come from opposite directions. In this case, the beams can contain redundant capabilities and changes in the beamlet intensities in one beam correlate highly with necessary changes in the opposite beam. Therefore, different combinations of the beamlet contributions from each of the beams could produce the same dosimetric results.

An example that illustrates both of the above forms of degeneracy is the tangential IMRT breast treatment shown in Figure II-11. This beam arrangement is commonly used in breast therapy due to its ability to spare the heart and lung. The beams are composed of 1 cm x 1 cm beamlets and optimized from a starting distribution of random beamlet intensities.

To show the degeneracy of the solution, the case was optimized from three

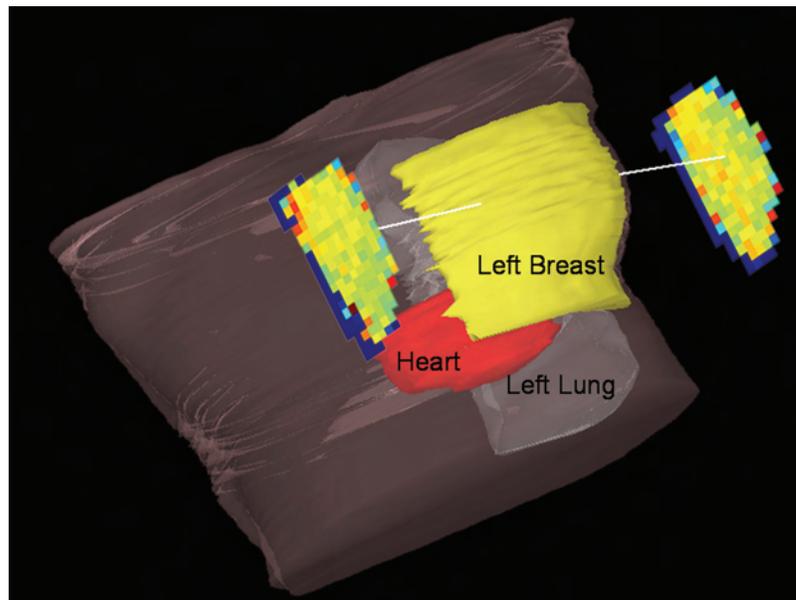


Figure II-11. A tangential left breast IMRT treatment. The left breast, heart, and left lung are shown in relation to the two beams.

different starting points. Each was simply a different random distribution of beamlet intensities. Figure II-12(a) shows the overlapping DVHs from each of the three optimized plans, demonstrating the dosimetric equivalence. Figure II-12(b) shows the different intensity modulated beams from each case. They have been lined up geometrically such that the circled regions roughly correspond to beamlets that would be affecting similar geometrical areas. In some cases, the circled regions show differing contributions from each beam, but may add up to similar values. On the other hand, some circled regions do not appear to follow any kind of trend, demonstrating that they

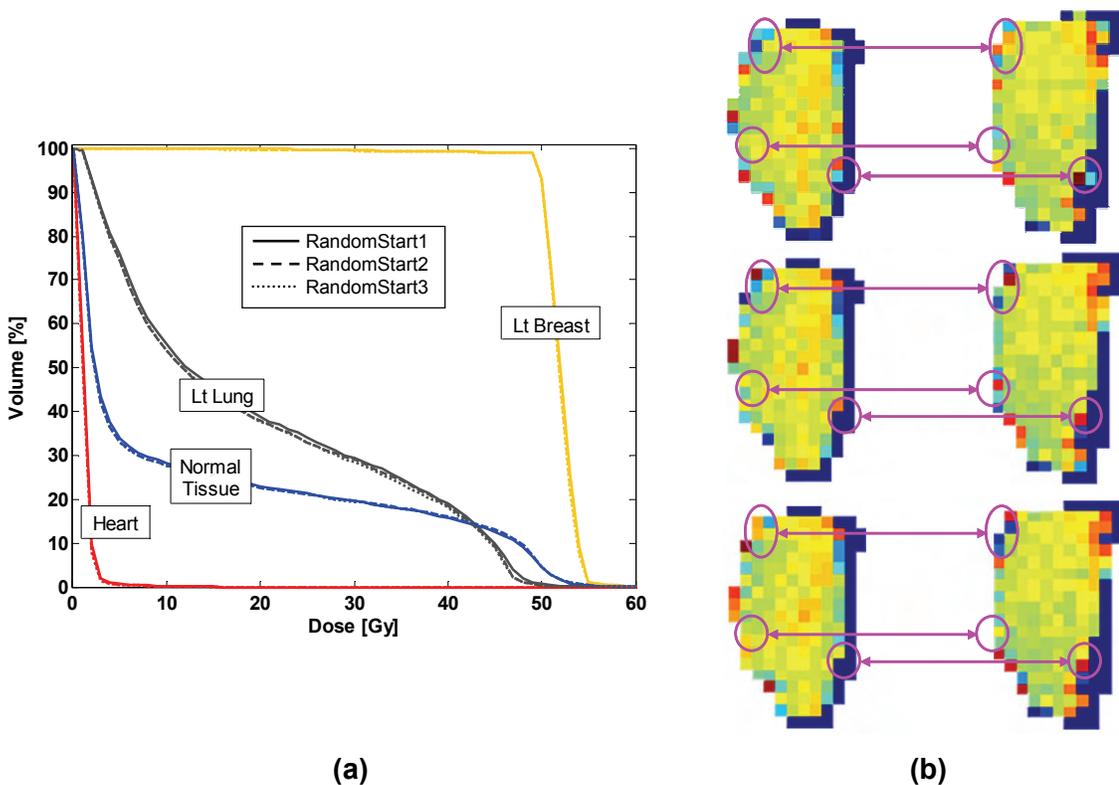


Figure II-12. (a) Dose-volume histograms for three different optimized breast plans starting at different random beamlet intensities, and (b) the corresponding beamlet distributions. The violet circles and arrows represent points in the field that roughly cover the same geometry in both beams.

are likely in areas that do not affect the objective function and can become noisy with no consequence (as Alber *et al.* suggest). A quick comparison of the 3 sets of beams clearly demonstrates that many different beamlet patterns are dosimetrically equivalent, and that certain beamlets can vary by large amounts without affecting the dosimetric results.

II.C.3. Optimization Method and Objective Function

It has been suggested that the three possible sources of unwanted “stochastic noise” in intensity modulated beams (IMBs) are the optimization technique, the cost function, and the definition of convergence of the cost function³⁸. In this section we review each of these issues and discuss them in the context of the optimization system, UMOpt, to determine the extent to which each may impact the complexity of IMRT beams.

II.C.3.1. Optimization Algorithm

The choice of the optimization algorithm used to solve inverse radiotherapy problems depends significantly on the cost function characteristics. The cost function is most commonly a collection of mathematically stated goals which are given importance weighting factors. Each goal will then correspond to a penalty for not reaching the goal. The purpose of the optimization algorithm is to find the variables or beamlet intensities that minimize the total penalty or total cost of the cost function. Depending on the shape of the cost function, this optimization problem can have multiple local minima, rather than one global minimum. If substantially different local minima are present, a probability-based stochastic algorithm, such as simulated annealing, may be used to find the global solution. Unfortunately, these algorithms can be extremely time-intensive, requiring an almost unmanageable number of iterations to reach the global solution. If multiple local minima do not exist, or the local minima do not vary significantly from each other, a deterministic algorithm, such as a gradient-descent technique, is a much more efficient choice. Deterministic algorithms are based on mathematics and rely on the problem and objective function to steer the search for a local solution. These algorithms can quickly converge to a solution, which is a positive aspect for IMRT planning since it already consumes much more time and resources than conventional or 3D conformal radiotherapy planning. In addition, deterministic algorithms have not been observed to introduce the additional “noise” into the intensity modulated beams that is seen when using a stochastic search method³⁸.

The possible number of objectives that can be formulated into a cost function is fairly large, ranging from simple dose-based objectives to biological models of tumor control probabilities (TCP) and normal tissue complication probabilities (NTCP). The use of dose-volume objectives given as “no more than x % of a structure can receive greater than y Gy” is also very common in the community. While Deasy has proven the existence of multiple local minima when using dose-volume constraints³⁹, several

authors have studied these local minima and concluded that they are not only rare, but also dosimetrically insignificant^{40,41}. In addition, convex approximations to these dose-volume objectives have been formulated and used successfully in gradient descent optimization algorithms. These observations are very important because they suggest that the annealing properties of stochastic algorithms are not required to find the global minimum of radiotherapy problems—or at least a solution that is clinically equivalent to the global minimum. This means that (i) more efficient gradient descent algorithms may be adequate for most IMRT planning, assuming that each new objective is proven to have either no or insignificant local minima, and (ii) the noise introduced into intensity modulated beamlet patterns by simulated annealing algorithms can likely be avoided.

In support of the degeneracy theory discussed by Alber *et al.*, Llacer *et al.* have observed that even though the multiple local minima observed in radiotherapy plans show negligible differences in the dosimetric results of the plan, the optimal beamlet intensity patterns vary widely. This evidence supports the assertion by Alber *et al.* that IMRT plans are degenerate, and some intervention should be taken to steer the optimization algorithm into a delivery-efficient plan. This is also evidenced by our breast optimization results in Figure II-12.

The role that the optimization algorithm plays in the complexity of the intensity modulated beams is studied briefly using an example case. A simulated annealing algorithm versus a deterministic quasi-Newton method is used to optimize the IMRT benchmark-like geometry shown in Figure II-3. In addition to the optimization method, the effect of the optimization convergence criteria and starting beamlet intensities on the optimal beam modulation are also studied.

In simulated annealing, several components dictate how the optimization process proceeds and when it stops. These parameters include a step size and its bounds, the number of cycles and iterations performed, and the annealing criteria, which include a starting “temperature” and several other parameters. The step-size is governed by a starting step-size, minimum step-size, maximum step-size, and step-size fraction. The latter determines the step-size in the subsequent cycle. Usually the step-size fraction is less than one so that the step-size slowly gets smaller in order to promote convergence. Each cycle is comprised of a certain number of iterations. Each iteration randomly changes the weight of one beamlet and accepts or rejects the change depending on whether it improves the cost function or according to some probability set by the temperature. At the University Michigan, a generally-accepted value for the number of

iterations per cycle is three times the number of beamlets. The step-size and fraction can affect the time to convergence, but the major parameter that affects the point to which the plan is optimized is the number of cycles. In this method, the number of cycles required to find the global minimum can be quite large, and it is difficult to determine when one is near the global minimum because cost may change very slowly in the final cycles due to the small-step size.

In the quasi-Newton method (one of the search algorithms employed inside UMOpt) there are two different convergence criteria. The first, “*factr*” is a user-set tolerance in the termination test for the algorithm. The iterations will stop when

$$\frac{f^k - f^{k+1}}{\max(f^k, f^{k+1}, 1)} \leq \text{factr} \times \text{eps}mch, \quad (\text{II.3})$$

where f^k is the cost after the k^{th} iteration and *eps*mch is the machine precision, which is automatically generated by the code. Typical values for *factr* on a computer with 15 digits of accuracy in double precision range from 1e12 for low accuracy to 1 for extremely high accuracy. The user can suppress this termination test by setting *factr* = 0. The second user-specified termination parameter, “*toler*”, is a double precision variable and will stop iterations when the following is met:

$$\max(|pg_i|) \leq \text{toler}, \text{ for } i = 1..n \quad (\text{II.4})$$

Here, pg_i is the i^{th} component of the projected cost function gradient with respect to the beamlet i . Similarly, the user can suppress this termination test by setting *toler* = 0. If both of these termination criteria are bypassed, then the optimization terminates when the line search cannot locate an adequate point after 20 function iterations. This happens at a point where the roundoff error begins to dominate the computation.

The optimization trials that were run to show the effect of the optimization algorithm and its parameters on the complexity of the beams are listed in Table II-1. The first comparison is one between plans optimized with quasi-Newton (QN) versus simulated annealing (SA). Since the QN method has been shown to be adequate for radiotherapy problems by more exhaustive studies, we just show this comparison as an example. The next plans analyze the effect of the QN convergence criteria (*factr* and

toler) on plan modulation starting from both uniform and random beamlet intensities.

Table II-1. Optimization Trials for IMRT Benchmark Phantom

#	Algorithm	Start	factr	toler
1	Simulated Annealing	Uniform 0	NA	NA
2	quasi-Newton	Uniform 0	0.0E+00	1.0E-01
3	quasi-Newton	Uniform 0	0.0E+00	1.0E-02
4	quasi-Newton	Uniform 0	0.0E+00	1.0E-03
5	quasi-Newton	Uniform 0	0.0E+00	1.0E-04
6	quasi-Newton	Uniform 0	0.0E+00	1.0E-05
7	quasi-Newton	Uniform 0	0.0E+00	1.0E-05
8	quasi-Newton	Uniform 0	1.0E+15	0.0E+00
9	quasi-Newton	Uniform 0	1.0E+14	0.0E+00
10	quasi-Newton	Uniform 0	1.0E+13	0.0E+00
11	quasi-Newton	Uniform 0	1.0E+12	0.0E+00
12	quasi-Newton	Uniform 0	1.0E+10	0.0E+00
13	quasi-Newton	Uniform 0	1.0E+08	0.0E+00
14	quasi-Newton	Uniform 0	0.0E+00	0.0E+00
15	quasi-Newton	Random	0.0E+00	1.0E-01
16	quasi-Newton	Random	0.0E+00	1.0E-02
17	quasi-Newton	Random	0.0E+00	1.0E-03
18	quasi-Newton	Random	0.0E+00	1.0E-04
19	quasi-Newton	Random	0.0E+00	1.0E-05
20	quasi-Newton	Random	0.0E+00	0.0E+00
21	quasi-Newton	Random2	0.0E+00	0.0E+00
22	quasi-Newton	Random3	0.0E+00	0.0E+00
23	quasi-Newton	Random4	0.0E+00	0.0E+00
24	quasi-Newton	Random5	0.0E+00	0.0E+00

Optimization with the simulated annealing algorithm took several days, and in the interest of time, the optimization was terminated when the cost changed by less than 0.0001 over 100 cycles with reasonable step sizes. A comparison of the SA and QN results is shown in Figures II-13 and II-14. In Figure II-13, there is a slight increase in dose to the normal tissue with the SA algorithm, which could be due to premature stopping of the algorithm. In the intensity modulated beams, it appears that the SA beams have more random noise and modulation, which is not a surprise due to the random search of SA and the previous results of Coolens *et al.* Figure II-14 compares the values of the PIMV, MU, and cost for each of the methods. These comparisons support the observation that the SA beams are noisier, with the PIMV and MU (especially) being higher in the SA plan. The cost values are very similar, demonstrating

that there was not a problem with the QN method falling into a local minima. The fact that the SA and QN plans are fairly similar, combined with the length of time required to optimize with SA, support the use of QN for routine IMRT planning to eliminate any noise that could be introduced due to the random SA search method.

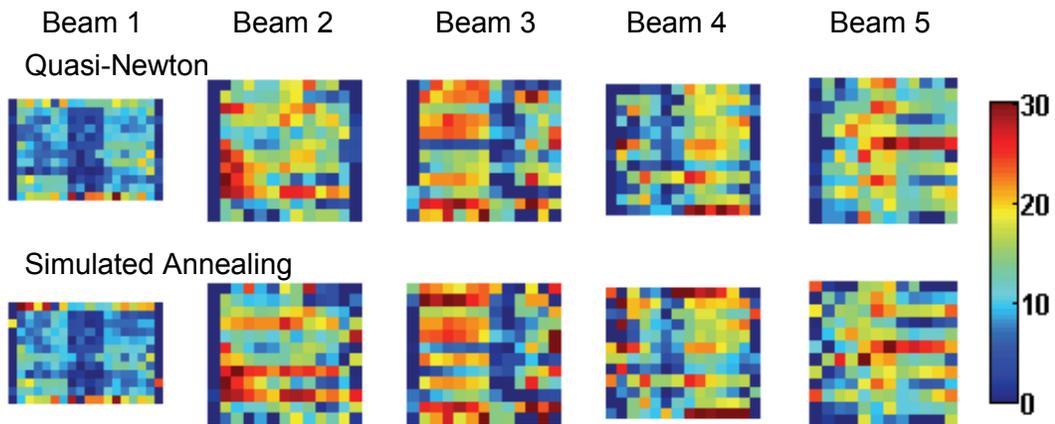
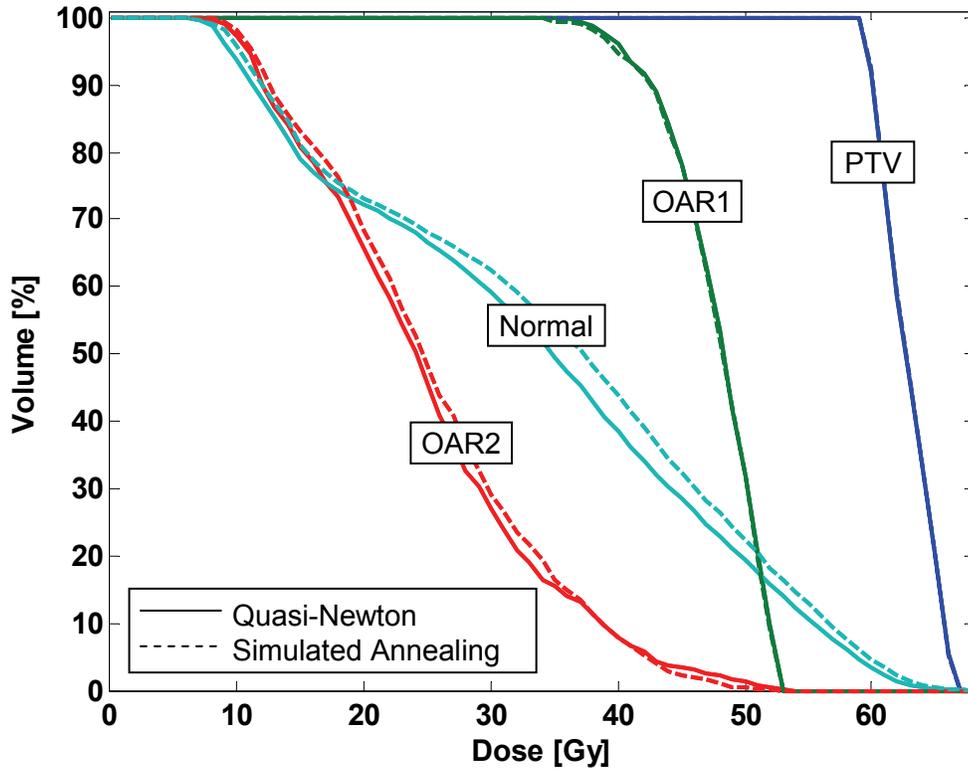


Figure II-13. Dose-volume histograms and intensity modulated beams for plans optimized with quasi-Newton and simulated annealing optimization algorithms.

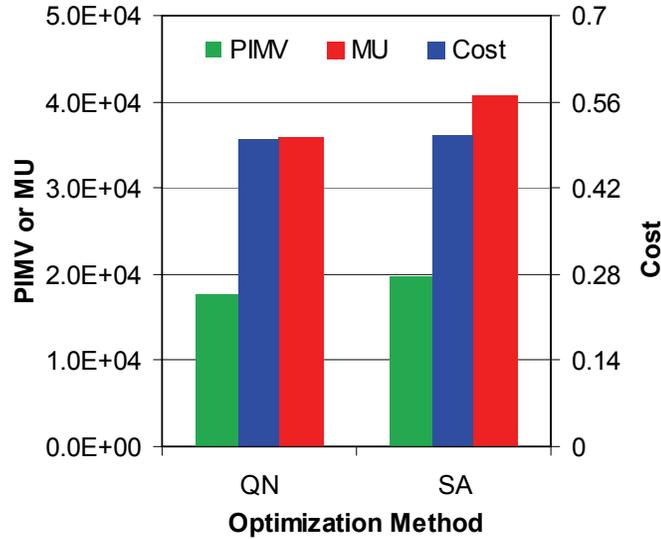


Figure II-14. The PIMV, MU and cost for plans optimized with the given the quasi-Newton vs. simulated annealing optimization methods.

The comparison between the two convergence parameters in the QN method revealed that they both perform similarly. This is expected since the optimization method itself should proceed in the exact same way with only a different stopping point. We did not observe any inconsistent behavior with either stopping criteria. Figure II-15

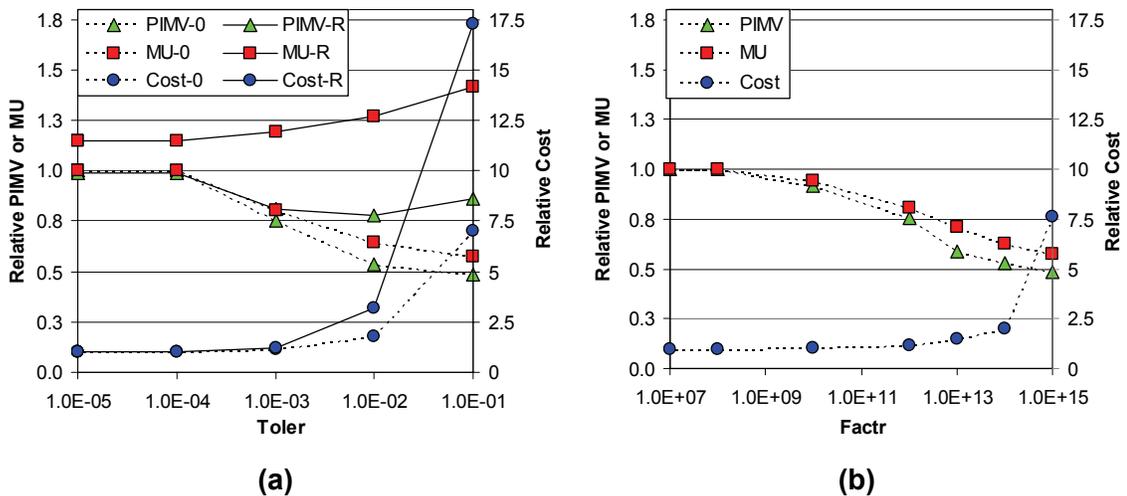


Figure II-15. Relative PIMV, MU and Cost for plans optimized with the given (a) Toler stopping criteria and starting at uniform beamlet intensities of 0 (dotted lines) or random beamlet intensities (solid lines), and (b) Factr stopping criteria starting at starting uniform beamlet intensities of zero. All values are relative to the fully converged plans with starting intensities of zero.

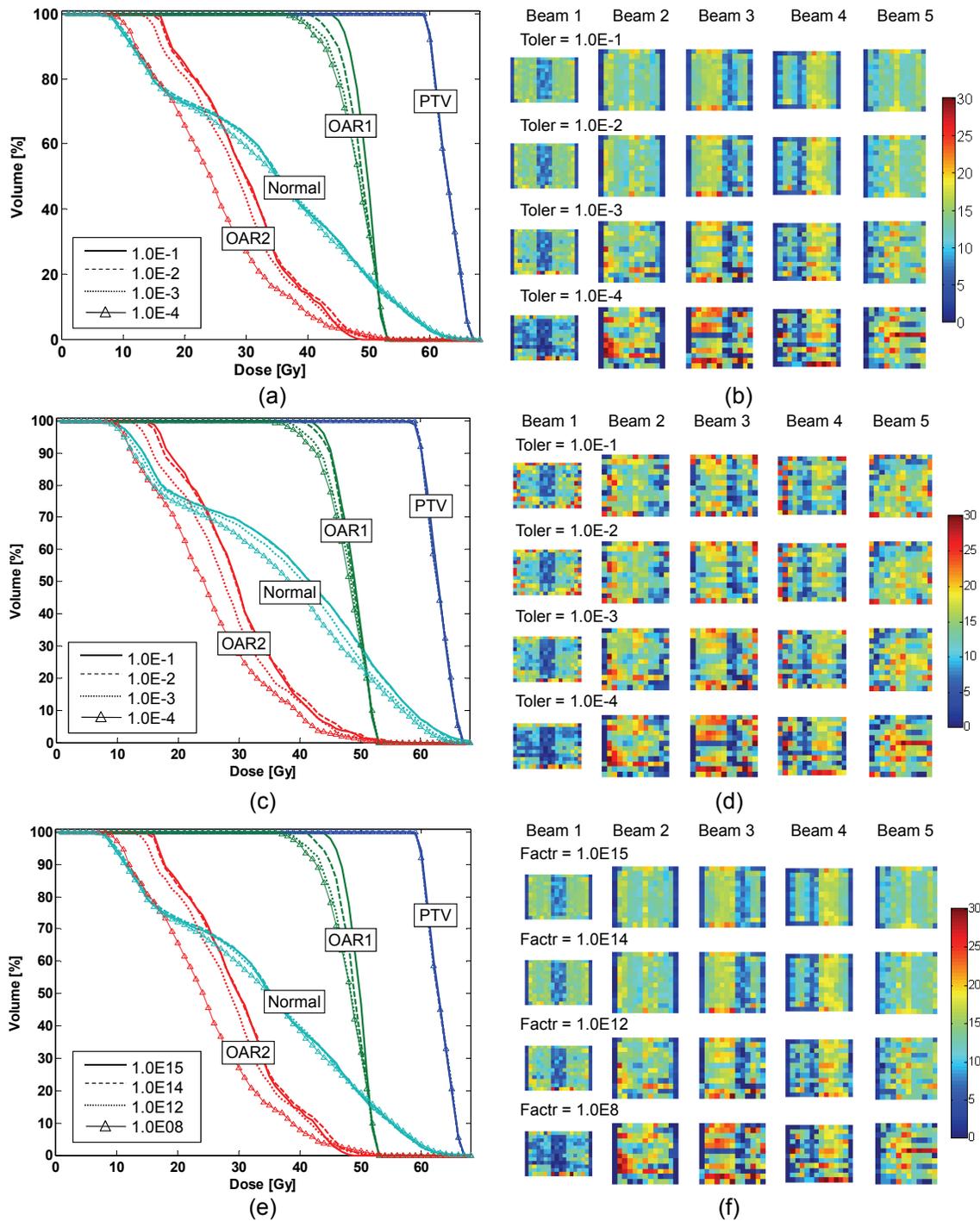


Figure II-16. DVHs and beams for the series of plans optimized with increasing (a and b) Toler stopping criteria and starting at uniform beamlet intensities of 0 (dotted lines) or (c and d) random beamlet intensities, and (e and f) Factr stopping criteria starting at starting uniform beamlet intensities of zero.

summarizes the different optimization trials run to test the stopping criteria (# 2-20 in Table II-1). Figure 15(a) shows the behavior of the PIMV, MU, and Cost at decreasing *toler* values when starting from uniform zero or random beamlet intensities, and Figure II-15(b) shows the same trends with decreasing *factr* value when starting from uniform zero beamlet intensities. These figures demonstrate that (i) when starting from uniform intensities, the MU and PIMV both increase as the cost decreases toward the minimum value, and (ii) with a random starting point, the MU and PIMV decrease as the cost decreases to the final solution.

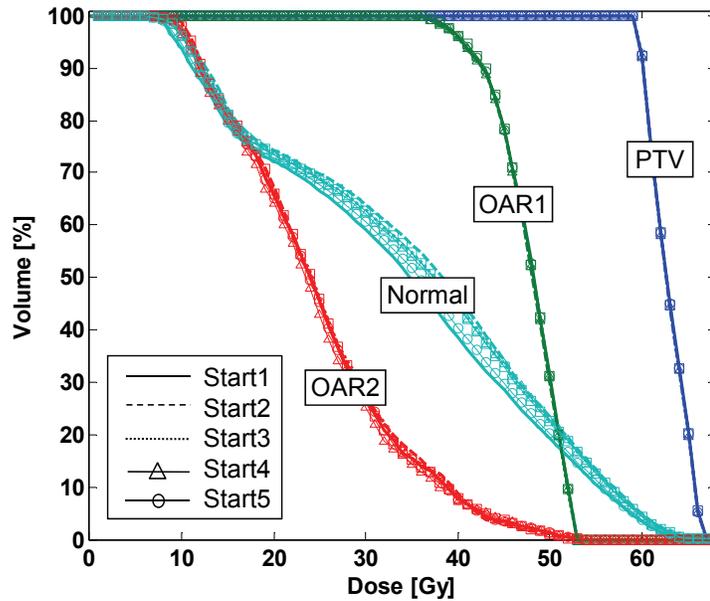


Figure II-17. Dose-volume histograms for the phantom when the optimized beams from 5 different random starting points.

Figure II-16 shows the DVHs and beams for the series of the plans in Figure II-15. Figure II-16(a) and (b) show that the DVHs improve as the uniform plan becomes more modulated, while (c) and (d) shows the opposite effect due to the random starting beamlet intensities. Figure II-16(e) and (f) simply demonstrates the parallel behavior of *toler* and *factr* when the other is set to zero.

An analysis of the examples shown in Figures II-15 and II-16 shows that modulation can be reduced if the optimization is stopped before full convergence when starting from uniform beamlet intensities. However, this premature stopping of the optimization algorithm can result in sub-optimal DVHs, and the degradation in the plan could be significant in terms of dose to the normal tissues.

Optimization trials 20-24 in Table II-1 compared the fully converged plans when starting for 5 different random starting intensities. This comparison was meant to study the effect of the starting point on the optimal beamlet intensities, as well as to see

Table II-2. Mean and σ of converged cost, MU, PIMV, and CC for 5 random starting beamlet intensities

	Cost	MU	PIMV	CC
Mean	0.51	38791	18049	0.89
σ	0.01	2205	547	0.03

whether there were multiple local minima or degeneracy observed in the beams. Table II-2 shows the mean and standard deviation of the converged cost, MU, PIMV, and average correlation coefficient for the five optimization instances. The cost value demonstrates that there are no significant local minima in terms of dosimetric quality, but the MU standard deviation shows that there is likely a large amount of degeneracy in the beamlet intensity patterns. The PIMV varies by less than the MU, and correlation coefficient between the plans averages a fairly high correlation of 0.89.

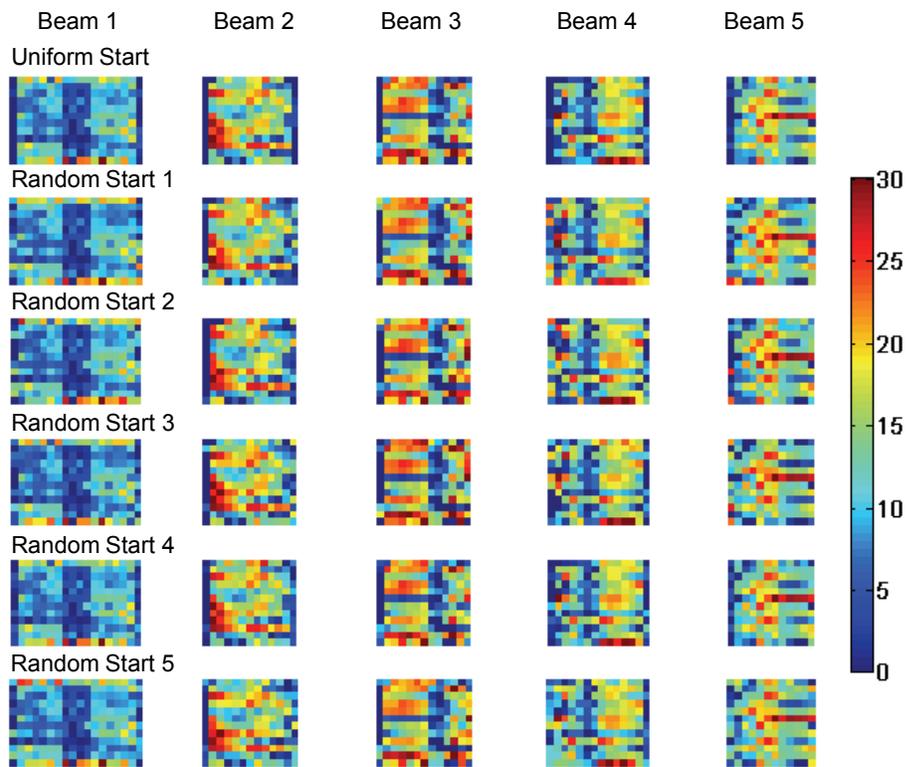


Figure II-18. IMRT beams for cases optimized to full convergence from different starting points. The uniform start is shown as a control for the 5 random starts.

Figure II-17 supports the conclusion that no significant dosimetric local minima were observed in these five plans, and agrees with the much more exhaustive testing of local minima by other authors that was discussed earlier. Here, the DVHs are very similar for all cases, while the beams in Figure II-18 actually vary by a large amount, especially near the edges of the field. Therefore, in Figure II-19, we have plotted a beamlet difference map between each of optimal beams started from random intensities and the optimal beamlet patterns when started from uniform beamlet intensities. This clearly demonstrates that the beamlets on the edge of the fields that do not overlap with the target are not a significant factor in determining the optimal cost. To some extent, it

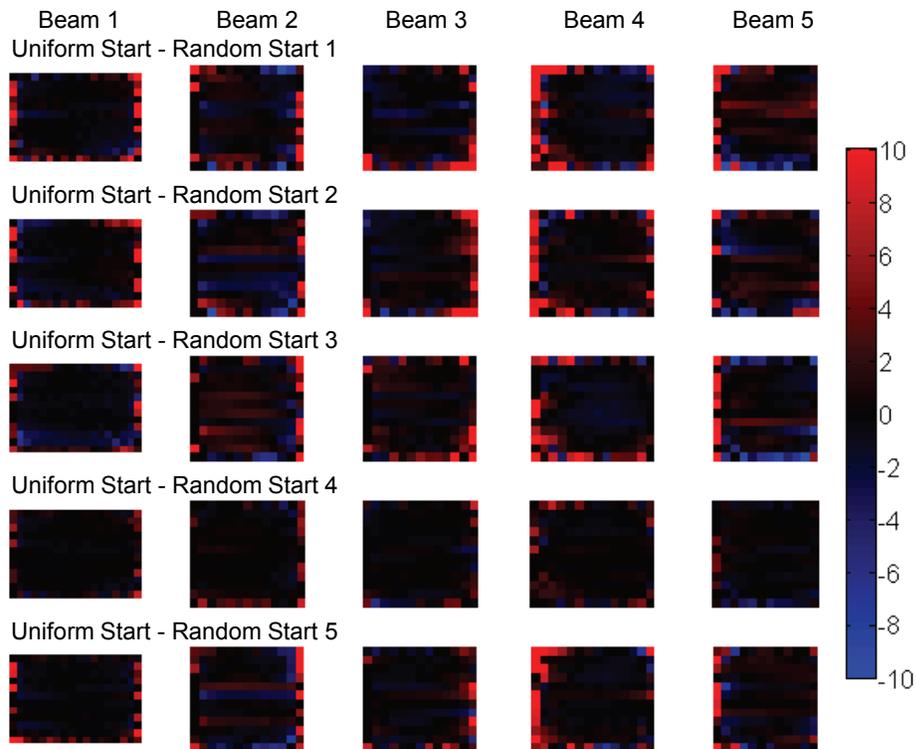


Figure II-19. Beamlet different displays between the uniform start and each of the random starts. The colors are cut off between -10 and 10 although the actual differences could be higher.

also appears that the beamlets in the center of Beam 2 may not have a large effect of the cost function. This supports the idea that IMRT problems are degenerate in the sense that there can be a variety of the beamlet patterns to reach the same cost because there are regions of the field that have a very minimal effect of the optimization process. In this case, it appears that the randomness of the starting intensities is an

artifact in the optimized beams because the optimizer had no incentive to remove it. This is further evidence that beamlet patterns can be smoothed without affecting the dosimetric quality.

II.C.3.2. Objective Function

The previous section highlighted the many objectives and goals that can be included in an IMRT cost function. Inside UMOpt, the treatment planner can choose a variety of “evaluators” to help build the cost function. Evaluators include minimum, maximum, or mean dose to a target or organ-at-risk. UMOpt also supports the use of dose-volume and biological objectives such as generalized equivalent uniform dose (gEUD), normal tissue complication probability (NTCP), and tumor control probability (TCP). Each evaluator can then be paired with a “modifier” in order to create a “costlet” that gives the penalty value for not meeting the intended objective. Then, all of the costlets are combined to form the cost function or objective function. A detailed study of the cost function components available in UMOpt is given by Kessler *et al.*⁷

Due to the variety of possible costlets that can be created, the cost function in UMOpt can take on many different forms and levels of complexity. The purpose of this section is not to analyze all possible cost functions, but to give several examples of the

Table II-3. Stepwise Brain Cost Function

Step	Objective(s) Added at Each Step	Weight
1	PTV1 Min Dose > 66.5 Gy	100
	PTV1 Max Dose < 73.5 Gy	100
	PTV2 Min Dose > 57 Gy	100
	PTV2 % Vol > 63 Gy = 10 %	10
	PTV2 Max Dose < 73.5 Gy	100
2	Optic Chiasm Max Dose < 55 Gy	100
	Optic Nerves Max Dose < 55 Gy	100
	Brainstem Max Dose < 60 Gy	100
3	Normal Tissue Max Dose < 63 Gy	10
4	Optic Chiasm Mean Dose < 0 Gy	0.001
	Optic Nerves Mean Dose < 0 Gy	0.001
	Brainstem Mean Dose < 0 Gy	0.001
5	Normal Tissue Mean Dose < 0 Gy	0.001
6	Replace Mean Dose < 0 Gy with Max Dose < 0 Gy	0.001

effect that the cost function can have on beam complexity. To demonstrate this, a brain IMRT case will be optimized in a stepwise fashion as outlined in Table II-3, recording the changes in complexity after set of costlets is added to the total cost function. The final costlets that will be added have the purpose of reducing any unnecessary dose to the normal tissues. A goal of mean or maximum dose < 0 Gy is unattainable in any structure near the target and is therefore given a low importance factor. While the goal is to reduce the unnecessary dose to the normal tissues, both costlets perform differently and could affect the intensity patterns in different ways. Next, the traditional dose-based cost function in Table II-3 will be replaced by a generalized equivalent uniform dose cost function. Equivalent uniform dose (EUD) is a metric that takes a heterogeneous dose distribution from a target or normal tissue structure and represents it by the uniform dose value that would have the same radiobiological effect⁴². The generalized EUD is given by:

$$EUD = \left(\frac{1}{N} \sum_i D_i^a \right)^{\frac{1}{a}} \quad (II.6)$$

where D_i is the dose in the i^{th} voxel of the structure, N is the total number of voxels, and a is a structure-specific parameter. The parameter a is positive for normal tissues and negative for tumors. Its value depends on the properties of the tissue, with $a = -5$ representing a responsive tumor and $a = -20$ representing a resistant tumor. For parallel-behaving normal tissues, a is near 1, and for serial-behaving normal structures, such as spinal cord, a is approximately 8. To compare the use of our conventional cost function to one employing only EUD, we replaced all of the costlets for each structure by a single EUD costlet for each. The values chosen are those values that are equivalent to the doses produced by the conventional cost function.

The results of the stepwise optimization in Table II-3 are shown in Figure II-20. On the left are the DVHs after each step of the optimization. As each costlet is included in the cost function, the plan becomes more complex. The addition of the fourth step, which asks for a minimization of the mean dose to the critical normal structures in the field creates competing interests between the target dose and normal tissue dose minimization. This conflict increases the complexity of the solution, which can be observed by the intensity modulated beams on the right.

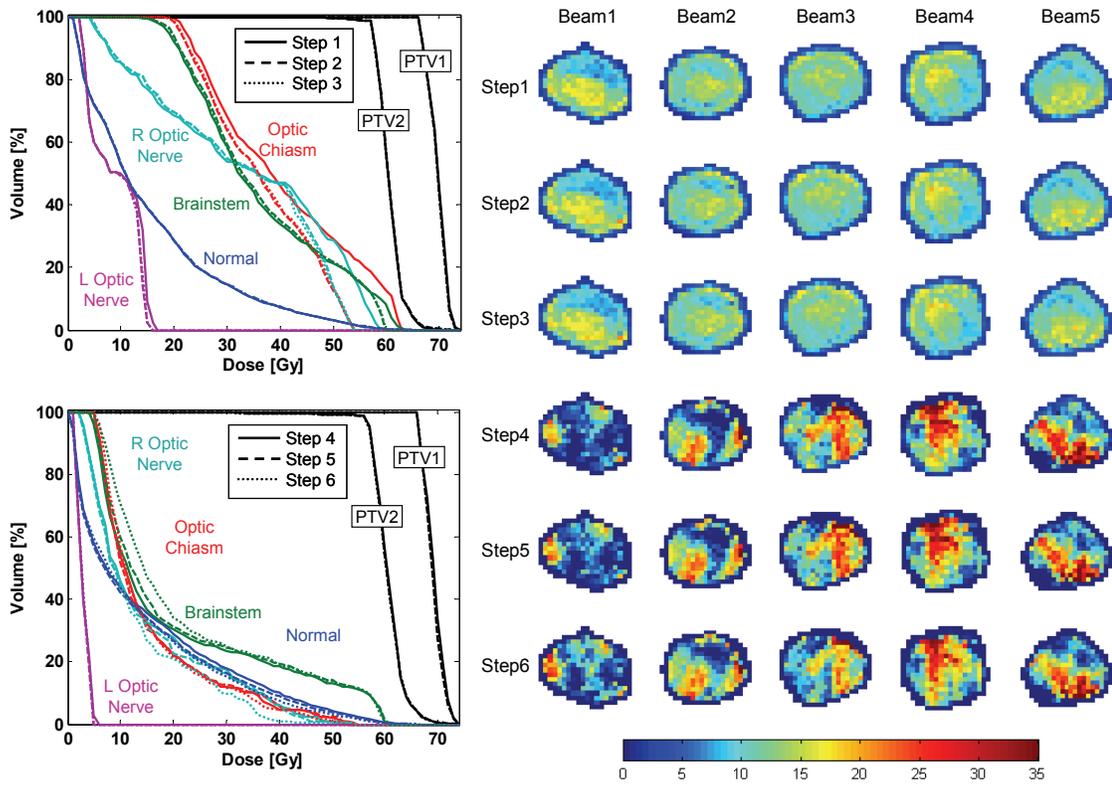


Figure II-20. The results from each step of the brain optimization outlined in Table 3. On the top left are DVHs for the first three steps of the optimization process and on the bottom left are DVHs from the final three steps. Steps 5 and 6 compare the use of Mean vs. Threshold costlets to minimize normal tissue dose.

Figure II-21 displays the effect that the different cost function components have on the plan intensity map variation (PIMV) and plan monitor units. The beam complexity rises significantly, as expected, after step 4. The shift from mean dose minimization to maximum dose minimization costlets (step 5 to 6) only slightly affects the plan DVHs and intensity modulated beams and PIMV. However, the MU decrease slightly with the use of the maximum dose costlets. This may be due to the fact that a decrease in the maximum dose decreases the maximum intensity of some of the beamlets, resulting in the need for slightly fewer MU.

Figure II-22 displays the difference in plan DVHs and intensity modulated beams when replacing the conventional cost function with an EUD-based cost function. The result is a significantly smoother plan with less homogeneity in the targets and more low dose and less high dose going to the critical normal structures. The MU required to deliver a 2 Gy fraction decrease from 991 to 299 MU. In an effort to further improve the

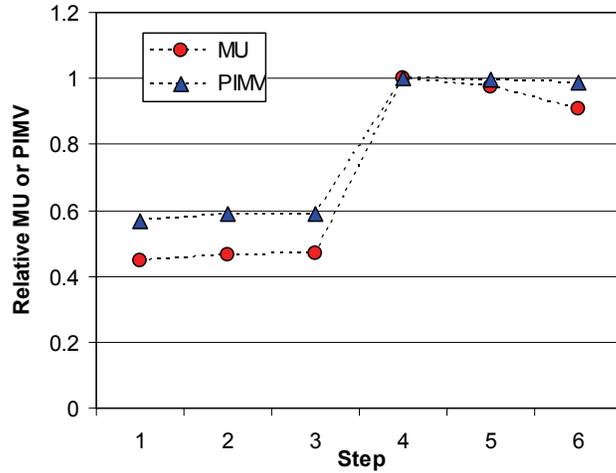


Figure II-21. The relative monitor units and plan intensity map variation after each step of the brain optimization outlined in Table II-3. Steps 5 and 6 compare the use of Mean vs. Threshold costlets to minimize normal tissue dose.

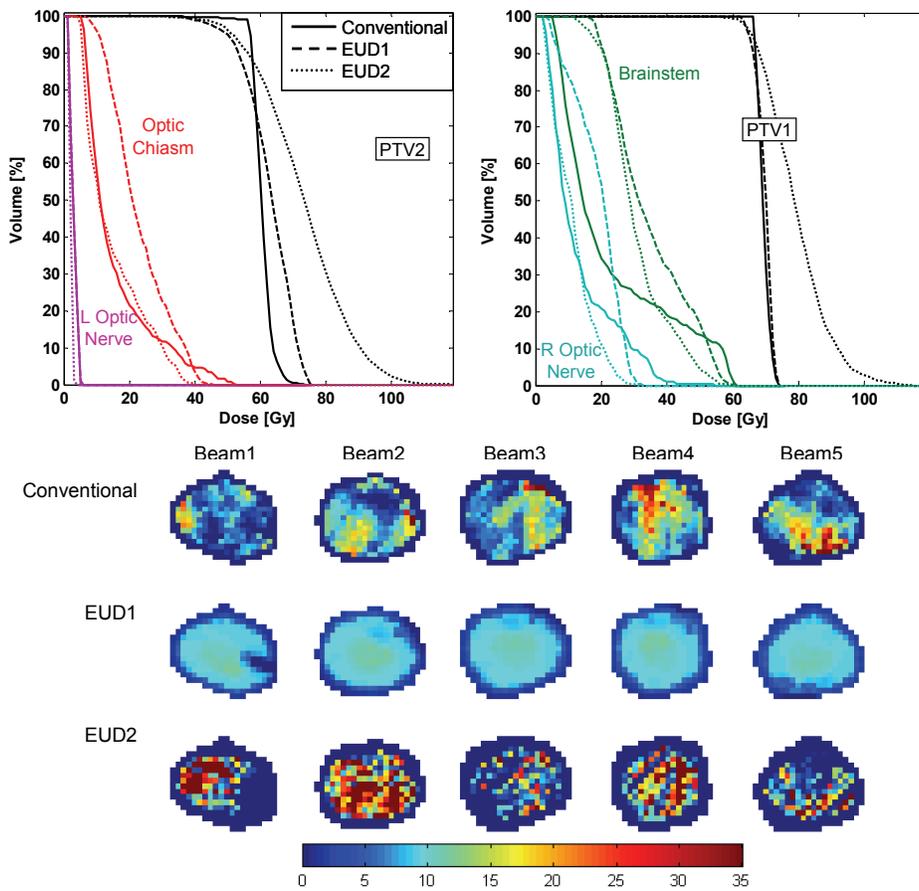


Figure II-22. DVHs and beams for the conventional cost function and the equivalent uniform dose (EUD1) cost function and the stricter EUD cost function (EUD2)

EUD plan, each of the EUD costlets was increased by 5 Gy for targets and decreased by 5 Gy by normal tissues. The results, shown as EUD2, reveal the sensitivity of the EUD formulation and the danger in defining the cost function in an unconventional manner. The monitor units increase by an order of magnitude (2110 MU/Fx) and the DVHs are unreasonable due to the large PTV heterogeneity. This also shows that it is very difficult to ascertain which types of cost function formulations may or may not promote smoother intensity patterns.

This chapter discusses how complex the IMRT planning process is, and demonstrates that many different parameters can affect the integrity of the planned IMRT fluence patterns. The most important part of the IMRT planning is to ensure that the optimization process results in a high quality accurate IMRT plan. Unfortunately, this high quality plan may be comprised of very complex IMRT beams that can have many negative implications. Some of the negative implications of complex IMRT plans such as poor delivery accuracy can be improved by other methods, such as improving the leaf sequencing algorithm. However, the actual complexity of the fields comes from a variety of sources including the discrete point sampling of regions of interest, the optimization method, and the objectives used to define the inverse planning cost function. We have reviewed the consequences of IMRT beam complexity and shown several examples from our optimization system. In addition, we have reviewed and investigated some of the known sources of IMRT beamlet complexity and also demonstrated that point sampling can also be a large source of IMRT beam complexity. This chapter has shown that proper point sampling and use of a deterministic optimization algorithm can help to reduce beam complexity. However, we have also shown that the majority of the modulation seen in IMRT fields is either simply from the plan trying to optimize tradeoffs between competing objectives, or from the degeneracy of the IMRT problem itself. Thus, reduction of IMRT beam complexity is a task that must be undertaken inside the IMRT planning system. The remainder of this work will therefore focus on a variety of different methods that we have developed, implemented, and analyzed to control IMRT beam complexity inside the optimization process, while still maintaining the dosimetric advantages that IMRT was invented to produce.

References

- ¹ "Intensity-modulated radiotherapy: current status and issues of interest," *Int J Radiat Oncol Biol Phys* **51**, 880-914 (2001).
- ² B. A. Fraass and D. L. McShan, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F.H. van der Giessen, H.J. van Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 273-276.
- ³ B. A. Fraass, D. L. McShan, and K. J. Weeks, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F H van_der_Giessen, H J van_Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 193-196.
- ⁴ B. A. Fraass, D. L. McShan, R. K. TenHaken, and K. M. Hutchins, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F H van_der_Giessen, H J van_Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 521-525.
- ⁵ D. L. McShan and B. A. Fraass, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F H van_der_Giessen, H J van_Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 41-44.
- ⁶ D. L. McShan, B. A. Fraass, and A. S. Lichter, "Full integration of the beam's eye view concept into computerized treatment planning," *Int J Radiat Oncol Biol Phys* **18**, 1485-1494 (1990).
- ⁷ M. L. Kessler, D. L. McShan, M. Epelman, K. A. Vineberg, A. Eisbruch, T. S. Lawrence, and B. A. Fraass, "Costlets: a generalized approach to cost functions for automated optimization of IMRT treatment plans," *Optimization and Engineering* **6**, 421-448 (2005).
- ⁸ J. H. Kim, N. Dogan, D. L. McShan, and M. L. Kessler, "An AVS-based system for optimization of conformal radiotherapy treatment plans," *Proceedings of 1995 International Advanced Visual Systems User and Developer Conference*, Boston MA, 1995.
- ⁹ T. R. Mackie, J. W. Scrimger, and J. J. Battista, "A convolution method of calculating dose for 15-MV x rays," *Med. Phys.* **12**, 188-196 (1985).
- ¹⁰ T. R. Bortfeld, D. L. Kahler, T. J. Waldron, and A. L. Boyer, "X-ray field compensation with multileaf collimators," *Int J Radiat Oncol Biol Phys* **28**, 723-730 (1994).
- ¹¹ E. J. Hall and C. Wu, "Radiation induced second cancers: The impact of 3D-CRT and IMRT," *Int J Radiat Oncol Biol Phys* **56**, 83-88 (2003).
- ¹² R. Mohan, M. Arnfield, S. Tong, Q. Wu, and J. Siebers, "The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy," *Med. Phys.* **27**, 1226-1237 (2000).
- ¹³ M. Alber, G. Meedt, F. Nusslin, and R. Reemtsen, "On the degeneracy of the IMRT optimization problem," *Med. Phys.* **29**, 2584-2589 (2002).
- ¹⁴ R. A. Siochi, "Modifications to the IMFAST leaf sequencing optimization algorithm," *Med Phys* **31**, 3267-3278 (2004).
- ¹⁵ D. Cao, M. A. Earl, S. Luan, and D. M. Shepard, "Continuous intensity map optimization (CIMO): a novel approach to leaf sequencing in step and shoot IMRT," *Med Phys* **33**, 859-867 (2006).

- ¹⁶ S. M. Crooks, L. F. McAven, D. F. Robinson, and L. Xing, "Minimizing delivery time and monitor units in static IMRT by leaf-sequencing," *Phys. Med. Biol.* **47**, 3105-3116 (2003).
- ¹⁷ M. Langer, V. Thai, and L. Papiez, "Improved leaf sequencing reduces segments or monitor units needed to delivery IMRT using multileaf collimators," *Med. Phys.* **28**, 2450-2458 (2001).
- ¹⁸ G. A. Ezzell and S. Chungbin, "The overshoot phenomenon in step-and-shoot IMRT delivery," *J Appl Clin Med Phys* **2**, 138-148 (2001).
- ¹⁹ O. A. Zeidan, J. G. Li, M. Ranade, A. M. Stell, and J. F. Dempsey, "Verification of step-and-shoot IMRT delivery using a fast video-based electronic portal imaging device," *Med Phys* **31**, 463-476 (2004).
- ²⁰ G. N. Grigorov, J. C. Chow, and R. B. Barnett, "Dosimetry limitations and a dose correction methodology for step-and-shoot IMRT," *Phys Med Biol* **51**, 637-652 (2006).
- ²¹ A. M. Stell, J. G. Li, O. A. Zeidan, and J. F. Dempsey, "An extensive log-file analysis of step-and-shoot intensity modulated radiation therapy segment delivery errors," *Med Phys* **31**, 1593-1602 (2004).
- ²² P. Zygmanski, J. H. Kung, S. B. Jiang, and L. Chin, "Dependence of fluence errors in dynamic IMRT on leaf-positional errors varying with time and leaf number," *Med Phys* **30**, 2736-2749 (2003).
- ²³ C. S. Chui, M. F. Chan, E. Yorke, S. Spirou, and C. C. Ling, "Delivery of intensity-modulated radiation therapy with a conventional multileaf collimator: comparison of dynamic and segmental methods," *Med Phys* **28**, 2441-2449 (2001).
- ²⁴ G. Nicolini, A. Fogliata, and L. Cozzi, "IMRT with the sliding window: comparison of the static and dynamic methods. Dosimetric and spectral analysis," *Radiother Oncol* **75**, 112-119 (2005).
- ²⁵ W. Que, J. Kung, and J. Dai, "'Tongue-and-groove' effect in intensity modulated radiotherapy with static multileaf collimator fields," *Phys Med Biol* **49**, 399-405 (2004).
- ²⁶ S. Kamath, S. Sahni, J. Palta, S. Ranka, and J. Li, "Optimal leaf sequencing with elimination of tongue-and-groove underdosage," *Phys Med Biol* **49**, N7-19 (2004).
- ²⁷ S. Luan, C. Wang, D. Z. Chen, X. S. Hu, S. A. Naqvi, X. Wu, and C. X. Yu, "An improved MLC segmentation algorithm and software for step-and-shoot IMRT delivery without tongue-and-groove error," *Med Phys* **33**, 1199-1212 (2006).
- ²⁸ E. J. Hall, "Intensity-modulated radiation therapy, protons, and the risk of second cancers," *Int J Radiat Oncol Biol Phys* **65**, 1-7 (2006).
- ²⁹ S. F. Kry, M. Salehpour, D. S. Followill, M. Stovall, D. A. Kuban, R. A. White, and Rosen, II, "Out-of-field photon and neutron dose equivalents from step-and-shoot intensity-modulated radiation therapy," *Int J Radiat Oncol Biol Phys* **62**, 1204-1216 (2005).
- ³⁰ J. Z. Wang, X. A. Li, W. D. D'Souza, and R. D. Stewart, "Impact of prolonged fraction delivery times on tumor control: A note of caution for intensity-modulated radiation therapy (IMRT)," *Int J Radiat Oncol Biol Phys* **57**, 543-552 (2003).

- ³¹ H. Paganetti, "Changes in tumor cell response due to prolonged dose delivery times in fractionated radiation therapy," *International Journal of Radiation Oncology, Biology, Physics* **63**, 892-900 (2005).
- ³² J. F. Fowler, J. S. Welsh, and S. P. Howard, "Loss of biological effect in prolonged fraction delivery," *International Journal of Radiation Oncology, Biology, Physics* **59**, 242-249 (2004).
- ³³ M. Schwarz, J. Van der Geer, M. Van Herk, J. V. Lebesque, B. J. Mijnheer, and E. M. Damen, "Impact of geometrical uncertainties on 3D CRT and IMRT dose distributions for lung cancer treatment," *Int J Radiat Oncol Biol Phys* **65**, 1260-1269 (2006).
- ³⁴ C. X. Yu, D. A. Jaffray, and J. W. Wong, "The effects of intra-fraction organ motion on the delivery of dynamic intensity modulation," *Phys Med Biol* **43**, 91-104 (1998).
- ³⁵ J. Seco, G. C. Sharp, J. Turcotte, D. Gierga, T. Bortfeld, and H. Paganetti, "Effects of organ motion on IMRT treatments with segments of few monitor units," *Med Phys* **34**, 923-934 (2007).
- ³⁶ W. A. Tomé and J. F. Fowler, "On cold spots in tumor subvolumes," *Med Phys* **29**, 1590-1598 (2002).
- ³⁷ J. Llacer, N. Agazaryan, T. D. Solberg, and C. Promberger, "Degeneracy, frequency response and filtering in IMRT optimization," *Phys. Med. Biol.* **49**, 2853-2880 (2004).
- ³⁸ C. Coolens, P. M. Evans, J. Seco, and S. Webb, "Analysis of stochastic noise in intensity-modulated beams," *Phys. Med. Biol.* **49**, 3857-3875 (2004).
- ³⁹ J. O. Deasy, "Multiple local minima in radiotherapy optimization problems with dose-volume constraints," *Med Phys* **24**, 1157-1161 (1997).
- ⁴⁰ Q. Wu and R. Mohan, "Multiple local minima in IMRT optimization based on dose-volume criteria," *Med Phys* **29**, 1514-1527 (2002).
- ⁴¹ J. Llacer, J. O. Deasy, T. R. Portfeld, T. D. Solberg, and C. Promberger, "Absence of multiple local minima effects in intensity modulated optimization with dose-volume constraints," *Phys Med Biol* **48**, 183-210 (2003).
- ⁴² A. Niemierko, "A generalized concept of equivalent uniform dose (EUD)," *Med Phys* **26**, 1100 (1999).

CHAPTER III.

REDUCING THE PROBLEM SCALE: OPTIMIZATION OF MATHEMATICAL SURFACES VERSUS INDIVIDUAL BEAMLET INTENSITIES

III.A. Motivation

In Chapter I, we described the large-scale optimization problem undertaken by planning intensity-modulated radiation therapy treatments, and in Chapter II we discussed some of the reasons why IMRT beams have a tendency, without intervention, to become discontinuous across a field and contain large intensity variations over small areas of the beam. Due to the degeneracy of many IMRT problems, it is possible that the high degree of complexity observed in some beamlet patterns is excessive and not necessary for a high quality IMRT plan. In fact, this complexity can result in lower quality IMRT plans in several areas, including treatment planning, delivery, and quality assurance. Thus, approaches to limit beam complexity have been pursued by many researchers, including beam smoothing¹⁻⁶ and beamlet restrictions⁷. Some of our developments in these areas are discussed in Chapters IV and V. Another possible method to reduce the unnecessary complexity of optimized intensity patterns is to explicitly reduce the degrees of freedom allowed in the optimization problem. One strategy that can achieve this is the use of mathematical surfaces (instead of individual beamlets) to represent and optimize IMRT beams. This technique allows for a potentially large reduction in the number of optimization variables and can prevent the optimized beamlet intensity patterns from becoming discontinuous, while still allowing for intensity gradients in regions where they are necessary to produce a quality IMRT plan. Markman *et al.* has studied the use of basis function parameters to optimize and represent beamlet intensity profiles in 2D, and has demonstrated the potential for reducing beam complexity and planning time without adversely affecting plan quality.⁸ Their work was a proof-of-principle using a simple 2D geometry with an approximate dose calculation algorithm.

This chapter discusses our independent investigation into the feasibility of basis function optimization for use in 3D inverse planning. The first section highlights the theory of basis function optimization and the implementation of basis function optimization into our IMRT treatment planning optimization infrastructure. The second section validates this implementation and studies several different basis function choices in a test phantom. Then, clinical examples in brain, prostate, and head/neck are presented.

III.B. Methods

III.B.1. Theory of Basis Function Optimization

As discussed in Chapter II, the most common form of IMRT optimization involves dividing each beam into a grid of smaller beams, or “beamlets”. Then, the weights of the individual beamlets are found that minimize an objective function, $f(\mathbf{b})$, where f can have many forms and \mathbf{b} is the vector of beamlet weights. A general form for f can be written as:

$$f = \sum_{m=1}^M p_m f_m , \quad (\text{III.1})$$

where M is the number of treatment planning goals, p_m is the importance weight of goal m and f_m is the objective function for goal m . Most commonly, f_m will be defined for a certain group of voxels that comprise a region of interest in the treatment plan, such as the planning target volume (PTV) or an organ-at-risk (OAR). Most objective functions are dose-based; an example of a minimum dose target objective is:

$$f_{ptv} = \frac{1}{N_{ptv}} \sum_{i=1}^{N_{ptv}} (d_t - d_i)^2 \text{ for all } d_i < d_t , \quad (\text{III.2})$$

where N_{ptv} is the number of voxels in the PTV, d_t is the minimum desired target dose, and d_i is the dose in the i^{th} PTV voxel. Similar dose-based objectives can be written for OARs. A variety of cost function forms in the UMOpt planning systems are given by Kessler *et al.*⁹. The dose in each voxel is the sum of the dose contributions to that voxel from each beamlet defined in the optimization system,

$$d_i = \sum_{j=1}^J D_{ij} b_j \quad (\text{III.3})$$

where D_{ij} is the dose to point i from beamlet j , b_j is the weight of beamlet j , and J is the total number of beamlets in the plan.

The goal of the optimization algorithm is to find the vector of beamlet weights, \mathbf{b} , that minimizes the total objective function value $f = f(\mathbf{b})$. Since beamlet dimensions usually coincide with the size of the leaves in the multi-leaf collimator on the treatment machine, they are generally defined to be on the order of 1 cm² or less. At the University of Michigan, linear accelerators with MLC leaves of either 1 cm or 0.5 cm width are used. Therefore, a typical 7 beam IMRT case with 5 cm by 5 cm beams and 0.5 cm² beamlet will have a total of 700 optimization variables. In larger treatment fields, such as those used in head-and-neck cancer, there can be well over 1000 beamlets. As stated in Chapter II, this leads to a large number of degrees of freedom in an IMRT optimization problem and increases the probability of the plan being degenerate. This can lead to increased “noise” in the beamlet profiles, resulting in inefficient treatment deliveries and many other potential complications (see Chapter II). Therefore, to reduce the degrees of freedom and prevent “spikey” beamlet patterns, a smaller set of optimization variables can be used. Increasing the size of the beamlets has been studied and found to be inadequate to meet the cost function goals at the field edges, as well as around interfaces between targets and normal tissues¹⁰.

Another way to reduce the number of optimization variables, while still preserving the possibility for slowly varying or steep gradients in essential areas, is to optimize the parameters, or coefficients for a mathematical surface instead of the individual beamlet intensities. To represent this new optimization problem, the objective function can be written as a function of surface parameters, $f(\mathbf{c})$ where \mathbf{c} is a vector of function coefficients. These coefficients are related to the individual beamlet values by the spatial location of each of the beamlets; the number of coefficients will depend on the surface or basis function set that is being used. For instance, the intensity of the beamlet at the (x,y) spatial position of a beam can be given by:

$$b(x,y) = c_1 g_1(x,y) + c_2 g_2(x,y) + c_3 g_3(x,y) + \dots \quad (\text{III.4})$$

Here, c_i is the optimized coefficient for basis function $g_i(x,y)$. An advantage to representing the beam as a linear combination of the basis functions is that the partial derivative, $\partial \mathbf{b} / \partial \mathbf{c}$, can be used to calculate the partial derivative of the cost function with respect to the basis function coefficients, $\partial f / \partial \mathbf{c}$. This calculation is necessary to use the quasi-Newton optimization algorithm to optimize the coefficients. Chapter II pointed out the advantages of this algorithm over stochastic algorithms such as simulated annealing in terms of speed and the fact that simulated annealing can introduce extra “noise” into the beamlet values. The latter may not be problem with basis function optimization, but the increase in speed is very important for this technique to be feasible in the planning of clinical radiotherapy treatments. Thus, if $\mathbf{b}(x,y)$ is a linear function of \mathbf{c} , then the cost function gradient with respect to the coefficients can be easily calculated by:

$$\frac{\partial f}{\partial \mathbf{c}} = \frac{\partial f}{\partial \mathbf{b}} \cdot \frac{\partial \mathbf{b}}{\partial \mathbf{c}}. \quad (\text{III.5})$$

There are variety of options for the number and structure of the basis functions, g . The number of basis functions chosen will depend on the complexity of patient geometry as well as the stringency of the cost function. The number of basis functions to be optimized should be sufficiently smaller than the number of beamlets in order the realize the benefits of the reduced number of variables and degrees of freedom in the plan. The main guideline for choosing a set of basis functions is,

$$\sum_i c_i g_i(x,y) \geq 0 \quad \forall (x,y) \in R, \quad (\text{III.6})$$

or, in other words, the weighted coefficient sum of the basis functions must be equal to great than zero in the bounds of the beam coordinates, R . This is a necessary requirement because the beamlet intensities must be equal to or greater than zero. Negative coefficients are acceptable so long as Equation II.6 is satisfied. It is possible to reset the beamlet weights to zero if they are calculated to be negative, but this introduces errors into the calculation of the cost function gradients in Equation III.5. If the errors are small enough so that the gradient is still well approximated, then the

quasi-Newton algorithm may still perform well. However, the gradients are poorly approximated, then it would be necessary to use a non-gradient optimization algorithm (i.e. simulated annealing).

In general, we have identified two different approaches to choosing g : (i) Choose a surface function $k(x,y)$ with pre-defined terms with leading coefficients that can serve as the optimization variables, or (ii) Choose a pre-defined set of local functions of finite support with fixed locations and optimize the amplitude of the functions so that a sum of the individual functions represents the optimal intensity map. An example of (i) would be the optimization of a two-dimensional polynomial. An example of (ii) would be a set of radial basis functions centered at different locations. Both polynomials and radial basis functions are very commonly used in surface fitting. In this work, we have used these two common basis sets to test the feasibility of basis function optimization for 3D IMRT planning. (Note: implementation details on several additional basis function sets which were studied in a preliminary way are included in the Appendix at the end of this chapter).

III.B.2. Polynomial Basis Functions

Polynomials are very common and widely-used fitting functions. In this application, the polynomial term coefficients are optimized to result in a mathematical surface that can represent the spatial beamlet intensity. Thus, the value of the beamlet at the (x,y) beam position can be defined as:

$$\begin{aligned}
 b(x,y) = & c_0 + c_1x + c_2y + \dots \\
 & c_3x^2 + c_4y^2 + c_5xy + \dots \\
 & c_6x^3 + c_7y^3 + c_8x^2y + c_9xy^2 + \dots \\
 & c_{10}x^4 + c_{11}y^4 + c_{12}x^3y + c_{13}xy^3 + c_{14}x^2y^2 + \dots
 \end{aligned}
 \tag{III.7}$$

where c_0 through c_{14} shown are the optimization variables for a 4th order polynomial basis function set. Figure III-1 illustrates an example beam that could be made by using a 2nd order polynomial. Polynomial functions are easily differentiable and are a simple choice to represent an IMRT beam. However, Equation III.7 does not require that $b(x,y) > 0$ so this property must be applied separately during the optimization process. This can adversely affect the accuracy of the gradient calculations used within the optimization process and therefore the method will also be tested with simulated

annealing to determine whether or not the gradient errors are large enough to adversely affect the performance of the quasi-Newton algorithm.

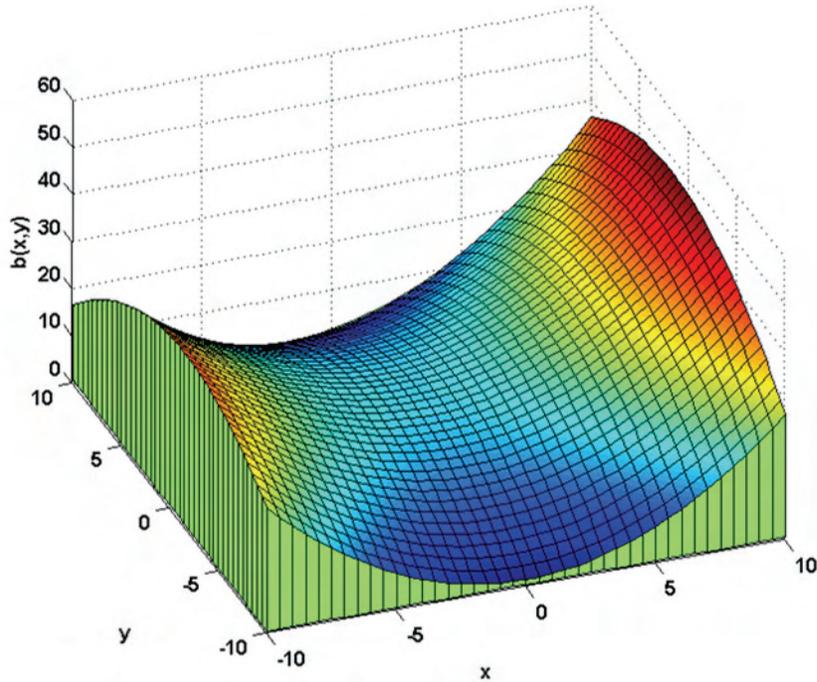


Figure III-1. An example of a 2D 2nd order polynomial that could be used to calculate the beamlet intensities. Here $\mathbf{c} = [16.5 \ 0.5 \ 0 \ 0.25 \ -0.15 \ 0.05]$ in Equation III.7.

III.B.3. Radial Basis Functions

A radial basis function (RBF) is defined as a real-valued function whose value depends only on the distance from the origin. They are commonly used in function fitting, time series prediction, and control. Radial basis functions are also commonly used as basis libraries in artificial neural networks. For use in defining beamlet fluence patterns, simple Gaussian RBFs are defined on a grid across each beam, and the weights of each of the individual functions serve as the optimization variables. An individual Gaussian RBF has the form:

$$G_i(x, y) = e^{-a \left[(x-t_{x,i})^2 + (y-t_{y,i})^2 \right]}, \quad (III.8)$$

where a is the width, or dilation of the function, $t_{x,i}$ is the x translation and $t_{y,i}$ is the y translation of the basis function. Figure III-2(a) displays a single Gaussian RBF centered at the origin, with $a = 0.25$ and an amplitude of 10. In order to represent the IMRT beam, the amplitudes of individual Gaussian RBFs will be optimized and the spatial beamlet intensity can be represented by

$$b(x,y) = \sum_{i=1}^N c_i G_i(x,y) , \quad (III.9)$$

where N is the number of basis functions and c_i is the amplitude of the i^{th} Gaussian RBF. In order to reduce the number of optimization variables, $t_{x,i}$ and $t_{y,i}$ are pre-determined before optimization so that positions of the RBFs are equally spaced across each beam. This means that that $t_{x,i}$ and $t_{y,i}$ lie on a uniform grid across the beam. In order to maintain $b(x,y)$ as a linear function of the optimization variables, the dilation of the basis functions, a , is also held constant for each beam. Figure III-2(b) shows a 3 x 3 RBF grid

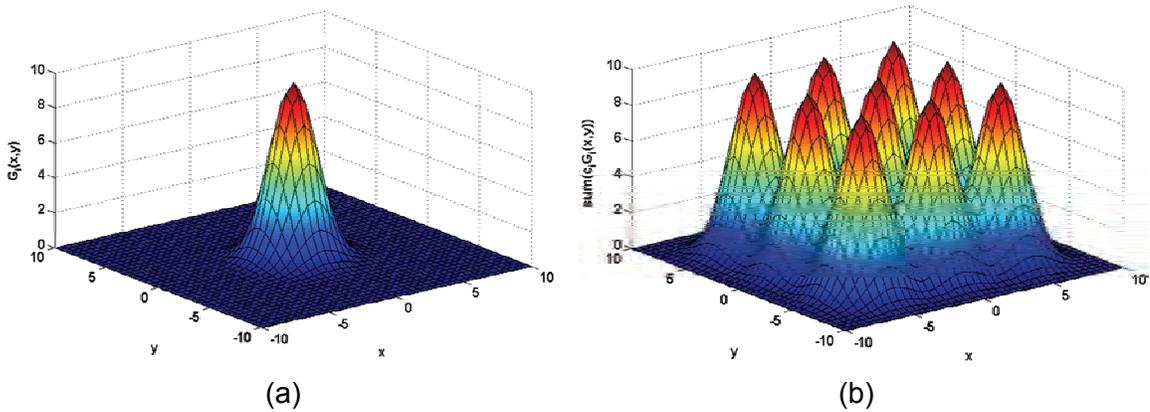


Figure III-2. (a) A single Gaussian radial basis function centered at (0,0) with $a = 0.25$ and amplitude 10, and (b) A grid of radial basis functions identical to (a), but centered at various (x,y) positions in the field.

with $a = 0.25$ and all $c_i = 10$, and Figure III-3 shows the same grid with varying amplitudes to demonstrate what a possible beam could look like. To determine a good choice for a , possible values were iterated through manually. A logical hypothesis is that a will be related to the grid spacing between basis functions. Therefore, an effort was made to determine the relationship between the number of basis functions and a so that it could be set automatically in the future.

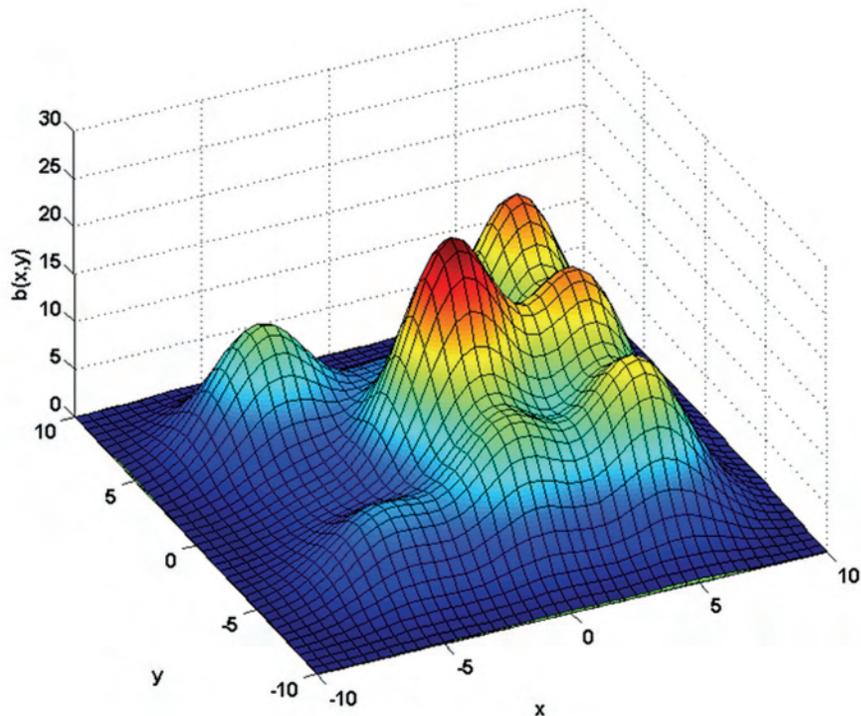


Figure III-3. An example of a beam that could be made optimizing the amplitudes of a 3 x 3 grid of Gaussian radial basis functions.

III.B.4. Implementation of Basis Function Optimization

To most efficiently study the use of basis function optimization at the University of Michigan, it was necessary to incorporate the necessary tools into the existing optimization framework, UMOpt. Many of the existing modules and data structures were used in the basis function implementation. Details regarding the code additions and changes are given in Appendix A. In general terms, the existing system was enhanced to perform optimization based on the basis function parameters or coefficients instead of beamlets. Figure III-1 illustrates an overview of how the basis function optimization routine is incorporated into the conventional beamlet optimization infrastructure, and Figure III-2 shows the interface that has been developed for basis function optimization in UMOpt. Starting at the top of Figure III-1, if basis function optimization is on, then UMOpt knows to create the parameters or optimization variables to be the basis function coefficient values, \mathbf{c} . In Figure III-2, this is specified by toggling the “Basis Fcn Opt On” button. In addition, the type of basis function to be used and number of coefficients per

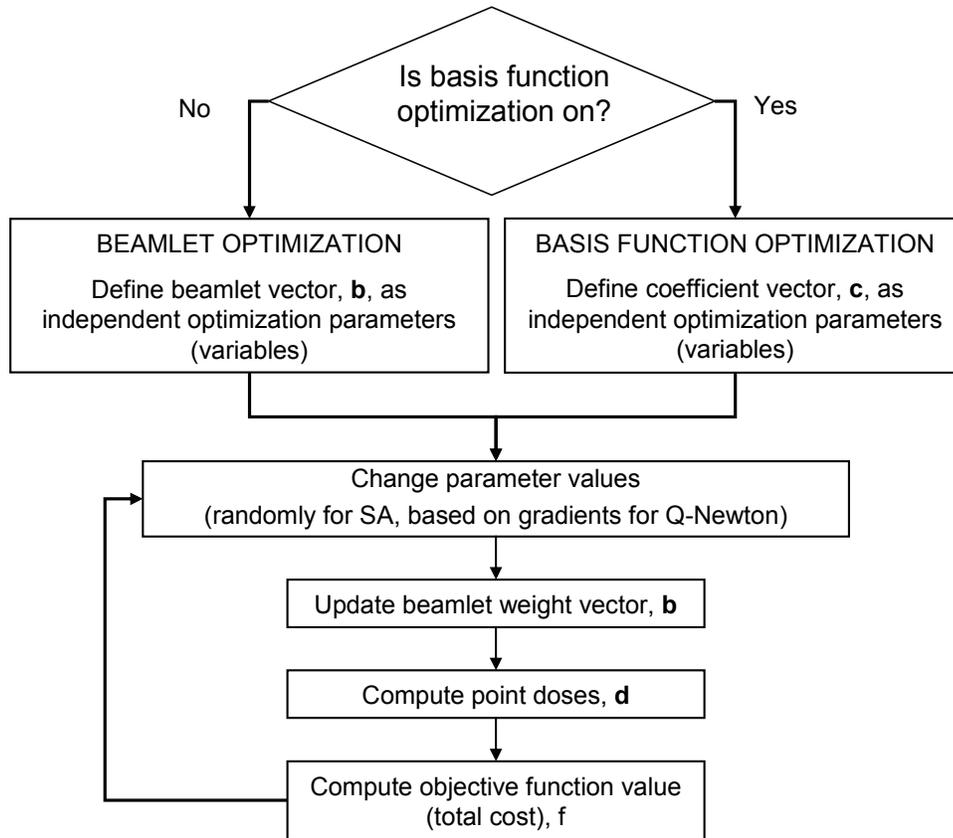


Figure III-4. Flow chart representing how basis function optimization fits into the conventional beamlet optimization infrastructure.

Table III-1. Basis function optimization parameters to be used in UMOpt

Basis Type	Description	# Coefficients per Beam	Details
1	Polynomial	6	2nd order
		10	3rd order
		15	4th order
		21	5th order
		28	6th order
		36	7th order
		45	8th order
6	Gaussian Radial Basis Functions	4	2 x 2 grid
		16	4 x 4 grid
		25	5 x 5 grid
		36	6 x 6 grid
		49	7 x 7 grid
		64	8 x 8 grid

beam must be specified. Each of the basis function types defined in UMOpt is shown in the Appendix, but the basis types and possible numbers of coefficients for the two main types that we are studying here (polynomials and Gaussian RBFs) are shown in Table III-1.

After specifying the basis type and number of coefficients, the optimization variables are created by using the “Create Parameters” button, which invokes a routine that initializes the optimization variables and specifies the indices and limits for each of the

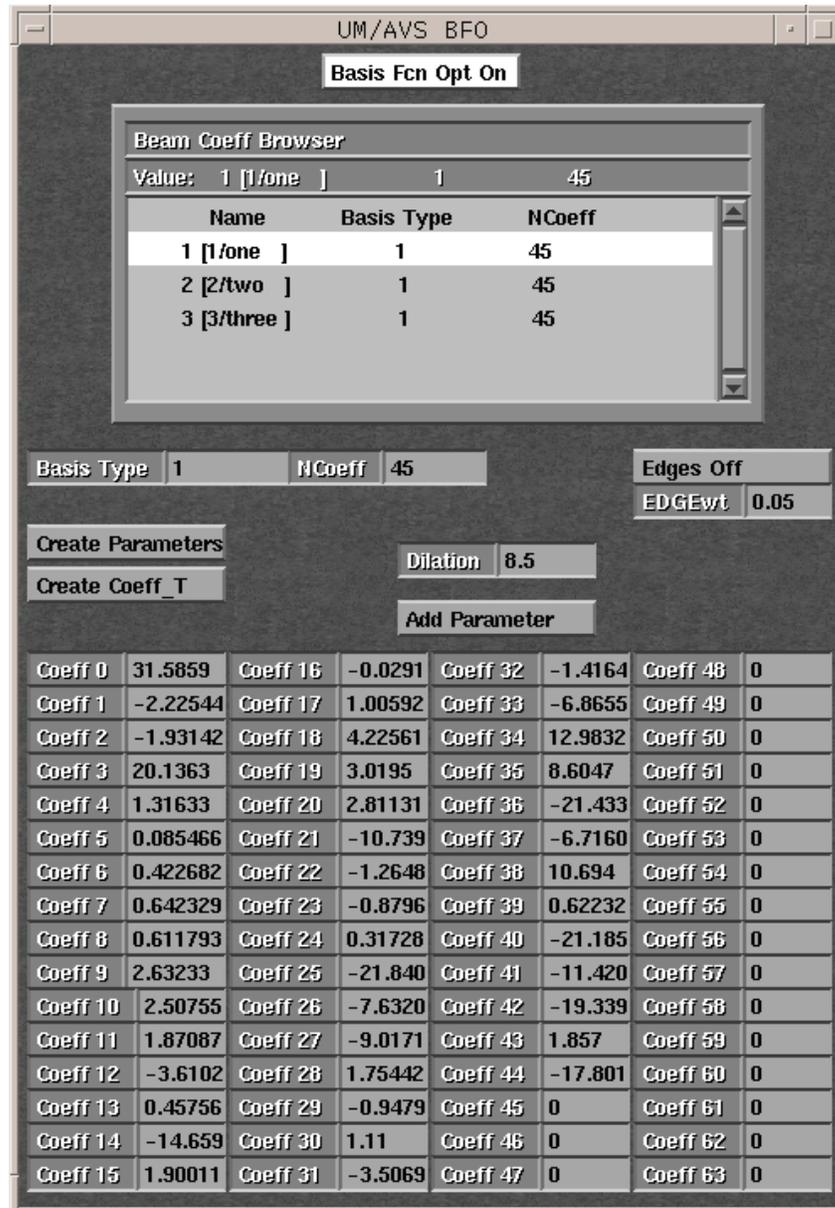


Figure III-5. Basis function optimization user interface that has been implemented into UMOpt.

variables. After initialization, the “Create Coeff_T” routine creates a transposed matrix of the basis function terms to be used in the optimization process to calculate the beamlet values and the gradients. For example, for a one-beam plan with J beamlets, the matrix for a 2nd order polynomial function would be:

$$\mathbf{M} = \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & y_1^2 & x_1y_1 \\ 1 & x_2 & y_2 & x_2^2 & y_2^2 & x_2y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_J & y_J & x_J^2 & y_J^2 & x_Jy_J \end{bmatrix} \quad (\text{III.10})$$

where x_i and y_i are the x and y coordinates of the i^{th} beamlet. Then, calculating the beamlet values is simple matrix multiplication, $\mathbf{b}=\mathbf{M}\mathbf{c}$:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_J \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & y_1^2 & x_1y_1 \\ 1 & x_2 & y_2 & x_2^2 & y_2^2 & x_2y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_J & y_J & x_J^2 & y_J^2 & x_Jy_J \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{bmatrix} \quad (\text{III.11})$$

Once the beamlet values are assigned according to the current coefficient values inside the optimization algorithm, the point doses are calculated according to Equation III.3, and the total objective function value cost is calculated as it is in beamlet-based optimization. To calculate the quasi-Newton optimization algorithm search direction, the partial derivative, $\partial f/\partial \mathbf{c}$ must be calculated. As mentioned previously, this is very simple to calculate for linear functions because $\mathbf{M} = \partial \mathbf{b}/\partial \mathbf{c}$. Inserting the 2nd order polynomial example into Equation III.5, we have:

$$\frac{\partial f}{\partial \mathbf{c}} = \begin{bmatrix} \frac{\partial f}{\partial b_1} \\ \frac{\partial f}{\partial b_2} \\ \vdots \\ \frac{\partial f}{\partial b_J} \end{bmatrix} \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & y_1^2 & x_1 y_1 \\ 1 & x_2 & y_2 & x_2^2 & y_2^2 & x_2 y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_J & y_J & x_J^2 & y_J^2 & x_J y_J \end{bmatrix}. \quad (\text{III.12})$$

Once the search direction and step size are calculated, the optimization process proceeds until the convergence criteria are met for the algorithm.

The “Add Parameter”, and “Edge Off” function buttons are covered in the discussion sections.

III.B.5. Treatment Planning and Analysis

To validate and investigate our implementation of basis function optimization for 3D inverse planning of IMRT treatments, several cases were optimized with three different sets of optimization variables, including (i) conventional beamlets, (ii) Gaussian radial basis function coefficients using 4^2 , 5^2 , 6^2 and 7^2 grids, and (iii) polynomial term coefficients for 2^{nd} , 4^{th} , 6^{th} , and 8^{th} order polynomials.

The initial evaluation of basis function optimization was performed on a phantom, and then three clinical examples (brain, prostate, and head/neck) were optimized using

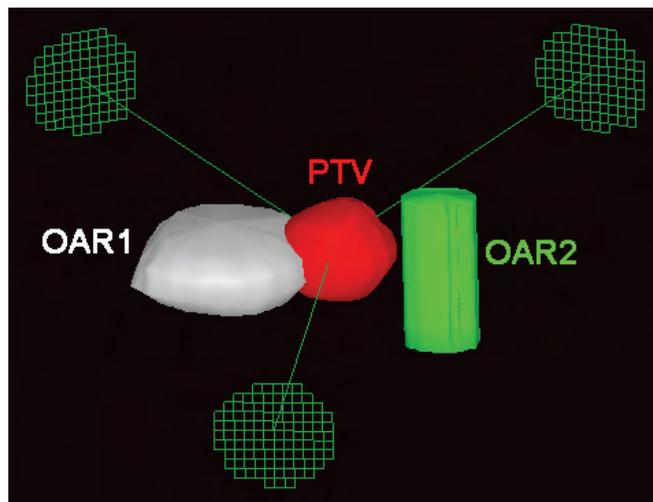


Figure III-6. Geometry and beam arrangement for the phantom case used in the basis function optimization characterization studies.

the same methods. The planning objectives and corresponding costlets applied in UMOpt for all three cases are shown in Table III-2, and the geometry and beam arrangement for the phantom case is shown in Figure III-6. The clinical plan objectives were chosen based on protocols in the Radiation Oncology Department at the University of Michigan, although the head/neck objective function did not include the swallowing structures, which are generally present for clinical cases. Five non-axial beams were used in the brain case, and the prostate and head/neck cases were optimized using 7 equally spaced axial beams. All cases were planned for a 6 MV linear accelerator (Varian Medical Systems, 21EX) with 120 leaf MLC (0.5 and 1.0 cm leaf widths). Dose calculations for the inverse planning system were performed by a convolution/superposition algorithm derived from the work of Mackie *et al*¹¹. All cases were optimized using a quasi-Newton-based search strategy, and care was taken to ensure that the optimized plans did not represent significant local minima of the cost function. Leaf sequencing for static MLC (SMLC) delivery was performed with an in-house-developed leaf sequencer based on the method reported by Bortfeld *et al*¹². Delivery sequences allow up to 250 segments per beam, with the goal of achieving a correspondence between planned and delivered intensities of 1 %.

The standard beamlet IMRT results served as a baseline for all plan comparisons. Quantities used for evaluation included dose metrics, DVHs, field modulation and complexity, and monitor units required for SMLC delivery.

Detailed analyses were performed for the different basis function sets to determine their smoothing properties and to gain more information on the pre-defined parameter values (such as the radial basis function dilation) for use in further development of those methods. These results are presented first, and then all methods are compared in the three clinical sites. Effort was made to determine which method is most desirable and efficient in terms of the number of variables needed to produce a satisfactory plan, the tradeoffs in plan quality, and other important factors such as delivery efficiency in terms of the number of monitor units required for delivery of the IMRT plan. Monitor units (MU), as described in Chapter II, are a measure of the linear accelerator output. At most institutions, linear accelerators are calibrated so that they deliver around 1 cGy/MU. However, IMRT delivery requires that large areas of the field are blocked during the delivery so that many more MU are required to deliver a 2 Gy fraction to the tumor than in conventional radiation therapy. This makes the delivery inefficient and can have potentially dangerous consequences to the patient, which are discussed in Chapter II.

Table III-2. Cost functions used in the phantom and clinical cases

Structure	Objectives	Costlet	Weight
Phantom Case			
PTV	Min Dose > 57 Gy	DVH_GE(57,100)	100
	Max Dose < 63 Gy	DVH_LE(63,0)	100
OAR1	Mean Dose < 15 Gy	MEAN.LE.15	10
	Max Dose < 63 Gy	DVH_LE(63,0)	1
	Max Dose < 0 Gy	DVH_LE(0,0)	0.001
OAR2	Max Dose < 20 Gy	DVH_LE(20,0)	10
	Max Dose < 0 Gy	DVH_LE(0,0)	0.001
Normal Tissue	Max Dose < 63 Gy	DVH_LE(63,0)	1
	Max Dose < 0 Gy	DVH_LE(0,0)	0.001
Brain Case			
PTV1	> 99% volume > 63 Gy	DVH_GE(63,99)	100
	Max Dose < 72 Gy	DVH_LE(72,0)	100
	< 1 % volume > 69 Gy	DVH_LE(69,1)	10
PTV2	Min Dose > 57.5 Gy	DVH_GE(57.5,100)	100
	Max Dose < 69 Gy	DVH_LE(69,0)	100
Optic Nerves	Max Dose < 60 Gy	DVH_LE(60,0)	100
	Mean Dose < 0 Gy	MEAN.LE.0	0.0005
Optic Chiasm	Max Dose < 60 Gy	DVH_LE(60,0)	100
	Mean Dose < 0 Gy	MEAN.LE.0	0.0005
Brainstem	Max Dose < 65 Gy	DVH_LE(65,0)	100
	Mean Dose < 0 Gy	MEAN.LE.0	0.0005
Normal Tissue	Max Dose < 72 Gy	DVH_LE(72,0)	1
	Mean Dose < 0 Gy	MEAN.LE.0	0.0001
Prostate Case			
Prostate + 5 mm	> 99% volume > 78.85 Gy	DVH_GE(78.85,99)	100
	< 1 % volume > 80 Gy	DVH_LE(80,1)	100
Rectum	Mean Dose < 25 Gy	MEAN.LE.25	10
	Max Dose < 80 Gy	DVH_LE(80,0)	1
	Max Dose < 0 Gy	DVH_LE(0,0)	0.001
Bladder	Max Dose < 80 Gy	DVH_LE(80,0)	1
	Max Dose < 0 Gy	DVH_LE(0,0)	0.001
Penile Bulb	Max Dose < 80 Gy	DVH_LE(80,0)	10
	Max Dose < 0 Gy	DVH_GE(0,0)	0.0005
Femora	Max Dose < 50 Gy	DVH_LE(50,0)	1
	Max Dose < 0 Gy	DVH_LE(0,0)	0.001
Normal Tissue	Max Dose < 77 Gy	DVH_LE(77,0)	1
	Max Dose < 0 Gy	DVH_LE(0,0)	0.0005

Head/Neck Case			
PTV1	> 99% volume > 69.3 Gy	DVH_GE(69.3,99)	100
	< 1 % volume > 77 Gy	DVH_LE(77,1)	100
PTV2	> 99% volume > 63.4 Gy	DVH_GE(63.4,99)	100
	< 1 % volume > 70.4 Gy	DVH_LE(70.4,1)	100
Rt Nodal PTV	> 99% volume > 63.4 Gy	DVH_GE(63.4,99)	100
	< 1 % volume > 70.4 Gy	DVH_LE(70.4,1)	100
Lt Nodal PTV	> 99% volume > 59.4 Gy	DVH_GE(59.4,99)	100
	< 1 % volume > 66 Gy	DVH_LE(66,1)	100
Spinal Cord	Max Dose < 50 Gy	DVH_LE(50,0)	1000
	Max Dose < 0 Gy	DVH_LE(0,0)	0.0001
Brainstem	Max Dose < 54 Gy	DVH_LE(54,0)	1000
	Max Dose < 0 Gy	DVH_LE(0,0)	0.0001
Mandible	Max Dose < 73.5 Gy	DVH_LE(73.5,0)	10
	Max Dose < 0 Gy	DVH_LE(0,0)	0.0001
Ipsilateral Parotid	Mean Dose < 26 Gy	MEAN.LE.26	1
	Max Dose < 0 Gy	DVH_LE(0,0)	0.0001
Oral Cavity	Mean Dose < 49 Gy	MEAN.LE.49	1
	Max Dose < 77 Gy	DVH_LE(77,0)	1
	Max Dose < 0 Gy	DVH_LE(0,0)	0.0001
Normal Tissue	Max Dose < 73.5 Gy	DVH_LE(73.5,0)	1
	Max Dose < 0 Gy	DVH_LE(0,0)	0.00001

III.C. Results

III.C.1. CT Phantom

The following two sections summarize the characterization of basis function optimization in the CT phantom using polynomial functions and Gaussian radial basis functions. Both methods were successful in producing adequate IMRT plans, with large reductions in the number of optimization variables, beam complexity, and monitor units. However, there are many differences between the two techniques in terms of their behavior with increasing numbers of optimization variables. There are several advantages and limitations that have been encountered as a result of reducing the degrees of freedom in the optimization problem. Both techniques are discussed below and compared at the end of the section.

III.C.1.1. Polynomial Basis Functions

Optimizing polynomials to represent the intensity modulated beams results in much differently-shaped beams than if one were to use beamlets or even a smoothing filter.

With optimized polynomial surfaces, the beams become very smooth and continuous, and the isolated high-intensity peaks that appear in beamlet plans are not present. In this simple-to-moderate complexity example, these smooth beams appear to be adequate to reach the optimization goals, and further, to produce acceptable plans that require many fewer MU to deliver than beamlet plans.

Figure III-7 shows the dose-volume histograms and intensity modulated beams obtained when using a series of different order polynomial functions. As polynomials of increasing order are used, the shape of the beam is only slightly more refined, and the general shape usually stays the same. This suggests that polynomials may not be capable of optimizing plans in which there are many local features that require

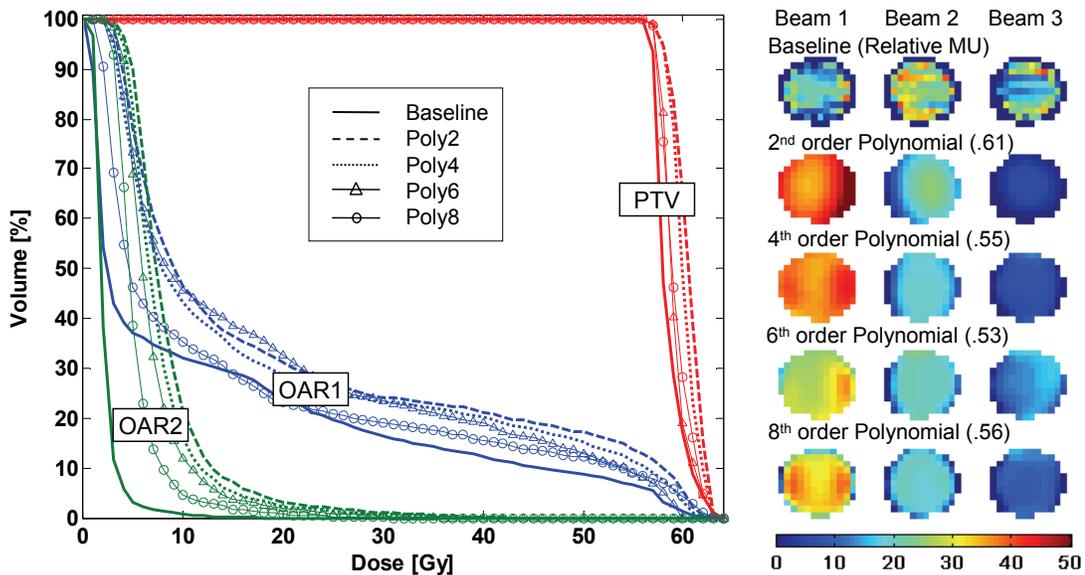


Figure III-7. Dose-volume histograms and the corresponding intensity modulated beams for plans optimized with various order polynomial functions. Also shown are the baseline IMRT results.

specialized beamlet patterns. There is also a very heavy weighting to Beam 1, which could create an undesirable distribution of normal tissue dose along beam 1. Because of the close proximity of the OARs to the PTV and the relative similarity of the importance of target coverage and OAR objectives, it was difficult for the lower-order polynomials to achieve enough modulation to meet both objectives, and therefore, concessions were made. This is clear in the slightly compromised target coverage in the

lower-order polynomial plan DVHs. There was an improvement in target coverage as well as a slight reduction in dose to the normal tissues at higher polynomial orders.

III.C.1.2. Radial Basis Functions

Optimizing a composite set of radial basis functions to obtain an intensity modulated beamlet pattern was feasible and allowed for greater beam modulation and higher intensity gradients than optimization with a polynomial function. However, the dosimetric improvement was not necessarily significantly different between the two functions. Figure III-8 displays the radial basis function optimization results as DVHs and intensity

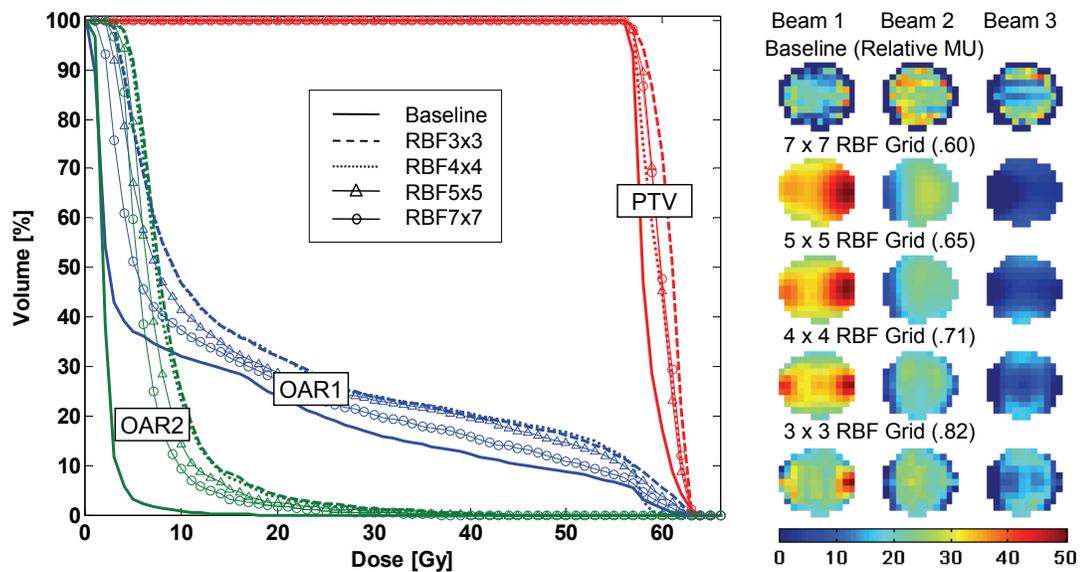


Figure III-8. Dose-volume histograms and the corresponding intensity modulated beams for plans optimized with various sized grids of radial basis functions. Also shown are the baseline IMRT results.

maps. When going from a high dilation 3 x 3 RBF grid to a 7 x 7 narrower dilation RBF grid, there is a slight improvement in target coverage and a decrease to the overall dose to the normal tissues. The beams become noticeably more modulated as the number of RBFs increases, but they are still much smoother than the baseline beamlet solution. There are also fewer areas of isolated high intensity in the field, although there are large areas of high intensity in beam 1.

As mentioned previously, the dilation of the RBFs or a in Equation III-8 was iterated through manually to determine a good value instead of making it an optimization parameter. This iteration took a minimum amount of extra time, and the final dilations

were analyzed to determine whether or not a trend could be found. Figure III-7 shows the four data points and two possible fits to the data. A 2nd order polynomial fit has an R² value of 1 and a simple power function fit has an R² value of nearly 1. These excellent fits demonstrate that the dilation certainly follows a trend, but more cases would have to be analyzed to determine which fit is best-suited for the problem. It is likely that similar geometries would fall into similar fits for the dilation, making the iteration time even less. These results agree with our original hypothesis that the optimal a value will be a function of the number of RBFs which is tied to the grid spacing between each RBF center. Figure III-9 shows that the number of basis functions increase and the space between each RBF is reduced, the width of the RBFs become narrower.

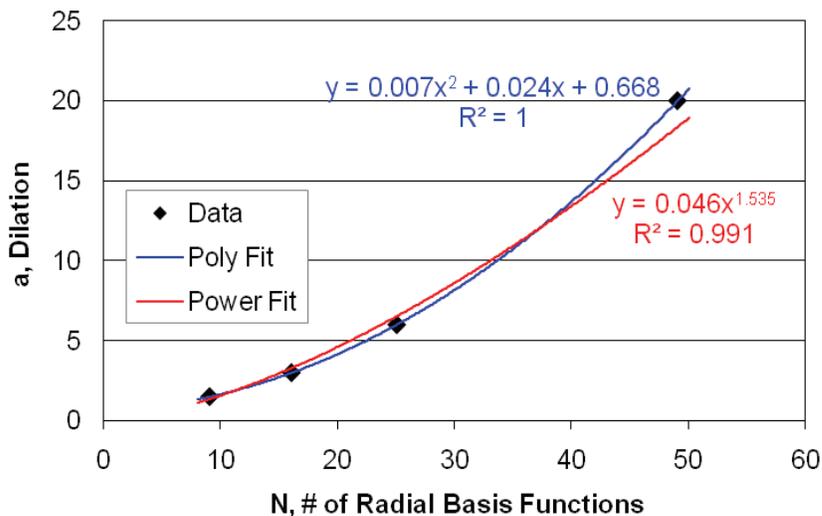


Figure III-9. The fitted dilation (a) for each radial basis function set plotted as a function of the number of RBFs in the set. The data were successfully fitted to a power function (shown in inset).

Both the polynomials and radial basis functions were successful at creating good quality IMRT plans in the phantom with many fewer optimization variables. The RBFs have a greater ability to achieve modulation where necessary, and have an advantage in sparing dose to the OAR1. They also appear to have slightly better target coverage with fewer variables. On the other hand, Figure III-8 shows that the polynomial functions required fewer monitor units for similar numbers of variables and cost. Figure III-8 shows that the MU requirements with increasing orders of polynomials do not necessarily increase. The MU requirements when using more populated RBF grids

shows a noticeable increasing trend. This is expected since the more RBFs that are used, the narrower the width of each function becomes, allowing for more localized

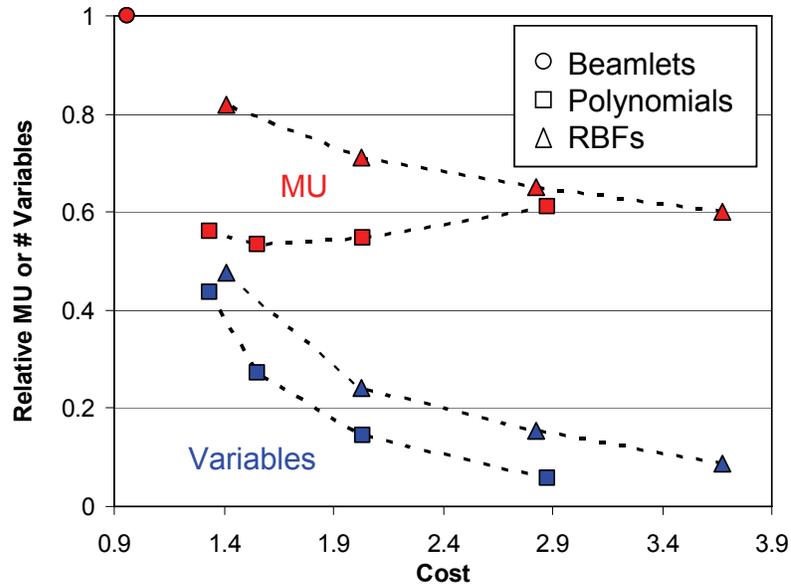


Figure III-10. The relative number of MU and optimization variables for each of the optimized cases as a function of the cost. The absolute number of total MU and variables in the baseline plan were 309, and 19,399, respectively.

changes in the modulation. As shown in Figure III-8, this feature allows for steeper gradients between the targets and normal tissues as compared to the polynomial plans suggesting that the RBFs may be more able to spare the organs-at-risk. However, large intensity gradients in the beam intensity also require additional monitor units.

For further comparison of the two methods, the minimum, mean, and maximum doses to the PTV, and the mean and maximum doses to OAR1, OAR2, and the remaining normal tissue is shown in Figure III-11. The minimum and maximum doses are reported as the minimum and maximum doses to at least 0.5 % of the volume. The results given here support the observations in the DVHs, which suggested that the biggest difference between the baseline and the basis function plans is in the normal tissue doses. While the mean dose in OAR1 is less with the baseline plan, each of the basis function plans kept the mean dose at or below the 20 Gy objective. However, only the higher-order polynomials and the two largest RBF grids were able to keep the OAR2 maximum dose below 35 Gy. The baseline beamlet plan was able to keep OAR2 well below this maximum dose. In addition, the mean dose to all of the normal structures as well as the PTV was higher in the basis function plans. It can also be noted that the

RBF 7x7 plan performed the best compared to the other basis function plans in terms of minimizing dose to normal tissues. All plans performed well in terms of target coverage; this is likely due to the design of the cost function.

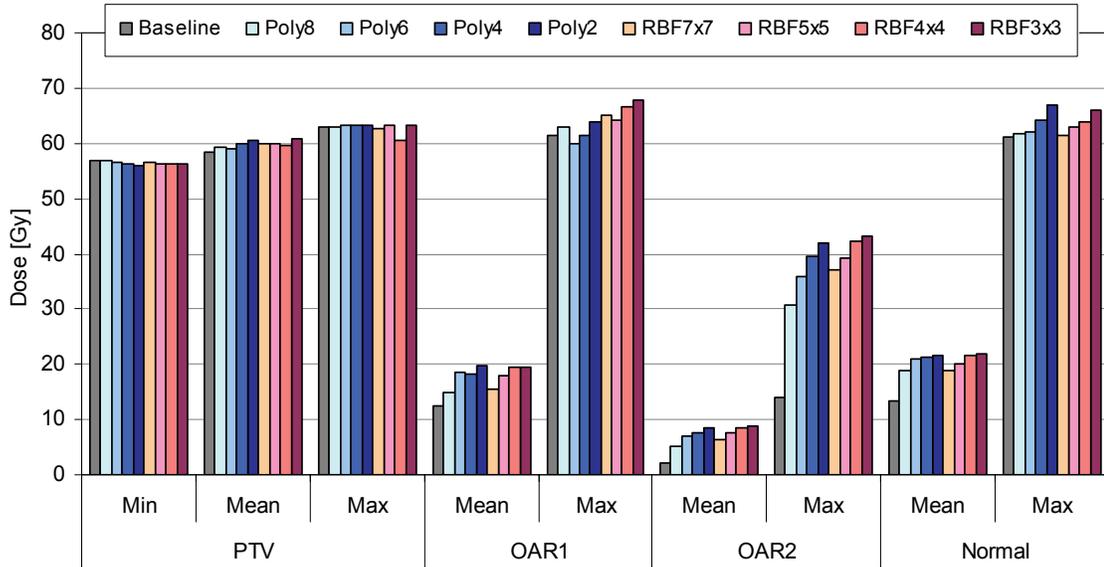


Figure III-11. Dose metrics for all optimization trials for the PTV, OAR1, OAR2, and normal tissue.

III.C.2. Clinical Examples

III.C.2.1. Brain

The brain case was optimized with each of the optimization methods – Baseline 2nd, 4th, 6th, and 8th order polynomials, and 3x3, 4x4, 5x5, and 7x7 RBF grids – and all plans were found to be clinically acceptable in the sense that all met the explicit inverse plan objectives. However, significant changes were observed in the overall dose received by the normal tissues. Figure III-12 shows the DVHs for the baseline, two polynomial, and two RBF optimizations. The 4th order polynomial and the 4x4 RBF cases had 75 and 80 variables, respectively, and the 8th order polynomial and 7x7 RBF plans had 225 and 245 variables each. The brainstem and normal tissue DVHs are shown on the right due to overlap with the right optic nerve and optic chiasm DVH on the left. Target coverage appears to be nearly identical for all optimization methods, but the minimization of dose to the brainstem, and chiasm is consistently better with radial basis functions than

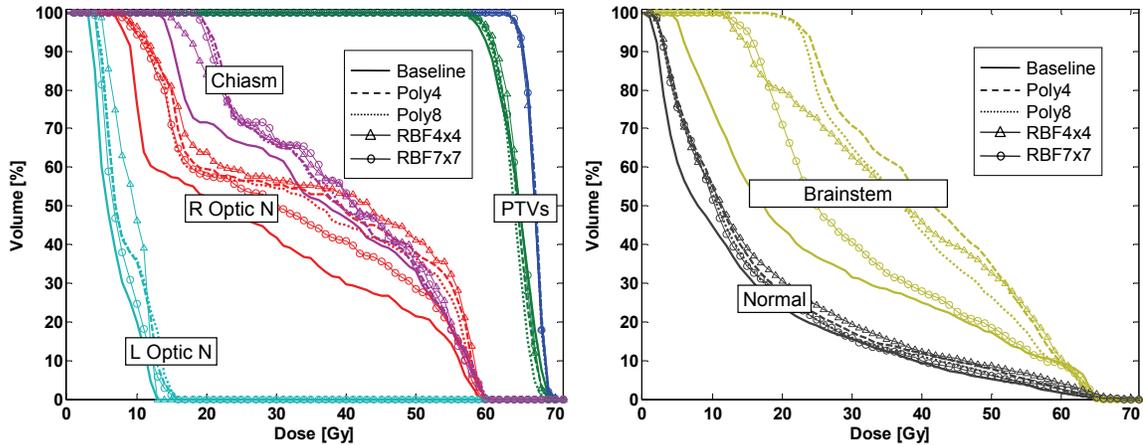


Figure III-12. Dose-volume histograms for the baseline beamlet plan in the brain in addition to plan optimized with 2nd and 4th order polynomial functions and 4 x 4 and 7 x 7 radial basis function grids.

polynomials. This is due to the increased resolution possible when optimizing with more localized functions such as the RBFs.

The standard deviation in several relevant dose metrics between all of the optimized plans for each of the structures are shown in Figures III-13. This shows that the biggest difference between the different optimization methods lies primarily in the normal tissue doses; this is partly a consequence of the design of the objective function. A closer look at the metrics with standard deviations equal to or greater than one can be

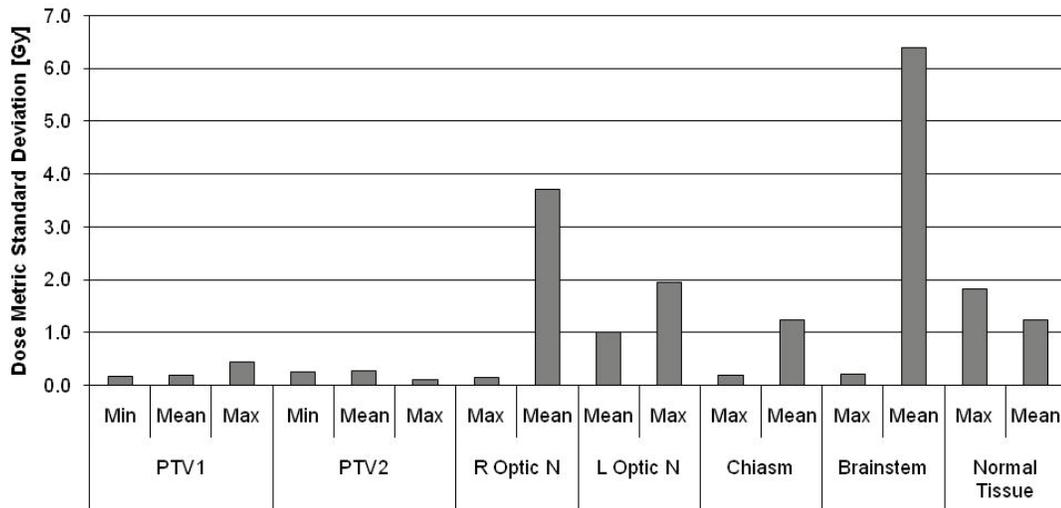


Figure III-13. The standard deviation in several relevant dose metrics between all optimization trials including baseline, 2nd, 4th, 6th, and 8th order polynomials, and 3x3, 4x4, 5x5, and 7x7 gaussian radial basis function grids. More details on the metrics with standard deviations equal to or greater than 1 Gy are shown in Figure III-14.

seen in Figure III-14. The large standard deviations observed in the RON and brainstem mean doses come from the differences between the basis function plans and baseline

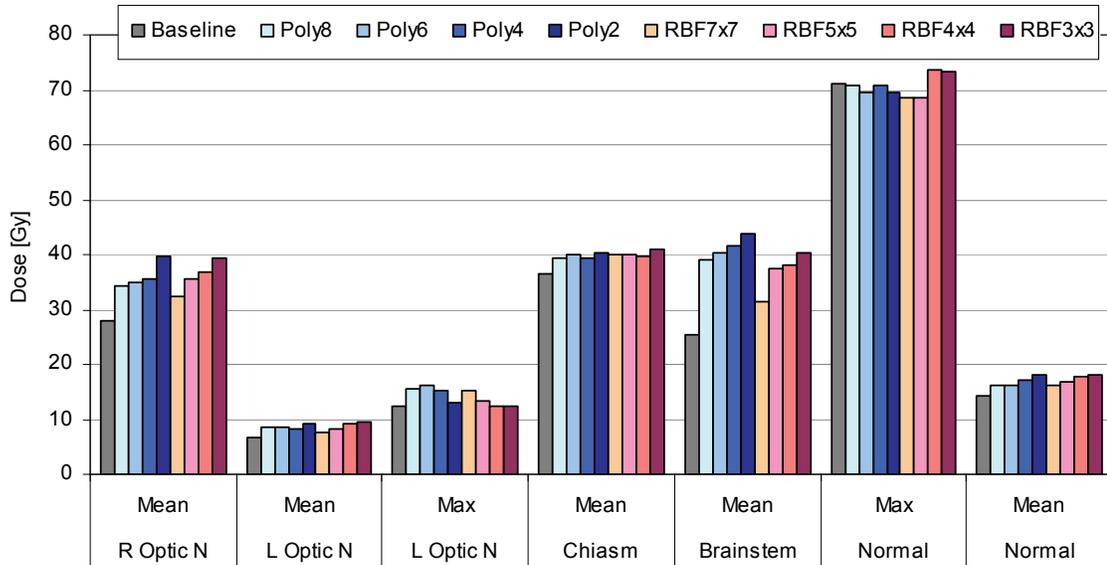


Figure III-14. Dose metrics for all optimization trials for the structures with large dose variations between methods.

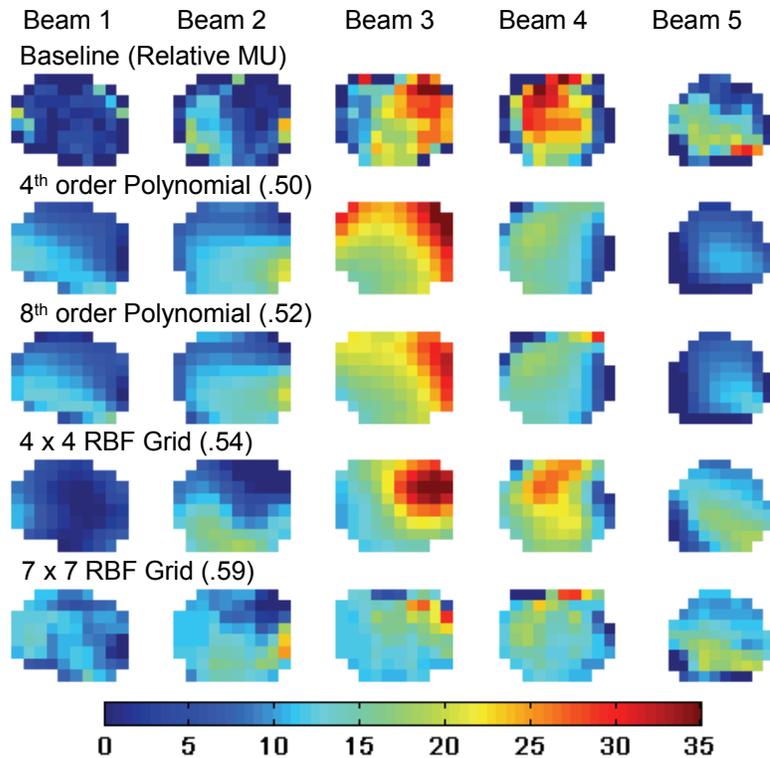


Figure III-15. Intensity modulated beams for the brain example. Baseline, 4th and 8th order polynomials, and 4x4 and 7x7 radial basis functions are shown.

plans, although the RBF 7x7 plans perform the best compared to the rest of the basis function plans.

The optimized intensity modulated beams are shown in Figure III-15 for each of the methods. It is clear that the beamlet plans exhibit the most modulation and “noise” in the optimized intensity patterns. As a result, they also require more monitor units to deliver. The relative number of MU to deliver the plan, and the number of optimization variables are shown relative to the baseline beamlet plan in Figure III-16. While the RBF plans exhibit superior normal tissue sparing over polynomials, they also result in more modulated fields that require more MU to deliver. The 7x7 RBF beams have a much higher degree of modulation than the rest of the basis function plans, but they are still much smoother than the baseline solution.

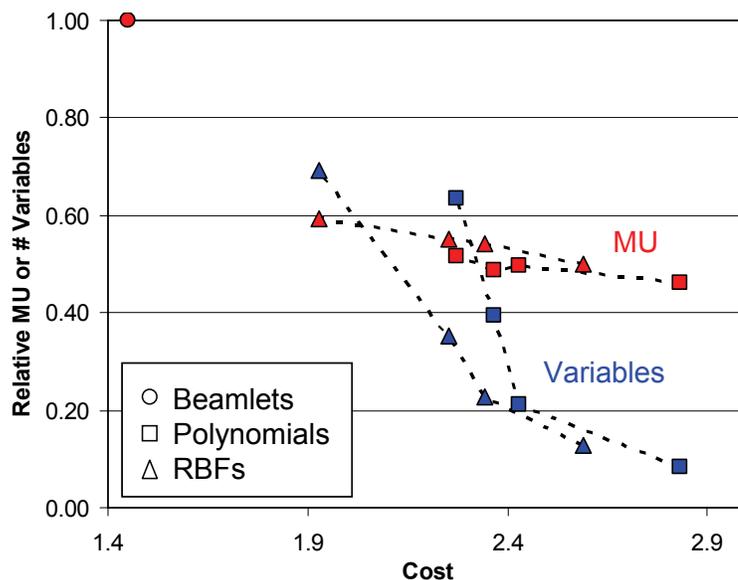


Figure III-16. The relative number of MU and optimization variables for each of the optimized brain plans. The absolute number of total MU and variables in the baseline plan were 354 and 26740, respectively.

III.C.2.2. Prostate

The prostate example was optimized successfully with both polynomials and radial basis functions. However, as in the brain, there were significant increases in mean dose to several of the normal tissues. Figure III-17 shows the DVHs for the baseline beamlet plan along with Poly4, Poly8, RBF4x4, and RBF7x7 plans. The target coverage is slightly degraded in the polynomial plans, although a 3% volume underdosing is allowed in the objectives. Again, the 7x7 RBF plan outperformed the remaining basis function

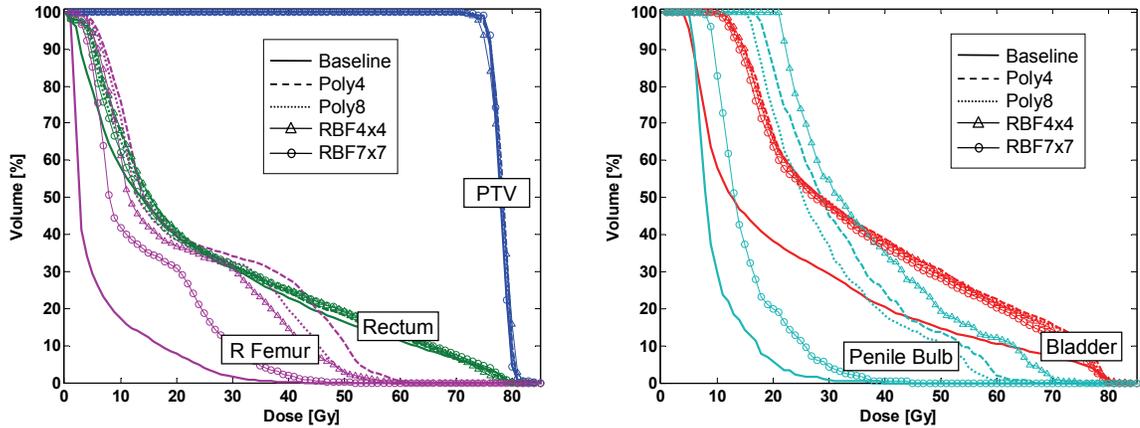


Figure III-17. Dose-volume histograms for the baseline beamlet plan in the prostate example in addition to plan optimized with 4th and 8th order polynomial functions and 4 x 4 and 7 x 7 radial basis function grids.

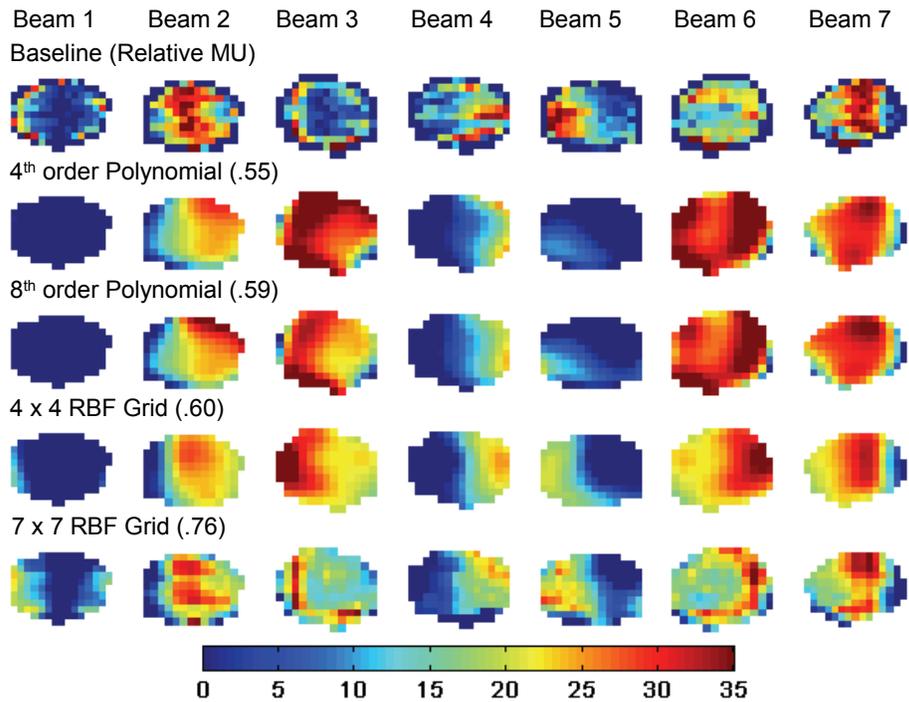


Figure III-18. Intensity modulated beams for the prostate example. Baseline, 4th and 8th order polynomials, and 4x4 and 7x7 radial basis functions are shown along with the relative MU requirements.

plans, due to its ability to create more localized modulation areas where needed. This helped reduce dose to the small volume penile bulb as well as sparing the femora (the left femur is not shown, but was similar to the right femur). The intensity modulated beams from each technique are shown in Figure III-18, along with the MU relative to

baseline. Again, we see little change with increasing polynomial order, but large differences as the number of variables in the RBF plan is increased. The trend in relative MU as the number variables increase can be seen in Figure III-19. Although the RBF7x7 has significantly more MU than the other cases, it represents a 24% reduction in MU compared to baseline, and the dosimetric results are superior to the other basis function plans.

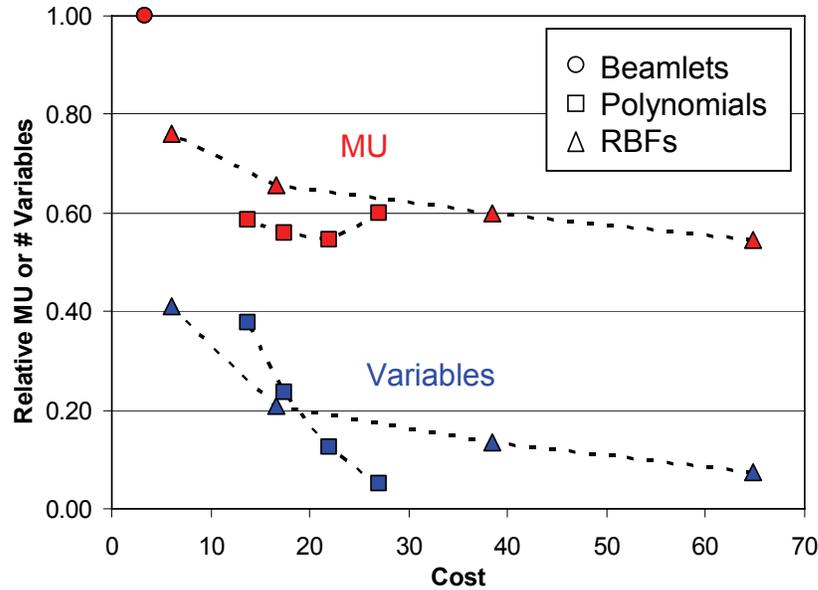


Figure III-19. The relative number of MU and optimization variables for each of the optimized prostate plans. The absolute number of total MU and variables in the baseline plan were 837, and 41,714, respectively.

III.C.2.3. Head/neck

Due to the complexity of the head/neck case and the previous results in the phantom and prostate, only the highest order polynomial and largest grid of radial basis functions were tested in this example. Both of the basis function plans had difficulties meeting the target objectives while sparing the numerous normal structures defined in the head/neck. Dose-volume histograms and intensity modulated beams for each of the optimization trials are shown in Figures III-20 and III-21. The increase in normal tissue dose in the basis function plans relative to the baseline plan is noticeable. The RBF plan is slightly better than the polynomial plan, although both are significantly worse than the baseline beamlet plan.

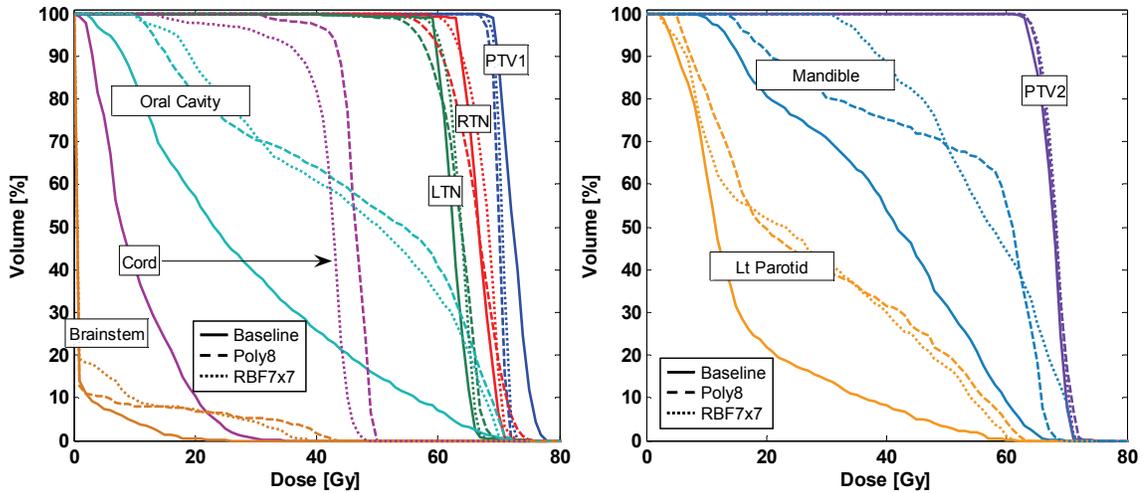


Figure III-20. Dose-volume histograms for the baseline beamlet plan in the head/neck example in addition to plan optimized with 8th order polynomial functions and a 7 x 7 radial basis function grid.

One of the most striking differences when using basis functions instead of beamlets is the increase in dose to the spinal cord. While the cord remains below its maximum dose limit of 50 Gy, the overall dose is increased by a large amount when using basis function optimization. In fact, most of the higher priority objectives were met by the basis function plans, leaving the biggest tradeoffs to occur with the normal tissue dose minimization costlets.

Figure III-21 reveals that the beamlet plan is very modulated, with several high intensity spikes and a large number of smaller fluctuations between neighboring beamlets. The polynomial function does a poor job approximating the general shape of the baseline plan, while the 7x7 RBF beams appear to be more intuitive, realizing that the areas of low intensity in the field correspond to the spinal cord location. Higher resolution RBF grids may be successful in the head/neck geometry, although there may still be issues with the beam edges. It is important to note that the basis function plans have more than an 80 % reduction in the number of variables compared to the baseline beamlet solution. Due to software constraints, the size of the RBF grid was limited, but further work to remove this constraint would be warranted to further study basis function optimization, especially in the RBF case.

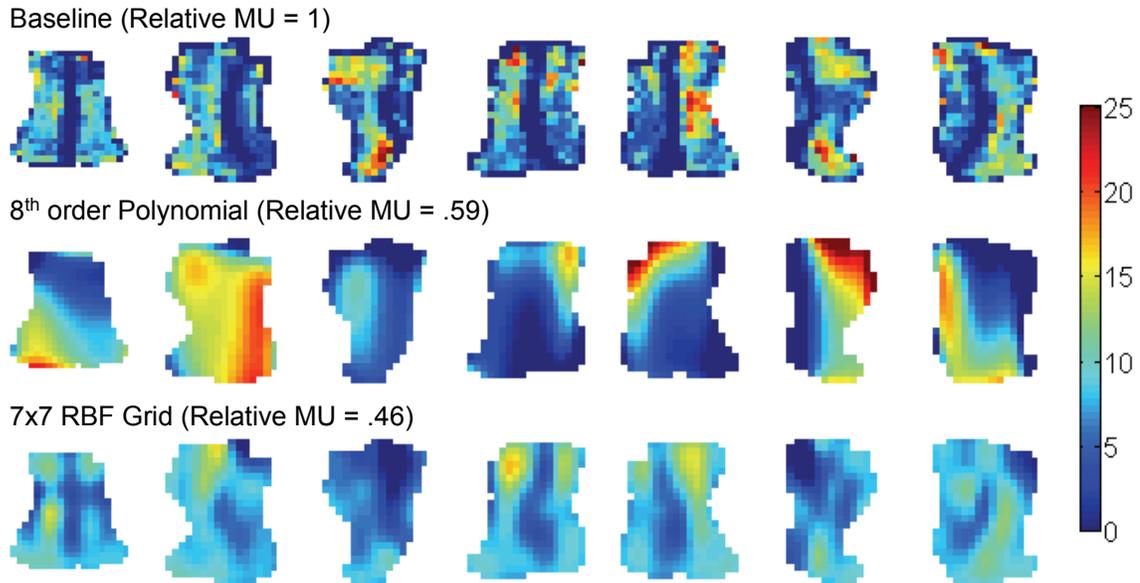


Figure III-21. Intensity modulated beams for the head/neck example. Baseline, 8th order polynomials, and 7x7 radial basis functions are shown along with the relative MU requirements.

III.D. Discussion

We have implemented and tested the use of mathematical basis functions to represent and optimize intensity patterns in inverse planned intensity modulated radiation therapy. We have shown results from the implementation of two basis function sets: polynomials and Gaussian radial basis function grids. In polynomial optimization, the leading coefficients in a polynomial function are optimized, and the beamlet values are assigned based on their spatial location according to the polynomial equation. In the Gaussian RBF implementation, an evenly-spaced grid of RBFs of equal dilation is initialized and the weights, or amplitudes of the individual RBFs are optimized such that the composite RBF set represents the beam. Basis function optimization with these two formulations was found to be feasible for 3D inverse IMRT planning, and is capable of producing acceptable IMRT plans in simple geometries with large reductions in beam modulation and MU. In the cases studied here, 25 – 50 % increases in delivery efficiency were seen when using basis function optimization. There is also a substantial decrease in the number of optimization variables as well as the time required per optimization (although this was not quantified).

Our implementation was characterized in a test phantom, and then studies were performed on several clinical examples. In the test case, both polynomial and radial

basis function optimization plans were found to be clinically comparable to beamlet plans and could be delivered with 18 - 44 % fewer monitor units. In the brain and prostate, a 7x7 grid of Gaussian radial basis function was the most successful at sparing normal tissues. The use of the basis function sets studied here was not adequate to plan a complicated head/neck IMRT case. However, the 7x7 RBF was able to meet most of the planning objectives with 80% fewer optimization variables. We believe that the use of higher resolution RBF grids is warranted for future study in head/neck and other more complicated geometries.

During this study, we identified several practical limitations of basis function optimization. First, the basis function sets described here usually sacrifice normal tissue sparing in one or more normal structures. The smaller the organ-at-risk, the more localized the modulation must be to adequately spare the structure. Because of this, only the 7x7 RBF plans were able to spare certain normal structures, such as the penile bulb in the prostate example. Even in this case, the beamlet plan was still able to spare the penile bulb to a much higher degree than the radial basis function plan. Another limitation of basis function optimization is the lack of resolution (high intensity gradients) at the edges of the IMRT fields. Functions such as polynomials are somewhat global, making it difficult to approximate a sharp gradient at the edge of the beam. This is responsible for some of the extra normal tissue dose observed in the basis function plans compared to the baseline beamlet plans. We have experimented with different edge modification functions to remedy this situation, but these efforts have been largely unsuccessful due to the fact that the composite function must stay positive to achieve good convergence with gradient-based optimization methods. The most practical way to ensure resolution at the edges of the field has been to simply exclude any beamlets that are more than 2-3 mm away from the edge of the target. If larger margins are defined in the beam than necessary, we developed a method to essentially tighten the margin and turn off the beamlets near the edge of the field. This is done by the “Edges Off” option in the basis function user interface shown in Figure III-5. If the option is selected then all beamlets whose centers are greater than a distance given by “EdgeWt” away from the target are set to zero. The use of this option can potentially affect the cost function gradient calculation with respect to the coefficient, but we have not observed a problem with convergence due to this.

Another practical limitation that was encountered during this work was a limit on the number of coefficients that could be used in the software. However, programming

techniques could overcome this limitation and is a good starting point for future work on this method. The more basis functions and coefficients that are used, the more important it becomes to be able to use a gradient descent optimization algorithm. The use of the simulated annealing algorithm to fully optimize the polynomial basis functions was not too costly because of the small number of coefficients to optimize. However, if more coefficients are used, simulated annealing will become too time-consuming to use this method clinically.

We have shown that basis function optimization is a promising method to reduce the degrees of freedom necessary for IMRT, resulting in reduced complexity in IMRT fields and reduced MU (and thus reduced leakage/transmission dose). We have shown that this method is able to be implemented within an existing optimization framework and also able to produce acceptable IMRT plans. Further research in this area, to locate improved basis function sets that add up to positive-valued beamlets and allow for more variables in the optimization software, is worthwhile. This method could be useful in its current state for simple geometries such as partial breast irradiation. Our experience with this method in optimization of very complex cases, such as head/neck, suggests that a significant increase in the number of basis functions would be necessary to equal the quality of beamlet optimization. Therefore, the high degrees of freedom of beamlet optimization may actually be necessary for a high-quality plan in many complex geometries.

Since we know that these cases can be prone to degeneracy, a better way to find a more delivery-efficient solution in the search space must be identified. In the next chapter, we return to the beamlet paradigm to study the use of both beamlet restrictions and modulation penalties during the optimization process.

A.III. Appendix for Chapter III

A.III.1. Alternate Basis Function Types

During the development of the basis function optimization for 3D inverse planning of IMRT, several basis function sets were implemented and evaluated preliminarily for use. The following is a summary of the basis function sets that are implemented in UMOpt, but were not discussed in the chapter contents.

A.III.1.2. Basis Type 2: Radial/Theta based edge function

Basis type 2 represents a radial based function that was implemented to deal with the problem of edge resolution. The formulation is very complex and was difficult to implement due to the data about the beamlets that is available in the optimization system. This basis type represents the beamlet intensity at (x,y) by

$$\begin{aligned}
 b(x,y) &= P(x,y)R(r,\theta) \\
 P(x,y) &= c_0 + c_1x + c_2y \\
 R(r,\theta) &= 1 - \exp\{S(\theta)[r - F(\theta)r_{\max}]\} \\
 S(\theta) &= c_3 \left\{ \sin\left[c_4(\theta - c_5)\right] + 1 \right\} \\
 F(\theta) &= c_6 \left\{ \sin\left[c_7(\theta - c_8)\right] + c_9 \right\}
 \end{aligned} \tag{A.III.1}$$

Each of the terms shown above were designed to represent a certain aspect of the beam shape. These representations are given in Table A-III-1, and the description of the coefficient variables in A.III.1 are given in Table A-III-2. The coefficient types are used to tell the optimizer what the limits of the coefficient values are.

Table A-III-1 The physical representation of each of the component of (A.III.1)

Component	Representation
P	General intensity with planar tilt
R	Edge fall off
S	Steepness of the fall off as a function of theta
F	Beginning fall off distance as a function of theta

Table A-III-2. Description of the coefficient properties in for basis type 2.

Coeff	Representation	Type	Min	Max
0	Mean Intensity Level	1	-Max Intensity	+ Max Intensity
1	X Wedge	1	- Max Intensity	+ Max Intensity
2	Y Wedge	1	- Max Intensity	+ Max Intensity
3	sin height	2	0	+ Max Intensity
4	sin period	3	-10	+ 10
5	sin horizontal shift	4	-pi	+pi
6	sin height	2	0	+ Max Intensity
7	sin period	3	-10	+ 10
8	sin horizontal shift	4	-pi	+pi
9	sin vertical shift	2	0	+ Max Intensity

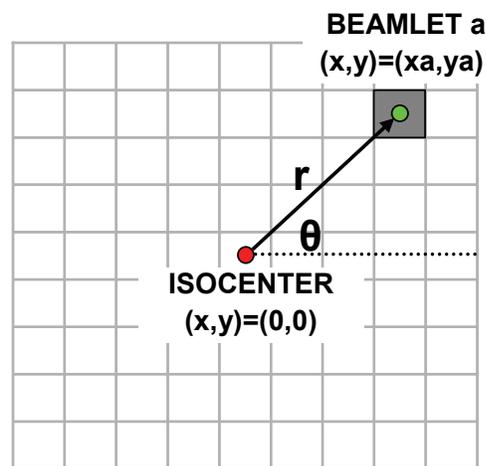


Figure A-III-1. An illustration of the notation and variables for r and θ calculated for the use of basis function type 2.

The above implementation is theoretically sound despite the difficult implementation. The parameters could benefit from further refinement and study although this basis type was not found to have an advantage over the more thoroughly studied radial basis functions or polynomials.

A.III.1.3. Basis Types 3 and 4: Mexican Hat Wavelets

Basis type 3 is a weighted sum of three different Mexican hat wavelets where the coefficients are the weights, and x and y dilations and translations of the function. An illustration of a Mexican Hat wavelet is shown in Figure A-III-2.

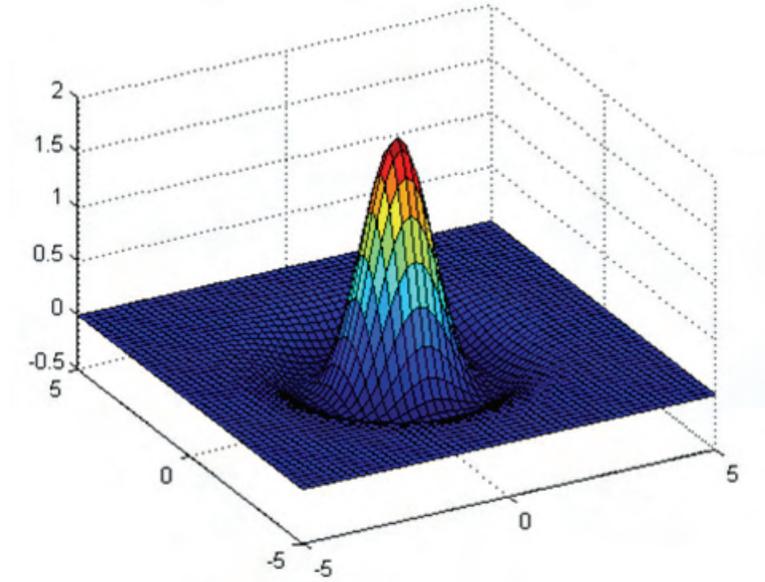


Figure A-III-2. A plot of a 2D Mexican hat wavelet.

The beamlet intensity at (x,y) is given by:

$$b(x,y) = \sum_{i=1}^N w_i \left[\left(2 - \left[\left(\frac{x-t_{x,i}}{a_{x,i}} \right)^2 + \left(\frac{y-t_{y,i}}{a_{y,i}} \right)^2 \right] \right) \exp \left(-\frac{1}{2} \left[\left(\frac{x-t_{x,i}}{a_{x,i}} \right)^2 + \left(\frac{y-t_{y,i}}{a_{y,i}} \right)^2 \right] \right) \right] \quad (\text{A.III.2})$$

where N is the number of Mexican Hat Bases, w_i is the weighting of the i -th base function, t_{xi} and t_{yi} , are the x and y translations of the i -th base, and a_{xi} and a_{yi} , are the x and y dilations of the i -th base. In terms of the coefficients, we can write:

$$b(x,y) = \sum_{i=1}^N c_{0i} \left[\left(2 - \left[\left(\frac{x-c_{1x,i}}{c_{2x,i}} \right)^2 + \left(\frac{y-c_{3y,i}}{c_{4y,i}} \right)^2 \right] \right) \exp \left(-\frac{1}{2} \left[\left(\frac{x-c_{1x,i}}{c_{2x,i}} \right)^2 + \left(\frac{y-c_{3y,i}}{c_{4y,i}} \right)^2 \right] \right) \right] \quad (\text{A.III.3})$$

where the coeffs are explained in Table A-III-3. This particular basis type did not perform as well as expected, but there may not have been enough coefficients to create the beams necessary.

Table A-III-3. Coeff descriptions for basis type 3.

Coeff (i=0..N)	Representation	Type	Min	Max
5i	Base Weight	1	-Max Intensity	+ Max Intensity
5i+1	X translation	3	- 10	+ 10
5i+2	X dilation	2	0	+ Max Intensity
5i+3	Y translation	3	-10	+ 10
5i+4	Y dilation	2	0	+ Max Intensity

Basis type 4 is a weighted of 15 pre-defined wavelets where the coeffs are the weights of the wavelets in the library. The format is the same as in A.III.3, but the coefficients are the weights and the other parameters are fixed for 15 pre-defined wavelets. Figure A-III-3 shows an example library of wavelets for basis type 4. This method could be useful if a large number of cases were tested to design a desirable library. It may be possible to apply a neural network or learning algorithm to choose the libraries that could work for different geometries.

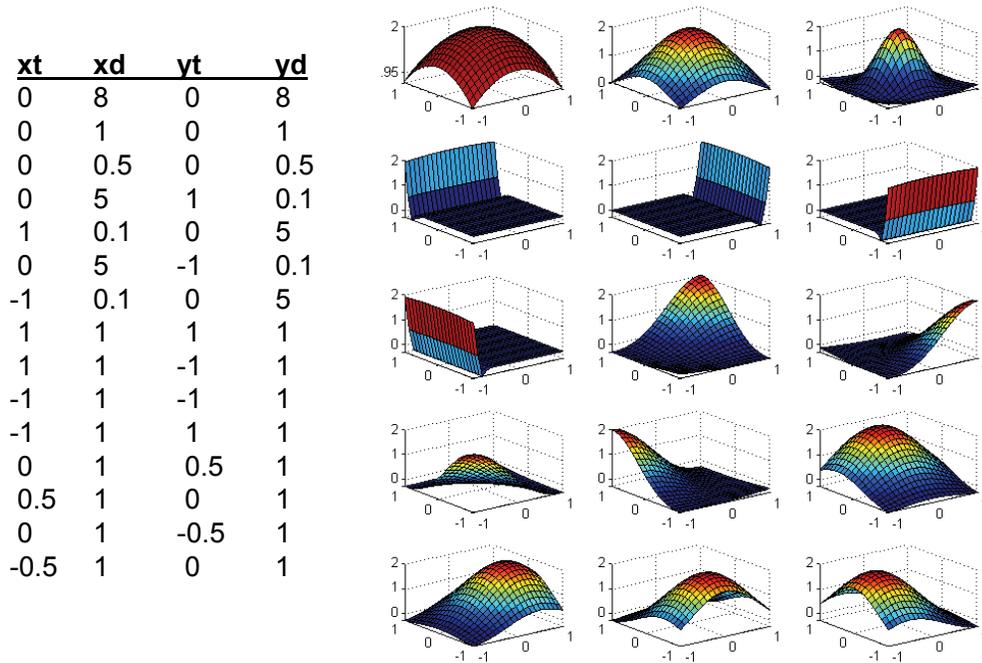


Figure A-III-3. A sample basis function library of Mexican hat wavelets for basis type 4.

A.III.1.4. Basis Type 9: Combined Polynomials + Radial Basis Functions

Basis type 9 is a combination of a 2nd order polynomial with a grid of small dilation radial basis functions. It was designed to achieve a beam's global shape through the polynomial terms, and also be able to represent local fluctuations with small dilation radial basis function. Basis type 9 currently uses a 7 x 7 RBF grid on top of the 2nd order polynomial. The beamlet intensity at (x,y) is calculated by

$$b(x,y) = c_0 + c_1x + c_2y + c_3x^2 + c_4y^2 + c_5xy + \sum_{i=6}^{55} c_i G_i(x,y) \quad (\text{A.III.4})$$

where the final term represents the 7 x 7 RBF grid, which is the same as the form in Eqn. (III.8) in the main chapter. Results from this basis type have been largely similar to the result from the Gaussian RBF grid alone, but more careful placement of the RBFs over the polynomial surface may improve the performance and allow for similar dosimetric results with fewer variables.

A.III.2. Optimization Code Details

The addition of the basis function capabilities into the existing UMOpt infrastructure required several code alterations, including the introduction of several new variables and

Table A-III-4. Description of basis function optimization-specific variables.

Structure	Variable Name(s)	Description
UMOPTIM_OPT	BASIS_FCN_OPT_ON	Allows user to choose basis function or beamlet optimization
	NBF	Number of basis functions (used with the library)
	MAX_B_GRADIENT	The largest beamlet gradient. Used in add_parameter routine
PARAM_OPT	COEFF_INDEX	Used so each parameter knows what coefficient type it is representing
BEAMLET_OPT	XC,YC,IXC,IYC,R,THETA	The coordinates of the beamlets (ixc, iyc, r, and theta can be normalized)
	GRADIENT	The beamlet gradients (used in q-newton method)
COEFF_T	CINDEX	Index for the coefficient transpose matrix
	CVALUE	The value of the coeff transpose matrix element (q-newton method)
BEAM_OPT	BASIS_TYPE	The type of basis function in use
	NCOEFF	The number of coeffs being used per beam for the given basis function type
	MAX_XC,MIN_XC, MAX_YC,MIN_YC,R_MAX	The max and min beamlet coordinates for each beam
BFLIB_OPT	XT,XD,YT,YD	The x and y translations and dilations of basis function library members
COEFF_OPT	ENABLED	Whether the coefficient is enabled or not
	WEIGHT	The weight of the coefficient
	COEFF_TYPE	The type of coefficient. Used for optimization limits and step sizes.

the modification of a few existing subroutines. Each new variable is briefly described in Table A-III-4. Additions of the variables and structures had to be done in both `[..include]umoptim_opt.h` and `[...include]umoptim_opt_struct.txt`. AVS variables and options are defined in `[...qnewton]anneal.c` and must be listed in `[...include]anneal.h`. Subroutines inside `anneal.c` that have been altered include `create_parameters`, `adjust_param`, and `update_parameters`. Slight modifications have also been made inside the main optimization method codes, `qnewton_method`, and `anneal_method`, and several other routines including `sum_beams`. Each of these routines now incorporates the use of basis function coefficients as parameters and allows the beamlet values to be updated based on the values of the basis function coefficients. There had to be slight changes in the anneal and quasi-newton routines as well as in the way the dose is summed.

References

- ¹ R. Mohan, M. Arnfield, S. Tong, Q. Wu, and J. Siebers, "The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy," *Med. Phys.* **27**, 1226-1237 (2000).
- ² S. Spirou, N. Fournier-Bidoz, J. Yang, C. Chui, and C. Ling, "Smoothing intensity-modulated beam profiles to improve the efficiency of delivery," *Med. Phys.* **28**, 2105-2112 (2001).
- ³ X. Sun and P. Xia, "A new smoothing procedure to reduce delivery segments for static MLC-based IMRT planning," *Med. Phys.* **43**, 2785-2794 (2004).
- ⁴ M. Alber and F. Nusslin, "Intensity modulated photon beams subject to a minimal surface smoothing constraint," *Phys. Med. Biol.* **45**, N49-N52 (2000).
- ⁵ Q. Hou, C. Zhang, Z. Wu, and Y. Chen, "A method to improve spatial resolution and smoothness of intensity profiles in IMRT treatment planning," *Med. Phys.* **31**, 1339-1347 (2004).
- ⁶ L. Ma, "Smoothing intensity-modulated treatment delivery under hardware constraints," *Med. Phys.* **29**, 2937-2295 (2002).
- ⁷ M. M. Coselmon, J. M. Moran, J. Radawski, and B. A. Fraass, "Improving IMRT delivery efficiency by applying intensity limits during inverse planning," *Med. Phys.* **32**, 1234-1245 (2005).
- ⁸ J. Markman, D. A. Low, A. W. Beavis, and J. O. Deasy, "Beyond bixels: Generalizing the optimization parameters for intensity modulated radiation therapy," *Med. Phys.* **29**, 2298-2304 (2002).
- ⁹ M. L. Kessler, D. L. McShan, M. Epelman, K. A. Vineberg, A. Eisbruch, T. S. Lawrence, and B. A. Fraass, "Costlets: a generalized approach to cost functions for automated optimization of IMRT treatment plans," *Optimization and Engineering* **6**, 421-448 (2005).
- ¹⁰ G. Zhang, Z. Jiang, D. Shepard, M. Earl, and C. Yu, "Effect of beamlet step-size on IMRT plan quality," *Med Phys* **32**, 3448-3454 (2005).
- ¹¹ T. R. Mackie, J. W. Scrimger, and J. J. Battista, "A convolution method of calculating dose for 15-MV x rays," *Med. Phys.* **12**, 188-196 (1985).
- ¹² T. R. Bortfeld, D. L. Kahler, T. J. Waldron, and A. L. Boyer, "X-ray field compensation with multileaf collimators," *Int J Radiat Oncol Biol Phys* **28**, 723-730 (1994).

CHAPTER IV.

RESTRICTING OR PENALIZING BEAM MODULATION DURING INVERSE PLANNING FOR IMRT

IV.A. Motivation

In Chapter III, we analyzed the merits of basis function optimization as a way to improve the continuity and smoothness of IMRT plans by representing beam intensities by surfaces instead of individual beamlets. This reduces the degrees of freedom available in the optimization problem and is a very promising technique to optimize simple IMRT plans, such as those required to treat prostate cancer. However, in more complicated treatment sites that require more sophisticated plans, such as the head and neck, it can be difficult to find a robust function or set of functions that can be optimized to represent the intensity pattern complexity that is necessary. An alternative option is to return to the beamlet paradigm and search for other ways to reduce beam complexity.

One possible strategy is to investigate whether some of the high intensity peaks that frequently occur in IMRT beams may result from limitations in the inverse planning optimization strategy and might not be essential for high quality plans. If so, then constraining the maximum allowable intensity for an IMRT plan may inhibit the optimization engine from pursuing an undesirable path that may be an artifact of the point-based inverse planning approach. Using a maximum intensity limit for the plan in the optimization process may result in beamlet patterns that lack potentially-unnecessary modulation and sharp spikes, while still allowing the optimization algorithm to make the proper tradeoffs between target and normal tissue doses. This approach would ideally produce a plan that will be (i) sequenced more accurately, (ii) delivered with fewer MUs, and (iii) less sensitive to positioning errors than an IMRT plan derived without maximum intensity limits, while still achieving the defined clinical objectives. Thus, maximum intensity limits can partly reduce the consequences of the beam complexity problem by removing large intensity spikes, but they do not prevent smaller fluctuations in

modulation between neighboring beamlets that can also increase MU. In addition, if a plan is very complex, or the maximum intensity limits are too strict, the use of these types of beamlet restrictions could potentially prevent the optimizer from reaching a quality IMRT plan. Therefore, it may be necessary to use “softer” restrictions such as smoothing the IMRT beams or penalizing modulation as a part of the cost function.

Smoothing procedures can either be applied inside or outside the optimization loop. Smoothing IMRT beams outside the optimization process may produce limited results or require re-optimization due to the plan degradation that occurs during post-optimization smoothing¹⁻³. Inside the optimization loop, smoothing can be done by (i) including smoothness criteria inside the objective function, or (ii) smoothing beams after each cycle or iteration⁴. The latter method is easier to implement, but it suffers from the same difficulty as applying interventions post-optimization: when any procedure is applied outside the cost function, the impact of that procedure on the cost function cannot be weighed according to the dosimetric consequences.

Implemented carefully, methods that penalize modulation as a part of the cost function can effectively improve delivery efficiency, while taking into account the dosimetric tradeoffs to be made with IMRT plan objectives. Spirou *et al.* have presented a comparison of smoothing, using a Savitzky-Golay filter along the direction of the MLC leaf travel, inside and outside the cost function⁴. They found that smoothing inside the cost function was superior in terms of producing sharper dose gradients, better dose homogeneity, and better critical organ sparing, especially for more complex cases. However, again, care must be taken to design the objective function in such a way that the smoothing process or modulation penalty does not dominate the solution and interfere with normal tissue dose limits or target coverage⁵.

As the radiotherapy community gains better knowledge of IMRT, it has become clear that the value of an IMRT plan should not be judged solely on its DVHs and dose metrics, but also on the efficiency of delivery and other potential problems associated with the plan, such as increased time and effort needed for quality assurance. This chapter examines two techniques to reduce IMRT beam complexity which can be applied during the optimization process. The first technique is the application of beamlet intensity limits that are placed as hard restrictions during optimization. The second technique is the use of beam modulation penalties as part of the inverse planning cost function. We discuss our development and implementation of two such penalties. All of

the methods are studied to characterize and compare their abilities to preserve or improve IMRT plan quality regarding (i) target and normal tissue DVHs and dose metrics, and (ii) reduced modulation and improved delivery efficiency.

IV.B. Methods

As discussed in previous chapters, the in-house treatment planning and optimization systems (UMPlan/UMOpt)⁶⁻¹² at the University of Michigan were developed with an infrastructure that allows for continual upgrades and improvements as radiotherapy technology advances. UMOpt, especially, is a versatile software package designed for 3-D beamlet-based inverse IMRT optimization, which allows for many different optimization parameters and limits and can also make use of specialized cost function components, or “costlets”. Using UMOpt, we have studied the impact of placing maximum intensity limits on the beamlet intensities in IMRT. In addition, three new costlets that aim to reduce IMRT beam complexity when included in an inverse plan objective function have been developed and implemented in UMOpt for further study. The three costlets do the following: (i) minimize deviation from a 1-D Savitzky-Golay filtered beam, where filtering is done only in the direction of the MLC travel, (ii) minimize deviation from a 2-D Savitzky-Golay filtered beam, or (iii) minimize a measure of the total plan modulation called the quadratic plan intensity map variation (PIMV_q)¹³. The theory and implementation of each of these methods is shown below.

IV.B.1. Maximum Beamlet Intensity Restrictions

Restricting the maximum intensity of beamlet values in IMRT planning is a simple idea that could have great benefit in terms of delivery efficiency. High intensity peaks that significantly increase the number of monitor units required to deliver a beam frequently occur in IMRT. If some of these features are not truly necessary to meet the cost function objectives, then the application of intensity limits could greatly reduce MU with little dosimetric effect on the IMRT plan. To test this idea, plans can be optimized using a series of assigned values of the maximum beamlet intensity. Plans can also be optimized without using a maximum intensity limit to represent the unconstrained solution. To better compare the intensity variations seen with different plans, since the required maximum intensity per beamlet is technique- and dose prescription-dependent, the maximum intensity ratio was chosen as a metric to permit direct comparison of plans

with different prescriptions or number of beams. The ‘maximum intensity ratio’ or MIR, is defined as

$$\text{MIR} = \frac{I_{\max} N_b}{D_t}, \quad (\text{IV.1})$$

where I_{\max} is the maximum intensity allowed for each beamlet, N_b is the number of beams in the plan, and D_t is the prescribed dose to the target volume. I_{\max} is set for a plan so that no beamlet defined in the plan can exceed the maximum intensity. Note that the utility of the maximum intensity ratio for comparing plans from treatment sites with large geometrical differences will be limited.

At different maximum intensity ratios, it is necessary to evaluate the amount of modulation present in the field to determine whether the intensity restrictions are effective at reducing the overall amount of the modulation in the field. To achieve this, we have defined a metric called the intensity map variation (introduced in Chapter II), which can be defined for each field and used to measure the modulation across a beam. The plan intensity map variation (PIMV) is calculated by summing the variation for each field and is defined for each plan as:

$$\text{PIMV} = \sum_{n=1}^{N_b} \left(\sum_{j=1}^{J-1} \sum_{k=1}^{K-1} \left[|b_{jk} - b_{j,k+1}| + |b_{jk} - b_{j+1,k}| + \dots \right] \right), \quad (\text{IV.2})$$

where N_b is again the number of beams in the plan, J is the maximum number of beamlets in the direction parallel to the motion of the MLC, K is the maximum number of beamlets in the direction perpendicular to the motion of the MLC, and b_{jk} is the intensity of the beamlet at the (j,k) grid position. The PIMV was chosen to be a measurement of the field modulation that is not biased by the sequencing algorithm chosen. For this study, each set of beamlets was defined as a regular grid. However, if a beam is defined as segments or an irregular grid, it would be necessary to apply a grid based on the smallest beamlet dimension and then use Equation IV.2 to calculate the PIMV.

In addition to measuring the overall field modulation, it is also important to determine to what extent the maximum intensity limits affect the beamlet intensity

pattern relative to the baseline beam. For a measure of the similarity between two intensity maps for plans at different maximum intensity ratios, the correlation coefficient for each intensity grid with respect to the unconstrained optimized intensity grid can be calculated. The correlation coefficient (use in Chapter II as well) between beams A and B, is given by

$$C_{I_A/I_B} = \frac{\sum_j \sum_k (b_{A,jk} - \bar{b}_A)(b_{B,jk} - \bar{b}_B)}{\sqrt{\left(\sum_j \sum_k (b_{A,jk} - \bar{b}_A)^2\right)\left(\sum_j \sum_k (b_{B,jk} - \bar{b}_B)^2\right)}}, \quad (IV.3)$$

where j and k are the dimensions of the intensity map, $I_{X,jk}$ is the intensity of grid element (j,k) when the maximum intensity ratio is X , and \bar{b}_X is the mean intensity of grid b_X . The correlation coefficient may vary from -1.0 to 1.0. A value of 1.0 means that the two patterns are perfectly linearly and positively correlated, while a value of -1.0 means that the two patterns are perfectly linearly and oppositely correlated. A high absolute number means there is a high level of correlation, while a small absolute number represents a weak correlation.

IV.B.2. Modulation Penalties

In Chapter II, we discussed the fact that IMRT treatment planning is, in essence, a large-scale optimization problem with a large number of independent variables. Alber *et al.* have shown that the majority of beamlets in the optimization problem are not in critical regions, and therefore they may not have a large effect on meeting the goals outlined in the objective function. Thus, without intervention, these beamlet patterns may become very noisy. Another source of complexity shown in Chapter II was that the random point sampling of the regions of interest can cause fluctuations in the beamlet intensities. Both of these issues cause modulation to occur in the IMRT field with potentially minimal dosimetric benefit. Therefore, it makes sense to apply another objective function component that lightly penalizes modulation in an effort to reach a smoother solution that can be delivered more efficiently. This section discusses our implementation of two different types of modulation penalties.

IV.B.2.1. Savitzky-Golay Filter Penalty

When computers became more instrumental in data analysis, Savitzky and Golay introduced the idea using a moving least squares fitting method to smooth noisy data without introducing large distortions¹⁴. Least squares fitting is a common mathematical procedure for finding the curve of best-fit to a set of data by minimizing the sum of the squares of the offsets of the data from the curve (or surface). By applying this method over small subset of data in a convolution procedure, the local variations can be well-modeled while smoothing the noise. In modern filtering processes, a Savitzky-Golay filter usually refers to polynomial-based least squares fitting. Two-dimensional Savitzky-Golay filters have been applied to smooth image and other 2D data sets in a variety of math, physics, and engineering applications. One dimensional Savitzky-Golay filtering has also been previously applied to smoothing IMRT beams by Spirou *et al.*⁴ Their work used moving least squares fitting with a 2nd degree polynomial to 5 element beamlet vectors in the direction of the MLC motion. An advantage in dosimetric quality and the

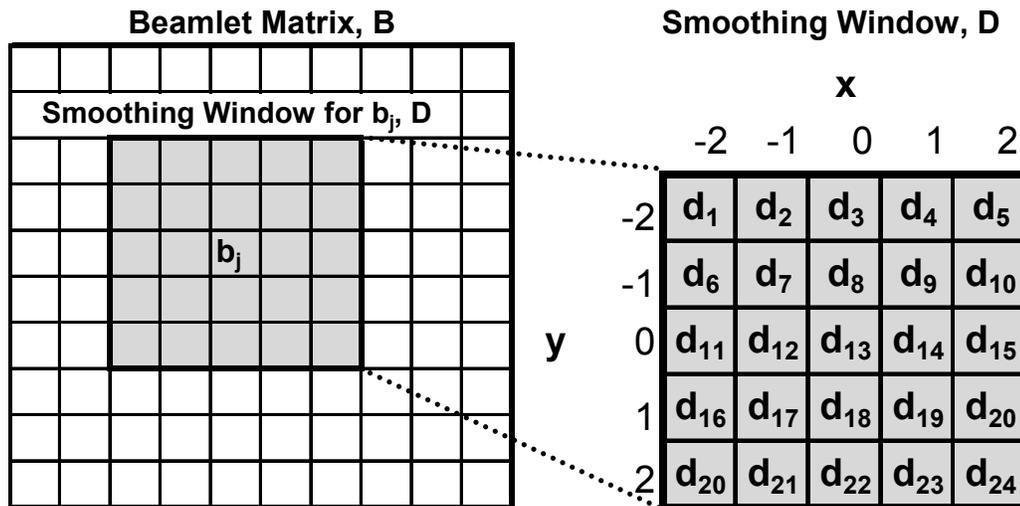


Figure IV-1. An illustration of a smoothing window, D centered at beamlet, b_j , from a beamlet intensity matrix, B . The elements within the smoothing window are denoted by a vector of intensities, d .

ability to obtain steeper intensity gradients was found when applying this procedure inside the cost function as opposed to smoothing after each iteration.

Our goal in studying Savitzky-Golay filtering for IMRT beams was to expand on the previous work by (i) using more complex and realistic clinical examples, and (ii) to

evaluate the use of a 2D filter as opposed to a filter that smoothes only in the direction of the MLC. To this end, both 1D and 2D Savitsky-Golay filtering has been implemented in UMOpt. The only difference between the two filters is the dimension of the smoothing window. The calculation of the filtered beam is quite simple using the moving least-squares approach. As a demonstration, the 2D filtering process starts with a vector of beamlet intensities \mathbf{b} , and the goal is to fit a local sub-set of the beamlets, which is denoted as \mathbf{d} . In this work, a 5 x 5 beamlet smoothing window is used. This is illustrated in Figure IV-1. Here, b_j is the beamlet whose value will be updated based on the fitted value derived from the filtering process. The beamlet matrix subset, \mathbf{D} , has a coordinate system centered at b_j . As shown in Figure IV-1, the values in \mathbf{D} are represented by the vector \mathbf{d} . The purpose of the filtering process is to fit the elements in \mathbf{D} to a 2nd order polynomial of the form:

$$s(x, y) = c_1 + c_2x + c_3y + c_4x^2 + c_5y^2 + c_6xy , \quad (\text{IV.4})$$

where \mathbf{c} is the vector of coefficients to be obtained. Determining \mathbf{c} is straightforward using the matrix equation,

$$\mathbf{M}\mathbf{c} = \mathbf{d} , \quad (\text{IV.5})$$

where \mathbf{M} is the matrix of the polynomial terms,

$$\mathbf{M} = \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & y_1^2 & x_1y_1 \\ 1 & x_2 & y_2 & x_2^2 & y_2^2 & x_2y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{25} & y_{25} & x_{25}^2 & y_{25}^2 & x_{25}y_{25} \end{bmatrix} . \quad (\text{IV.6})$$

In Equation IV.6, (x_i, y_i) are the coordinates of d_i as shown in Figure IV-1. Equation IV.5, can be solved for \mathbf{d} by multiplying by the matrix transpose:

$$\mathbf{M}^T\mathbf{M}\mathbf{c} = \mathbf{M}^T\mathbf{d} . \quad (\text{IV.7})$$

Then, inversion results in

$$\mathbf{c} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{d}, \quad (\text{IV.8})$$

which can be solved to directly obtain the polynomial coefficients, \mathbf{c} , for fitting the subset of beamlet data, \mathbf{D} . The result from Equation IV.8 can then be introduced back into Equation VI.6 to obtain the smoothed beamlet value for the original beamlet b_j , which will be referred to as s_j . This process is carried out for all beamlets in a beam to obtain a vector of smoothed beamlets, \mathbf{s} .

Replacing \mathbf{b} by \mathbf{s} would adversely affect the cost function, and thus degrade the IMRT plan unless the plan was very simple. Therefore, either reoptimization must be performed after \mathbf{s} is applied or the deviation between \mathbf{b} and \mathbf{s} can be penalized inside the cost function. As mentioned previously, the best dosimetric results were obtained by Spirou *et al.* when using the penalty method. Similarly, we have implemented a Savitzky-Golay costlet into UMOpt, which is referred to as the SG penalty:

$$\text{SG Penalty} = p \times \sum_{j=1}^J (b_j - s_j)^2 \quad (\text{IV.9})$$

Here, p is the weight assigned to the smoothing importance, J is the number of beamlets in the plan, b_j is the value of the j^{th} beamlet, and s_j is the value of the j^{th} beamlet after the smoothing operation. Both 1D and 2D SG penalties have been implemented in UMOpt, the only difference being the size of the smoothing window, which is illustrated in Figure IV-2. In the SG1D penalty, the filter is applied only in the direction of MLC leaf motion, and a 5 beamlet vector was chosen as the smoothing window so that our results could be compared to the Spirou work. Different-sized smoothing windows were tested, but did not perform as well as the 5 element window (to be discussed later).

Operating under the hypothesis that a 2D smoothing operation would be more beneficial for a real patient geometry than a 1D smoothing operation, the SG2D penalty was used to promote modulation reduction in all directions, not just in the direction parallel to MLC travel. In the SG2D penalty, a 5 x 5 beamlet grid is used as the smoothing window, in a process described below.

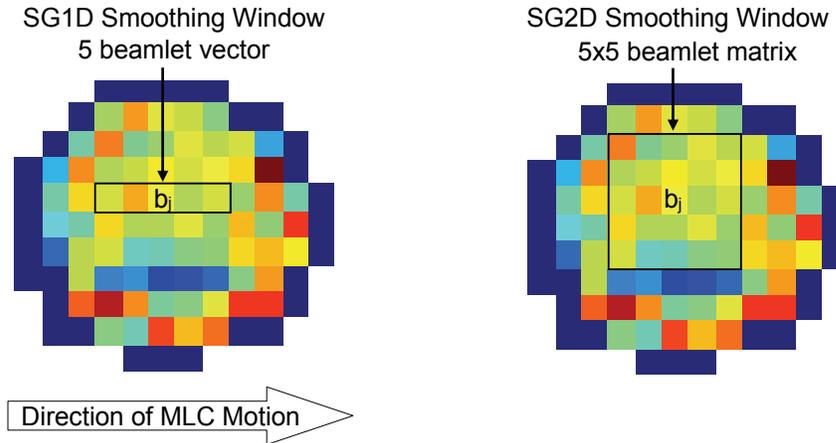


Figure IV-2. An illustration of the SG1D and SG2D smoothing windows, \mathbf{D} centered at beamlet, b_j , from a beamlet intensity pattern. The elements within the smoothing window are inside the black rectangle. Note the direction of MLC motion is shown to be parallel to the SG1D smoothing window.

To use the SG penalties during optimization with the quasi-Newton optimization algorithm, the partial derivative of the SG penalty with respect to the beamlets must be calculated. The dosimetric objective function gradients, $\partial f / \partial \mathbf{b}$, are calculated using automatic differentiation software (ADOL-C¹⁵) in UMOpt¹⁶. The SG penalty gradients were calculated in *Mathematica* (Wolfram Research, Inc., Chicago, IL), and hand-coded into the UMOpt system to be added to the existing gradients.

IV.B.2.2. Plan Intensity Map Variation Penalty

The plan intensity map variation (PIMV) in Equation IV.2 is a measure of overall beam modulation. To reduce modulation and smooth the beam as part of the cost function, a costlet was constructed to minimize a quadratic form of the PIMV. The quadratic PIMV costlet, or PIMV_q penalty is,

$$\text{PIMV}_q \text{ Penalty} = p \times \text{PIMV}_q, \quad (\text{IV.10})$$

which can be added into any weighted sum cost function, similar to the SG penalties. Again, p is the weight assigned to the smoothing importance, and PIMV_q is defined as

$$\text{PIMV}_q = \sum_{n=1}^{N_b} \left(\sum_{j=1}^{J-1} \sum_{k=1}^{K-1} \left[(b_{jk} - b_{j,k+1})^2 + (b_{jk} - b_{j+1,k})^2 + \dots + \frac{1}{2}(b_{jk} - b_{j+1,k+1})^2 + \frac{1}{2}(b_{jk} - b_{j+1,k-1})^2 \right] \right), \quad (\text{IV.11})$$

where N_b is the number of beams in a plan, J is the maximum number of beamlets in the direction parallel to the motion of the multi-leaf collimator (MLC), K is the maximum number of beamlets in the direction perpendicular to the motion of the MLC, and b_{jk} is the intensity of the beamlet at the (j,k) grid position. The four terms in Equation IV.11 allow all differences between each neighboring beamlet, including the diagonal beamlets, to be taken into account, and the quadratic version was chosen to simplify the calculation of the gradient during the optimization procedure. As in the SG penalties, the partial derivative of the PIMV_q penalty with respect to \mathbf{b} was derived using *Mathematica* (Wolfram Research, Inc., Chicago, IL) and added to the existing objective function gradients, which are calculated automatically.

The weights used in Equation VI.9 and VI.10 were varied to demonstrate the potential impact of the beam modulation penalties on the optimal intensity modulated beams as the importance of smoothness was increased.

IV.B.3. Technique Comparisons and Analysis

Each of the beam modulation reduction techniques were either already present (maximum intensity limits) or implemented into UMPlan and UMOpt.

In order to compare maximum intensity limits and modulation penalization in the cost function, each technique was used in the optimization of several IMRT plans, including a simple CT phantom and clinical example cases in the brain, prostate and head/neck. To judge the full effect of the different modulation reduction methods, plans were optimized at a variety of maximum intensity ratios and with a range of weights (p values in Equation IV.9 and IV.10) applied to the modulation penalties. A characterization of each technique was performed in the phantom, and then comparisons were made between each of the techniques using the clinical examples. In all comparisons, plans were judged with respect to the unconstrained baseline IMRT plan. These standard plans were obtained with no beam modulation penalties, and no restrictions were placed on the allowed values of the beamlet intensities. The basis for comparisons included dose metrics, DVHs, field modulation and complexity, and

efficiency of SMLC delivery. Also, the correlation coefficient was determined between intensity maps with varying maximum intensity ratios.

All cases were planned for a 6 MV linear accelerator (Varian Medical Systems, 21EX) with 120 leaf MLC (0.5 and 1.0 cm leaf widths). Dose calculations for the inverse planning system were performed by a convolution/superposition algorithm derived from the work of Mackie *et al*¹⁷. All cases were optimized using a quasi-Newton-based search strategy. Plans were initialized with several different starting intensity patterns to ensure that the optimized plans did not represent local minima of the cost function.

After optimization, leaf sequencing for static MLC (SMLC) delivery was performed with an in-house-developed leaf sequencer based on the method reported by Bortfeld *et al*.¹⁸ Delivery sequences allow up to 250 segments per beam, with the goal of achieving a correspondence between planned and delivered intensities of 1 %.

IV.B.4. Case Studies

As mentioned previously, each of the modulation penalties was validated on a CT phantom and then tested on nine clinical cases—three each for brain, prostate and head/neck. The geometry and beam arrangement for the simple CT phantom are shown in Figure IV-3, and the high priority planning objectives for the phantom and clinical cases are shown in Table IV-1. In addition, each plan included a lower priority objective

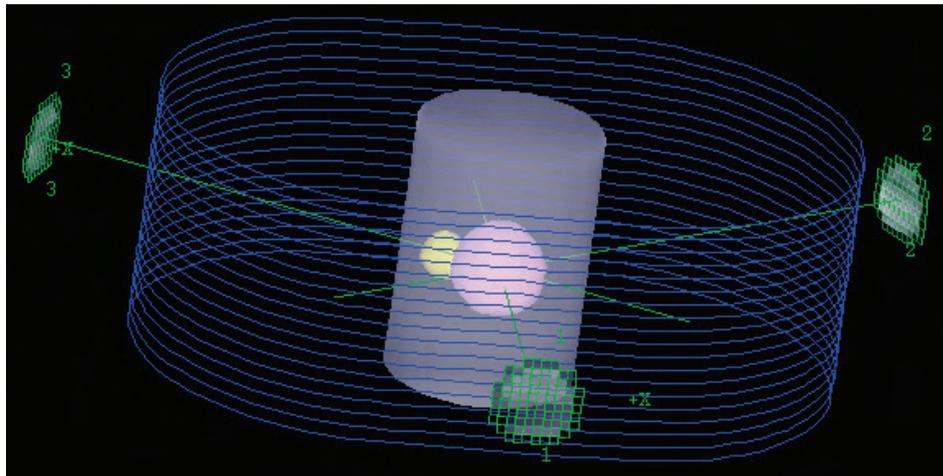


Figure IV-3. Illustration of the CT phantom geometry and beam placement. The PTV is the violet center sphere, “Small” is the yellow off-center sphere, and “Cylinder” is the cylindrical gray structure. The blue external contours were not used in the cost function to save calculation time.

Table IV-1. Inverse planning objectives for the CT phantom and clinical cases.

CT Phantom Inverse Plan Objectives	
Structure	Objectives
PTV	≥ 60 Gy; ≤ 66 Gy
PTV + 1 cm	≥ 50 Gy; ≤ 66 Gy; $\leq 20\% \geq 60$ Gy
Small	$\leq 15\% \geq 25$ Gy
Small + 1 cm	≤ 66 Gy; Mean ≤ 20 Gy
Cylinder	≤ 70 Gy

Brain Inverse Plan Objectives	
Structure	Objective
PTV1	66 Gy (min 95%, max 105%, and 1 % volume up to 110%)
PTV2	60 Gy (min 95%, max 105% of PTV1 with PTV1 coverage priority)
Optic Nerves	≤ 60 Gy
Optic Chiasm	≤ 60 Gy
Brainstem	≤ 65 Gy

Prostate Inverse Plan Objectives	
Structure	Objectives
Prostate + 3 mm	Mean = 75.85 Gy +/- 3 %; ≤ 0.5 cc down to 93 %; ≤ 0.5 cc up to 115 %
Rectum	$\leq 15\% \geq 80$ Gy; $\leq 25\% \geq 75$ Gy; $\leq 35\% \geq 70$ Gy; $\leq 50\% \geq 65$ Gy
Bladder	$\leq 15\% \geq 80$ Gy; $\leq 25\% \geq 75$ Gy; $\leq 35\% \geq 70$ Gy; $\leq 50\% \geq 65$ Gy
Femora	Mean ≤ 50 Gy; $\leq 10\% \geq 52$ Gy
Penile Bulb	Mean ≤ 52.5 Gy; $\leq 15\% \geq 70$ Gy
Uninvolved Tissue	Max $\leq 100\%$ Rx Dose; Mean ≤ 52.5 Gy; $\leq 15\% \geq 70$ Gy

Head/Neck Inverse Plan Objectives	
Structure	Objectives
PTV	70 Gy +/- 5 %
Nodal Boost PTV	70 Gy +/- 5 %
High Risk Nodal PTV	66 Gy +/- 5 %
Low Risk Nodal PTV	60 Gy +/- 5 %
Spinal Cord	≤ 45 Gy
Spinal Cord + 5 mm	≤ 50 Gy
Brainstem	≤ 54 Gy
Contralateral Parotid	Mean ≤ 26 Gy
Ipsilateral Parotid	Mean ≤ 26 Gy (if possible, otherwise minimize)
Mandible	≤ 70 Gy
Submandibulars	Minimize dose
Oral Cavity	≤ 70 Gy

of minimizing overall dose to all of the normal structures and uninvolved normal tissues. These objectives were given a low weight, so as to not interfere with meeting the highly-weighted objectives in Table IV-1. The clinical plan objectives were chosen based on in-house IMRT protocols.

Brain cases were optimized with four or five beams originally placed by a dosimetrist, prostate plans were optimized with 9 equally-spaced axial beams, and head/neck cases were optimized using 7 equally-spaced axial beams.

Optimization results for each of the different techniques were compared using dose metrics, DVHs, field modulation and complexity, and efficiency of SMLC delivery. The results from the modulation penalty validation in the CT phantom are presented first, and then all methods are compared in the three clinical sites. Our analysis of these results attempted to determine which method produced the highest quality IMRT plans, taking into account the target and normal tissue planning objectives and other important factors such as overall beam complexity and delivery efficiency.

IV.C. Results

IV.C.1. Characterization of Methods in the CT Phantom

The CT phantom case (Figure IV-3) was optimized with the baseline objective function described in Table IV-1, and then optimization was performed including maximum intensity limits and the three different modulation penalties. Each individual method is characterized below, and the methods are compared.

IV.C.1.1. Maximum Intensity Limits

The application of beamlet intensity restrictions during the optimization of the phantom case successfully improved delivery efficiency as long as reasonable maximum intensity levels were applied. At extreme intensity limits, the optimization was severely hindered and unable to reach the dosimetric objectives. Figure IV-4 shows the trend in monitor units, mean correlation coefficient, and plan intensity map variation as the maximum intensity ratio (MIR, Equation IV.1) decreases. The MIR value is shown below the data points and the x-axis is given in terms of total cost. This is a measurement of plan quality according to the objective function. As the MIR decreases, the total cost increases because the maximum intensity limit increasingly interferes with the plan reaching the optimal quality possible with unconstrained beamlet values. At the same

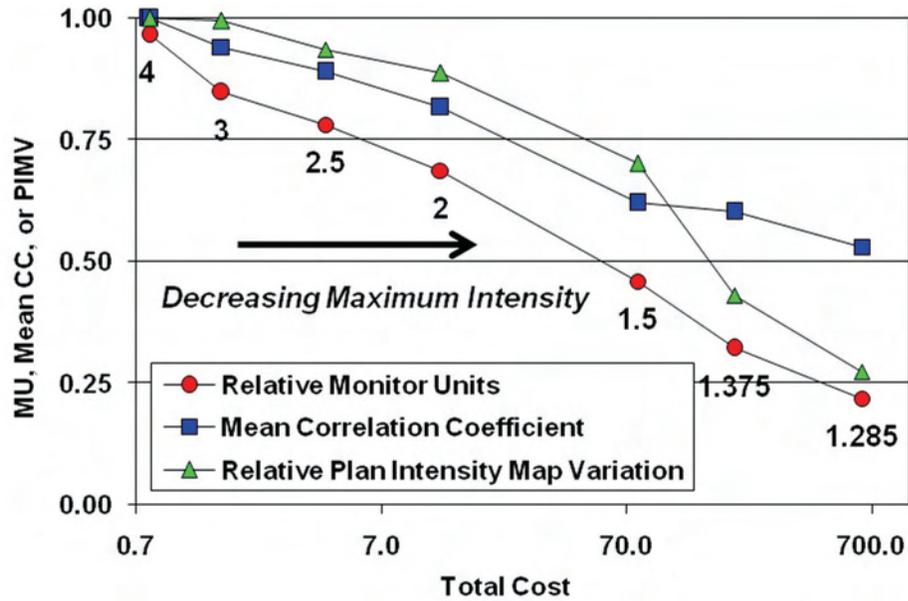


Figure IV-4. The relative monitor units, mean correlation coefficient, and relative plan intensity map variation as a function of total cost for the CT phantom when optimized with different maximum intensity limits on the beamlets. The maximum intensity ratio for each plan is shown below the lines. All results are relative to baseline plan which had a maximum intensity ratio of 4. The correlation coefficient value was averaged over all beams.

time, MU and PIMV decrease as the fields become flatter and more uniform. The correlation coefficient with respect to the baseline plan, averaged over each beam, also decreases with decreasing MIR. This demonstrates that the intensity limit forces the beam intensity pattern further and further from the baseline pattern. However, at high MIR values, the mean correlation coefficient remains fairly high. It can also be noted that the plan becomes very sensitive to the MIR value as it approaches 1. The cost, MU, CC, and PIMV change substantially from 1.5 to 1.285, suggesting that the MIR is reaching its lower limit to be able to achieve the necessary target coverage. This is supported by Figure IV-5 which shows qualitative results in the form of dose-volume histograms and the beamlet intensity patterns for the Baseline (MIR = 4), and MIR = 2.5, 2, 1.375, and 1.285. At the lower MIR values, the beams become very uniform and at the lowest MIR tested, the target coverage begins to suffer. Conversely, at the higher MIR values, the beams remain highly modulated with evident cut-off points in the beamlet intensities.

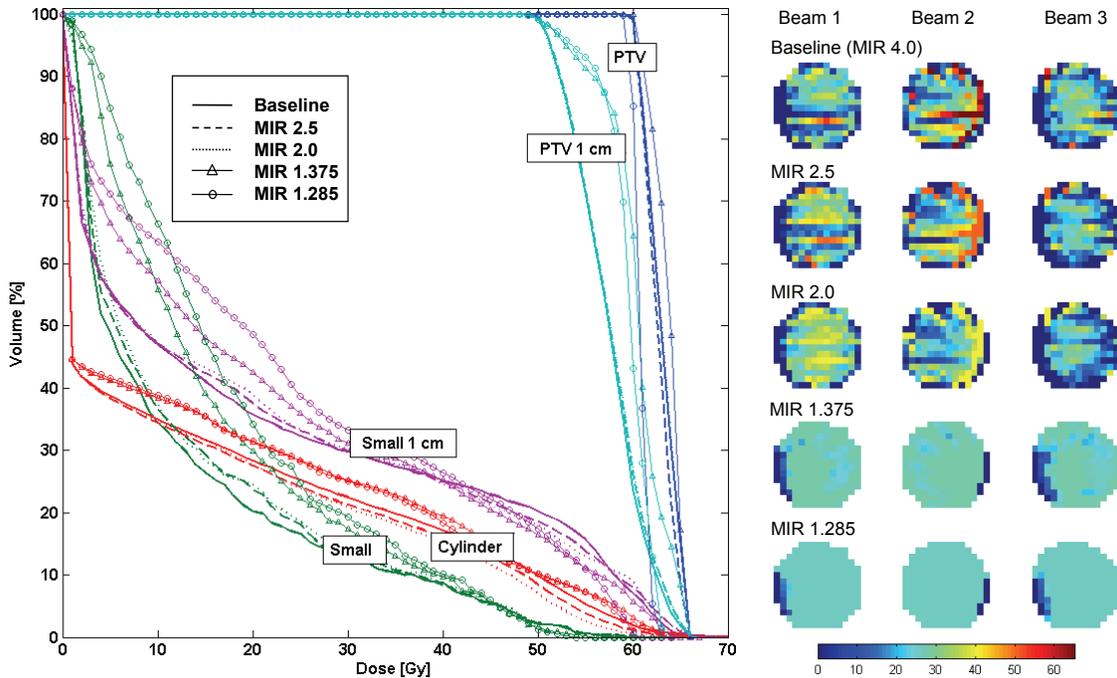


Figure IV-5. DVHs and beams for plans at various maximum intensity ratios (MIR) in the CT phantom.

Judging by the DVHs, the dosimetric effect of the maximum intensity limit is small when there is little restriction. As the MIR decreases, there is an increase in the amount of low dose delivered to the normal tissues and an increase in mean dose to the normal tissues and the PTV expansion. Still, it appears clear that some of the unrestrained beamlet intensities are unnecessary and detrimental to the delivery efficiency. The plan at a MIR of 2 is very similar to the baseline plan and can be delivered with a 31 % reduction in MU. These results suggest that the application of a moderate maximum intensity limit could be a simple and practical way to improve delivery efficiency.

IV.C.1.2. Savitzky-Golay Penalties

Figure IV-6 illustrates the trend in MU, mean CC, and PIMV as the 1D and 2D Savitzky-Golay (SG) modulation penalty weights increase and objective function value also increases. Similar to the maximum intensity limited plans, the SG plans move away from the baseline solution with the MU, CC, and PIMV dropping as the SG penalty increases. Unlike the maximum intensity limited plans however, the MU, CC, and PIMV decrease more quickly and result in lower values at similar cost values.

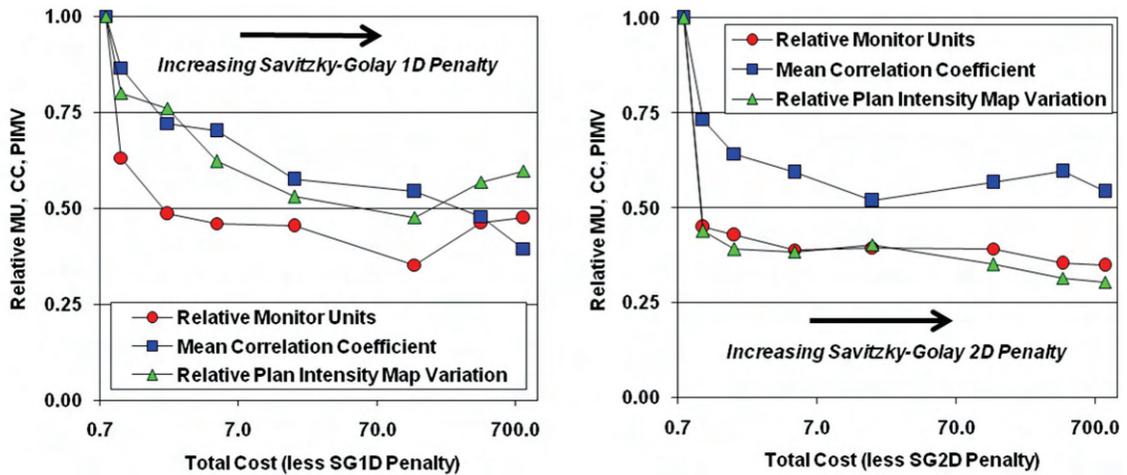


Figure IV-6. The relative MU, mean correlation coefficient, and relative PIMV as a function of total cost for the CT phantom when optimized with the SG1D or SG2D penalty in the objective function. All results are relative to baseline plan. The correlation coefficient value was averaged over all beams.

Compared to each other, the SG1D PIMV is much higher than the SG2D plans, while the SG1D MU are only slightly higher. This can be explained by the fact that the SG1D penalty promotes smoothness only in the direction parallel to the MLC motion. This aids in MU reduction, but it still allows for high intensity gradients between each row of beamlets. The mean CC also falls off more slowly when the SG1D penalty is used, which could also be explained by the additional modulation being able to better mimic the intensity patterns in the baseline plan. The CC falls off very quickly at higher weights accompanied by undesirable increases in the MU and PIMV. This could be a result of the optimizer being dominated by fitting the beam to the SG smoothed beam at higher weights. Therefore, this penalty may be less useful at high weights, where non-conformance to the filtered beams is highly penalized.

Figure IV-7 and IV-8 display the DVHs and intensity modulated beams for the SG1D plans and SG2D plans, respectively. There is a noticeable difference in the trend of the intensity patterns between the 1D and 2D penalties. The SG1D beams display almost step-like behavior in the direction perpendicular to the MLC motion. This is because there is no penalty for modulation in this direction. As Spirou *et al.* suggest, a 5 beamlet smoothing window was used. The use of a larger smoothing window resulted in an exaggeration of this step function observation, with each row of beamlets having an almost identical weight. Conversely, the use of a smaller window resulted in minimal smoothing and minimal reduction in MU as compared to the baseline plan.

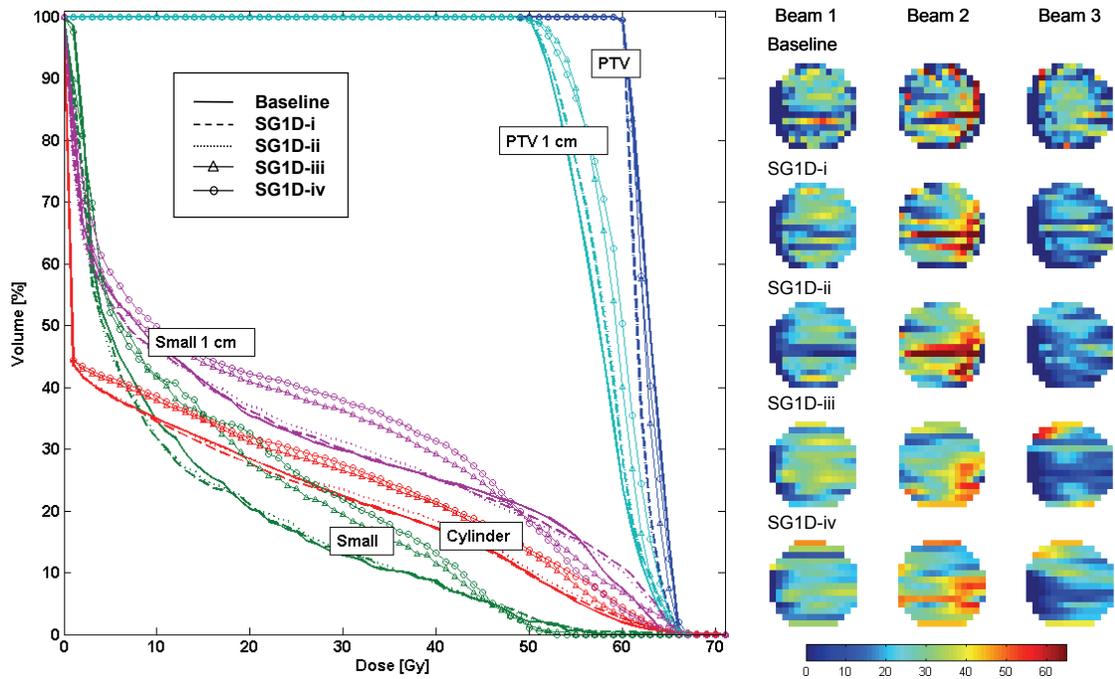


Figure IV-7. DVHs and beams for plans at various SG1D penalty weights in the CT phantom. Weights in increasing order are given by i-iv.

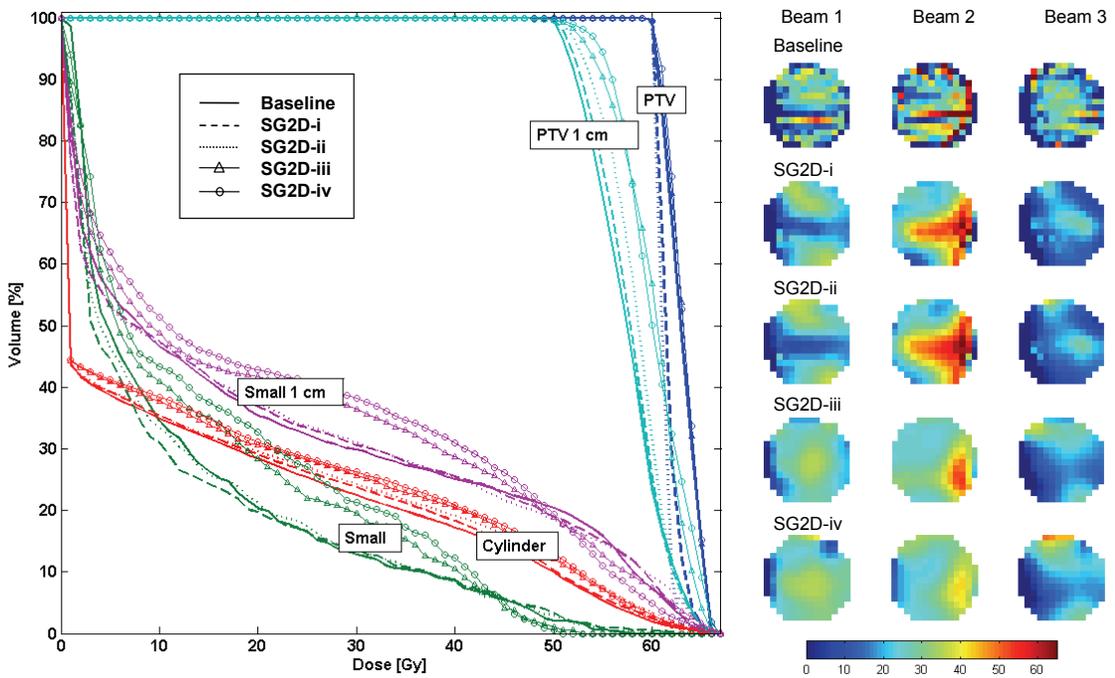


Figure IV-8. DVHs and beams for plans at various SG2D penalty weights in the CT phantom. Weights in increasing order are given by i-iv.

In addition to the potentially detrimental step-like behavior in the intensity patterns, there is also a break-down in the modulation trend at higher SG1D penalty weights with the iii and iv beams being quite distinct. Also, several rows of beamlets appear to be very uniform.

Although there are some negative features seen at higher SG1D penalty weights, the target coverage is still maintained due to the importance of target coverage in the cost function. This is also true in Figure. IV-8. The main difference between the SG1D and SG2D results is in the beams. The use of the SG2D penalty produced more highly smoothed versions of the original plan with large reductions in (i) the amount of the modulation, (ii) the number of high intensity beamlets, and (iii) the size of beamlet-to-beamlet variations. The SG2D beams are much smoother overall, since smoothing is promoted in two directions instead of one. The change in beam modulation as the SG2D penalty weight increases is more intuitive and predictable than in the SG1D penalty. The DVHs for both sets of plans are similar. In general, the SG2D plans are smoother and can be delivered with fewer MU at similar cost values and DVHs.

IV.C.1.3. Plan Intensity Map Variation Penalty

The quadratic plan intensity map variation penalty ($PIMV_q$) gave similar results to the SG2D penalty in terms of the reduction in MU, CC, and PIMV at increasing penalty weights (Figure IV-9). As would be expected, the PIMV metric is the lowest at similar cost values when using the $PIMV_q$ penalty compared to the other methods. In addition, the behavior of each of the curves in Figure IV-9 is very consistent at increasing weights, and the MU is the lowest of all methods at similar cost values. The net cost increases because as the modulation penalty increases, the plan deviates further from the baseline optimized plan. To minimize the modulation penalty, tradeoffs are usually made with the lower priority objective of minimizing total dose to the normal tissues. This is indicated by the slowly increasing net cost at the lower penalty weights. As the weight further increases, there is a steeper increase in the net cost as the modulation penalties begin to dominate the optimization and various high priority objectives are violated. In a clinical case, one would likely not choose a smoothing penalty weight that interfered with the high priority objectives. The extent to which the modulation penalty would be allowed to trade off with lower priority objectives would have to take into account the increase in normal tissue dose versus the reduction in MU and beam complexity.

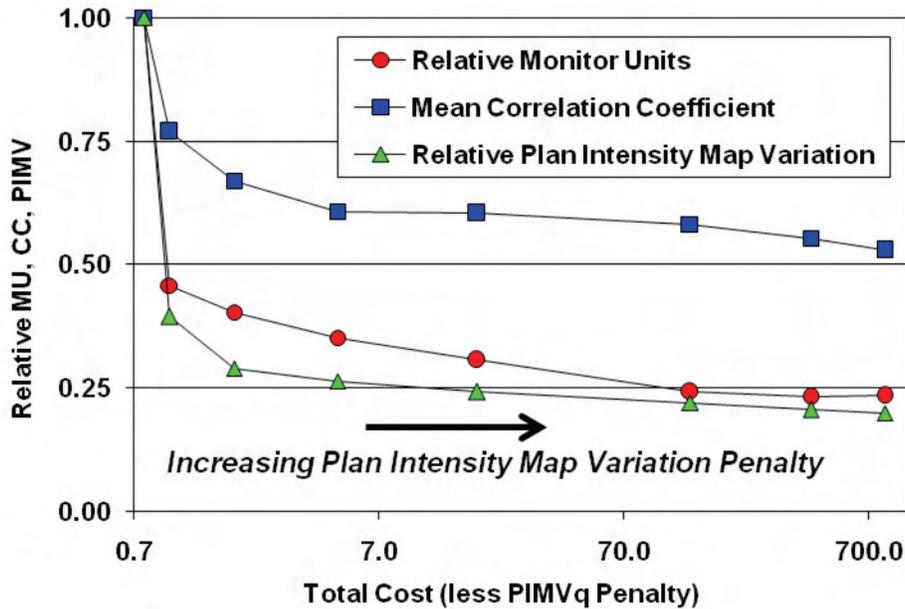


Figure IV-9. The relative MU, mean correlation coefficient, and relative PIMV as a function of total cost for the CT phantom when optimized with the PIMV_q penalty in the objective function. All results are relative to baseline plan and the correlation coefficient value was averaged over all beams.

In addition to the desirable behavior of the MU and PIMV when using the PIMV_q penalty, there is also an additional advantage of directly penalizing the modulation without having to compute a filtered beam and then assign a deviation-based penalty. Thus, when using the PIMV_q penalty, it would be very unusual to observe the MU or PIMV increasing with increasing penalty weights, as in the SG1D plans.

Figure IV-10 shows the DVHs and beams for the PIMV_q penalized plans. While the DVHs are similar to the SG penalties, there are striking differences in the intensity modulated beams. The PIMV_q penalized fields are relatively uniform, with the second beam delivering the majority of the intensity. In these beams, the modulation was high only in areas near the overlap of the PTV and the “Small” structure. Looking at the large differences in the beam intensity patterns from each of the different modulation penalties, it appears that the phantom case has a large solution space, i.e., there are many different beam combinations that lead to very similar DVH results.

For this CT phantom example, the PIMV_q penalty achieved the same total objective function values with lower degrees of beam complexity and MU than the SG penalties. However, tradeoffs are still being made within the dose objectives, and each

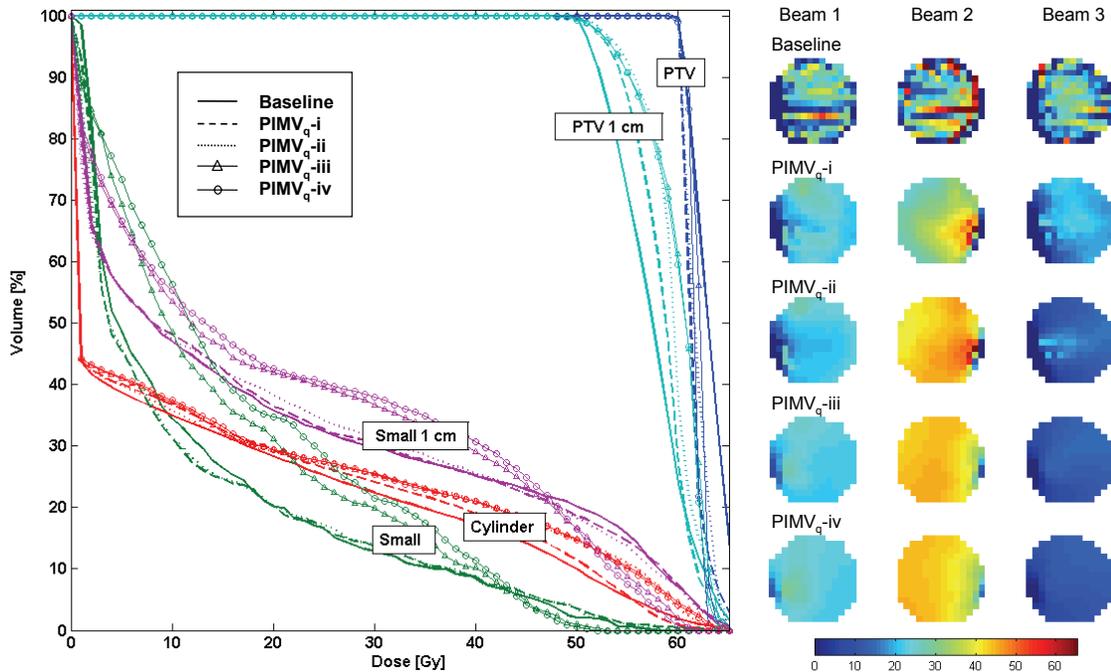


Figure IV-10. DVHs and beams for plans at various PIMV_q penalty weights in the CT phantom. Weights in increasing order are given by i-iv.

of the modulation penalties causes slightly different tradeoffs to be made in the cost function. Thus, each of the penalties may have merit in different geometries and in situations where planners may have different tradeoff preferences. It is likely that the SG1D and SG2D penalties will be useful in cases that require high intensity gradients across a field, and that the PIMV_q penalty will be most applicable because it generally provides the same objective value for the least amount of beam complexity. These ideas will recur in the following discussion of our results from the brain, prostate, and head/neck trials.

IV.C.2. Clinical Studies

Nine clinical cases were optimized using each of the three modulation penalties in the objective function at varying weights. The results were compared to plans using the baseline objective functions without or with maximum beamlet intensity limits. The inclusion of the modulation penalties at reasonable weights did not compromise the clinical planning objectives and generally resulted in smoother, less complex intensity patterns that could be delivered with significant reductions in monitor units compared to the baseline IMRT plans. This gain in delivery efficiency was sometimes at the expense

of increased dose to the normal tissues. Therefore, in addition to the results presenting the maximum MU reductions while still meeting the clinical objectives, we have included the maximum MU reduction while maintaining all normal tissue mean doses achieved with the baseline cost function. This was done by adding mean dose costlets along with the modulation penalties to the baseline cost function at the mean dose levels achieved by the baseline plan optimization. These results will be discussed for each individual treatment site.

IV.C.2.1. Brain

In the three brain cases tested, each beam complexity reduction technique produced plans that met the high priority dose objectives and could be delivered with a significant MU reduction compared to the baseline IMRT plan. Figure IV-11 shows all of the optimization runs plotted as relative MU versus net cost. In each plan shown, all of the high priority objectives in Table IV-1 are met. The increase in delivery efficiency seen in the brain is usually gained as a result of a tradeoff between the modulation penalty (or

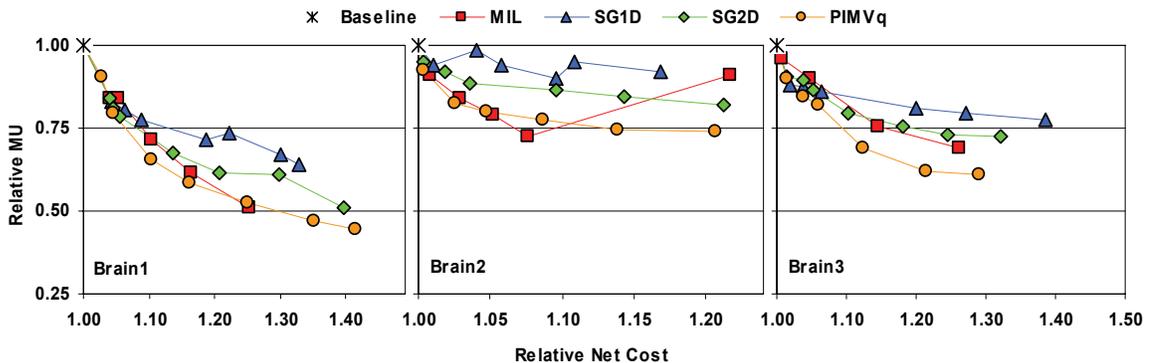


Figure IV-11. Plots illustrating relative MU versus net cost tradeoffs in the brain cases when using each of the modulation penalties compared to a standard cost function and maximum intensity limits. All values are relative to the standard plan.

maximum beamlet intensity) and the low-priority objective of minimizing overall dose to the normal tissues. This can be seen in Figure IV-12 which shows, for all cases, the optimized DVHs with the baseline cost function, with maximum intensity limits, and with each of the modulation penalties applied. The plans shown are at moderate modulation penalty weights and are very similar in net cost. The intensity modulated beams from each of the techniques are shown in (b), (d), and (f). A reduction in overall beam modulation from the baseline cost function to the other techniques can be seen, and the

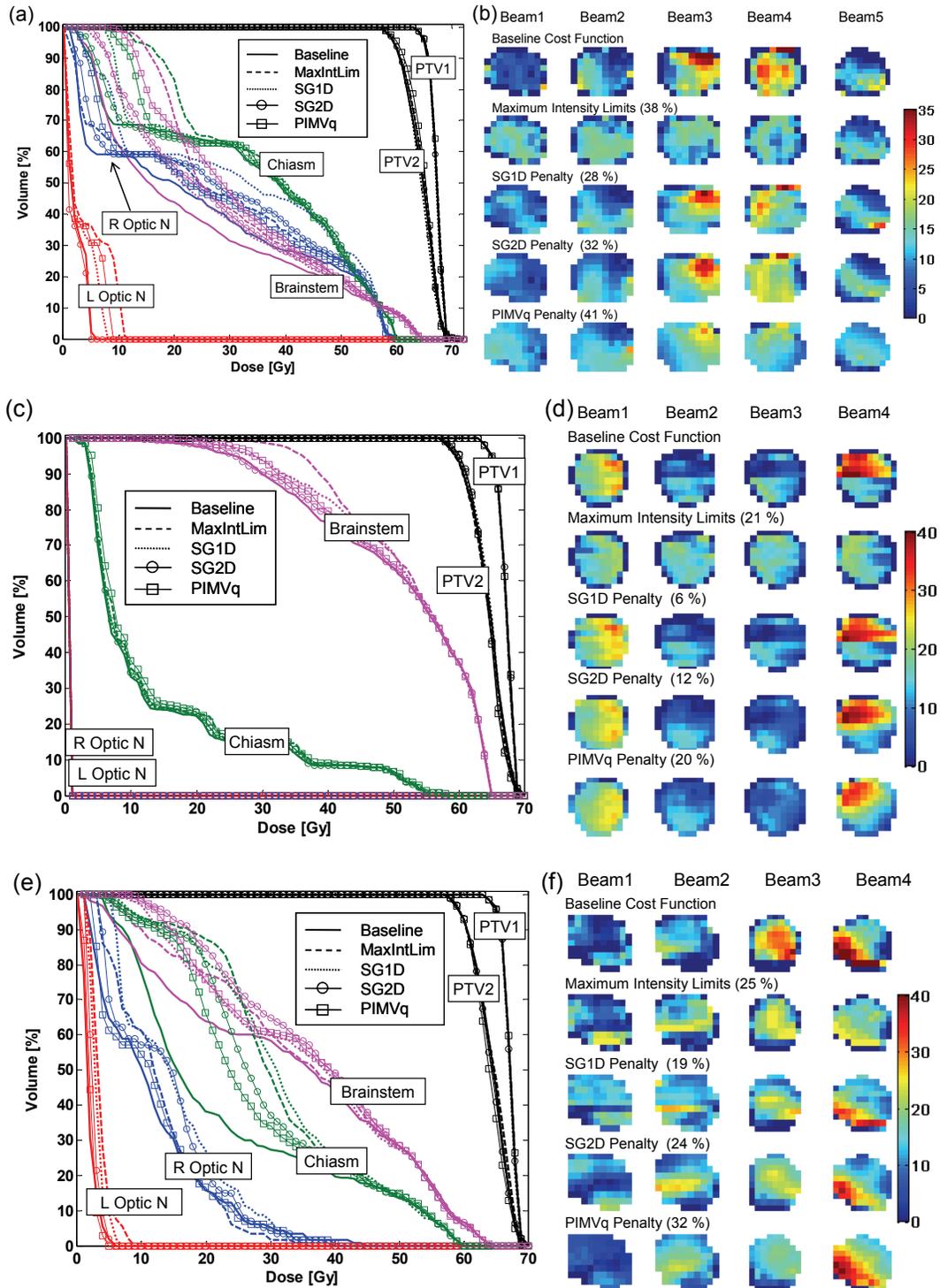


Figure IV-12. DVHs and corresponding beams shown optimized with each of the beam modulation reduction methods for (a,b) Brain 1, (c,d) Brain 2, and (e,f) Brain 3. SMLC MU reductions are in parentheses.

MU reductions are shown in parentheses. The same general trend in relative MU was observed in all brain cases with the PIMV_q modulation penalty, consistently producing the largest reduction in MU at similar objective function values, followed by maximum intensity limits, the SG2D penalty, and then the SG1D penalty (Figure IV-11).

The maximum possible MU reductions for each method compared to the baseline plan are shown in Table IV-2 for each of the brain cases. The maximum possible MU reduction is defined as the maximum MU reduction possible while still meeting all of the high priority clinical planning objectives in Table IV-1. In practice, these plans may not be acceptable to the physician because there can be significant mean dose increases in the normal tissues. Therefore, also included is the more modest maximum MU reductions possible when maintaining the mean normal tissue doses from the baseline plans. Realistic MU reductions for clinical practice may lie between these two values, and would be based on the individual plan and physician tradeoff preferences. The smaller decreases in MU for *brain2* may be explained by a lack of modulation in the fields for the original cost function, as the plan was relatively simple without involvement of the optic nerves. The SG penalties may not have resulted in a large MU reduction because the original beams already closely resembled the filtered beam. Since the PIMV_q plans were penalized based on the total modulation, they have the greatest MU decrease, along with the relatively flat maximum intensity limited plans.

Table IV-2. Maximum MU reductions possible in the brain using each optimization technique while (i) meeting high priority plan objectives, and (ii) maintaining mean normal tissue doses from the baseline plan.

Optimization Technique	<i>Brain1</i>		<i>Brain2</i>		<i>Brain3</i>	
	(i) only	(i) and (ii)	(i) only	(i) and (ii)	(i) only	(i) and (ii)
Maximum Intensity Limits	49.0%	9.4%	27.3%	2.5%	31.1%	4.7%
SG1D Filter Penalty	35.8%	3.9%	10.1%	0.0%	22.4%	13.5%
SG2D Filter Penalty	49.2%	12.7%	18.1%	3.6%	27.5%	12.9%
PIMV _q Penalty	56.7%	16.4%	25.8%	12.8%	39.1%	13.1%

IV.C.2.2. Prostate

In the three prostate cases tested, each technique produced plans that met the high priority dose volume objectives, and demonstrated significant reductions in MU. Relative

MU are plotted versus net cost in Figure IV-13 for all of the prostate cases tested. In *prostate1*, the MU reductions were largest for plans using maximum intensity limits or the PIMV_q penalty. In *prostate2* and *prostate3*, the PIMV_q plans consistently required the fewest MU at a similar net cost and demonstrated the most predictable and consistent behavior at higher weights compared to the Savitzky-Golay filter-based modulation penalties. All techniques, however, produced acceptable plans with large MU reductions. The main tradeoff for improved delivery efficiency was, again, a slight increase in overall dose to the normal tissues. This can be seen in the DVH

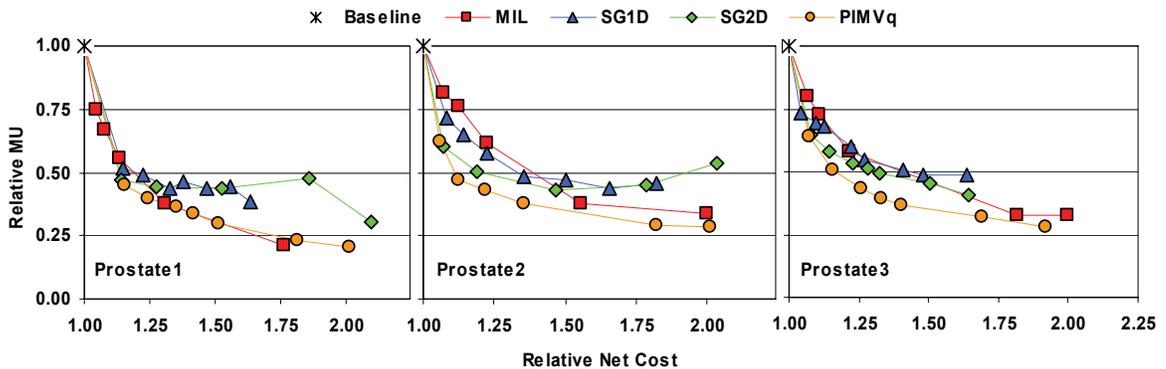


Figure IV-13. Plots illustrating relative MU versus net cost tradeoffs in the prostate cases when using each of the modulation penalties compared to a baseline cost function and maximum intensity limits. All values are relative to the standard plan.

comparisons for all methods in Figure IV-14 (a), (c), and (e). All plans shown met the objectives in Table IV-1, and the plans shown for each patient have similar values of net cost. The DVHs for the plans with modulation penalties are all very similar, while the maximum intensity limited plan had the highest dose to the rectum along with reduced dose to the femora. This demonstrates the availability of plan tradeoffs possible to achieve the same objective function value.

Four of the nine beams from each of the optimized plans are shown in Figure IV-14 (b), (d), and (f). Here the maximum intensity limited beams look very different from the rest of the beams, supporting the large shape differences that were observed in the plan DVHs. The reduction in overall beam complexity for the non-baseline plans is apparent in the intensity maps, and the MU percent reductions compared to the baseline cost function are shown in parentheses.

To further illustrate the tradeoff between beam smoothing and normal tissue mean dose, Figure IV-15 shows mean doses plotted as a function of relative SMLC MU

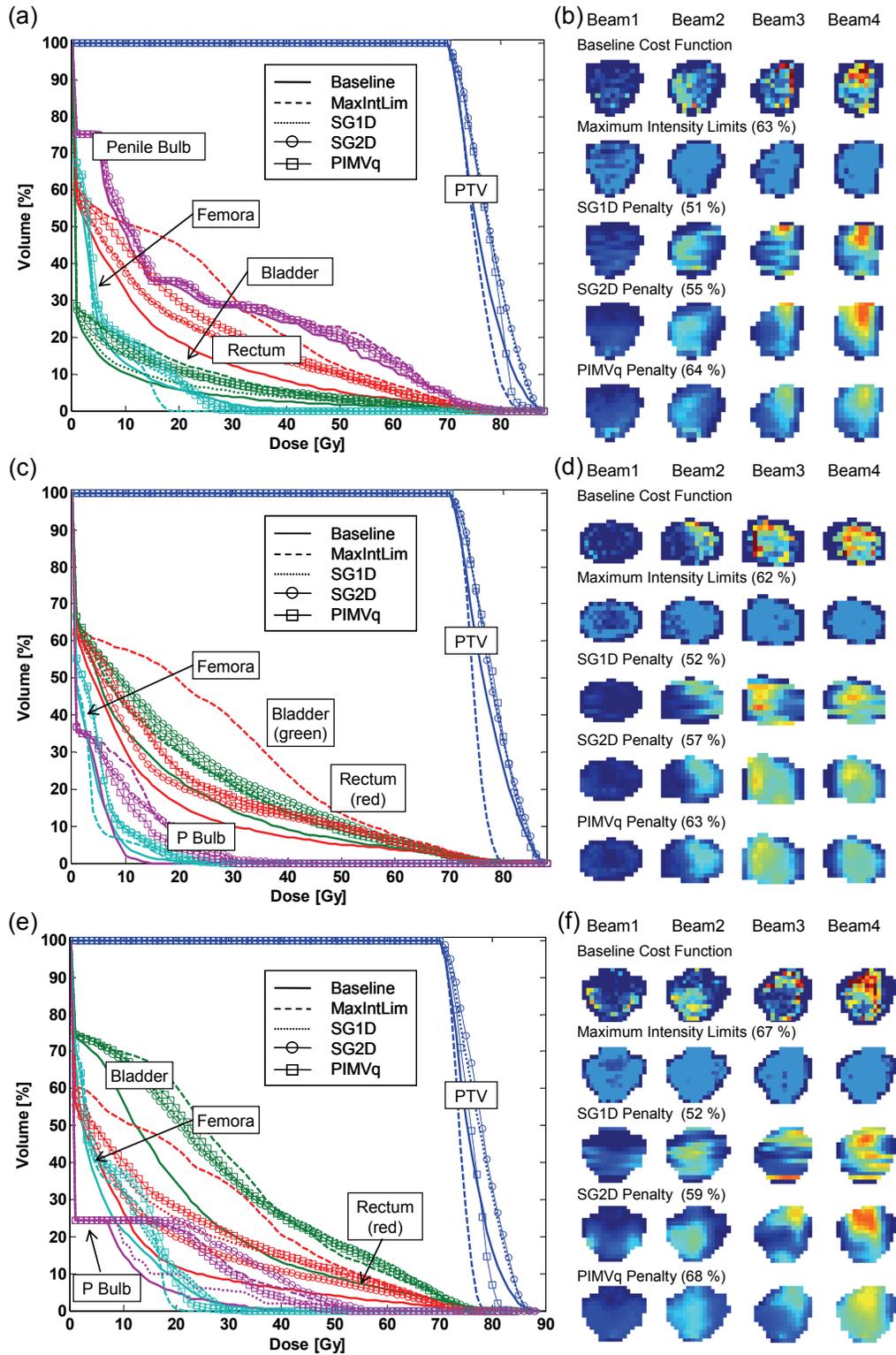


Figure IV-14. DVHs and corresponding beams 1-4 out of 9 shown optimized with each of the beam modulation reduction methods for (a,b) Prostate 1, (c,d) Prostate 2, and (e,f) Prostate 3. SMLC MU reductions are in parentheses.

for each of the normal structures in each prostate case. The PTV mean and minimum doses are also shown. This demonstrates that the minimum dose to the target stays constant as the modulation penalties increase. However, the mean doses to the normal tissues increase as the MU decrease. This is more prominent in structures that are

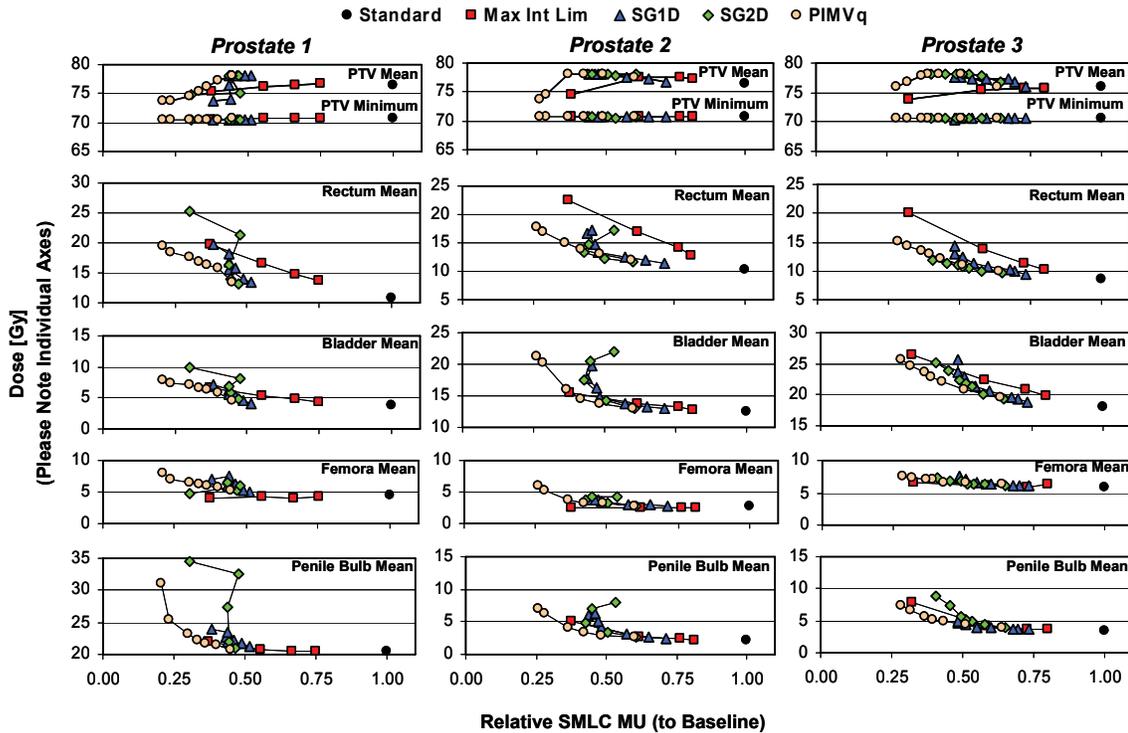


Figure IV-15. Doses in each prostate case plotted as a function of SMLC MU when using each of the optimization techniques. Trends in the PTV minimum and mean dose are shown along with normal tissue mean doses.

Table IV-3. Maximum MU reductions possible in the prostate using each optimization technique while (i) meeting high priority plan objectives, and (ii) maintaining mean normal tissue doses from the baseline plan.

Optimization Technique	Prostate1		Prostate2		Prostate3	
	(i) only	(i) and (ii)	(i) only	(i) and (ii)	(i) only	(i) and (ii)
Maximum Intensity Limits	62.6%	20.9%	62.4%	12.3%	67.6%	12.1%
SG1D Filter Penalty	61.8%	12.3%	56.1%	2.2%	51.6%	8.8%
SG2D Filter Penalty	69.6%	26.3%	57.2%	16.2%	59.4%	21.9%
PIMVq Penalty	79.2%	26.6%	71.5%	18.9%	71.7%	26.4%

close to the target, such as the rectum and bladder. The plots for *prostate1* and *prostate2* also show the inconsistent behavior of the SG penalties at higher weights.

Table IV-3 shows the maximum MU reductions possible when applying each of the modulation reduction methods. The first column shows the maximum MU reduction possible while still meeting all of the high priority plan objectives, but allowing for tradeoffs with the lower priority objective of minimizing overall dose to the normal

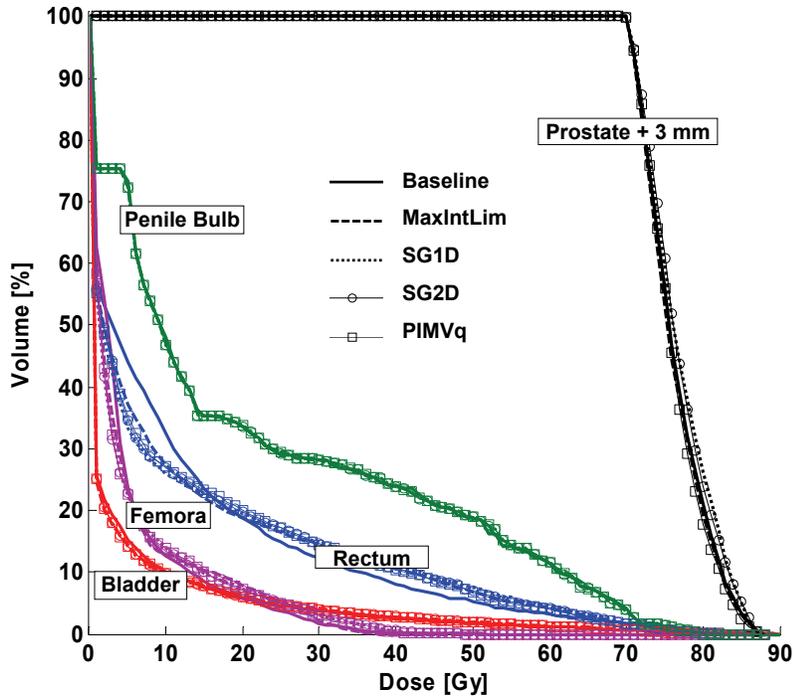


Figure IV-16. *Prostate1* DVHs shown optimized with each of the beam modulation reduction methods and maximum intensity limits when including mean dose costlets for the normal tissues equal to the mean doses achieved by the baseline plan.

tissues. The second column shows the MU reduction possible while meeting the high priority objectives and maintaining the mean normal tissue doses achieved in the baseline plan. The DVHs for the *prostate1* plans that correspond to the latter are shown in Figure IV-16. There is a slight change in the shape of the rectum DVH, but the rest of the normal tissue and target DVHs are unchanged. Thus, MU reductions occur (on the order of 18.9 % - 26.6 %) when applying the PIMV_q penalty in the prostate with essentially no loss in clinical quality compared to the baseline plans. These MU reductions are actually similar to the reductions seen in what appear to be “worse” plans in Figure IV-14.

This demonstrates one of the disadvantages of using a weighted sum cost function and shows why one must be explicit in the definition of the objectives. If dose to the normal tissues is weakly penalized, as it usually is in a weighed sum cost function so as to not interfere with the more important objectives, the plan will have little incentive to reduce that dose. In Figure IV-16, where costlets that limit the normal tissue mean doses to those we know were already achievable in the baseline plan are included, those limits were able to be adhered to while still reducing the MU. It should be noted that the considerable MU decreases observed in the prostate compared to the brain are likely a function of both baseline plan complexity and the reduced beamlet size (0.5 cm by 0.5 cm as compared to the 1 cm by 1 cm) used in the prostate plans.

IV.C.2.3. Head/Neck

The head/neck body site included the highest number of targets and normal structures in the cost function. The cases tested had large primary PTVs and nodal volumes, requiring large treatment fields that encompassed substantial (if not all) volumes of many important normal structures. Consequently, there was little room to improve plans based

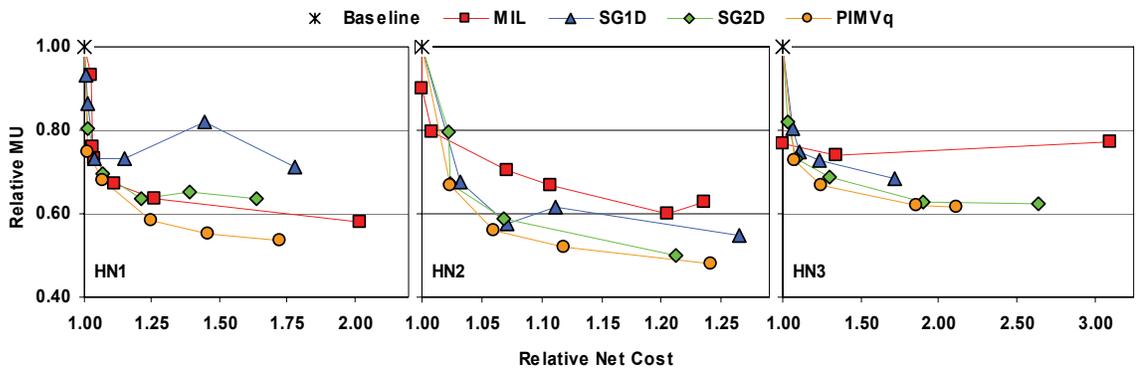


Figure IV-17. Plots illustrating relative MU versus net cost tradeoffs in the head/neck cases when using each of the modulation penalties compared to a standard cost function and maximum intensity limits. All values are relative to the standard plan.

upon the addition of a beam modulation penalty while not interfering with the high priority dose objectives. Therefore, the MU reductions observed (Figure IV-17) for the three head/neck plans were smaller and plateaued more quickly than those in the prostate. Still, it was possible to reduce modulation and MU with little loss in plan quality according to the DVHs and dose metrics. As in the other body sites, target coverage was not

compromised, due to the design of the cost function. Plans became unacceptable as the modulation penalty grew too large, and in these cases, the first thing to be violated was generally the sparing of one parotid or the spinal cord maximum dose. The PIMV_q method penalizes all modulation, no matter the location; therefore PIMV_q penalized plans have the largest mean spinal cord doses. While a small change may not be clinically important, larger increases in mean cord dose, such as those obtained with higher penalty weights, may warrant a change in the cost function, or use of a smoothing mechanism that would not penalize modulation near the spinal cord. All other mean doses and DVHs are very consistent throughout the optimization runs, with mean doses varying, in most cases, by less than 1 Gy.

The PIMV_q penalty produced the highest MU reductions, followed by the maximum intensity limits, SG2D, and SG1D penalties, with the SG1D penalty producing the most inconsistent MU reductions (see Figure IV-17). The highest MU reductions possible while still meeting all of the high priority objectives are shown in Table IV-4. As in the other sites, it was also possible to achieve substantial MU reductions while maintaining the mean normal tissue doses achieved in the baseline plan (Table IV-4).

Table IV-4. Maximum MU reductions possible in the head/neck using each optimization technique while (i) meeting high priority plan objectives, and (ii) maintaining mean normal tissue doses from the baseline plan.

Optimization Technique	<i>HN1</i>		<i>HN2</i>		<i>HN3</i>	
	(i) only	(i) and (ii)	(i) only	(i) and (ii)	(i) only	(i) and (ii)
Maximum Intensity Limits	41.8%	26.9%	40.0%	37.1%	22.6%	18.3%
SG1D Filter Penalty	28.8%	28.8%	45.3%	36.2%	31.4%	25.7%
SG2D Filter Penalty	36.5%	34.0%	50.8%	50.8%	37.5%	29.5%
PIMV _q Penalty	47.4%	35.6%	52.3%	52.3%	38.5%	33.1%

HN2 had additional boost targets within each nodal volume, making it difficult to meet all of the objectives. Many plans had difficulties in satisfying the target dose homogeneity requirements due to the number of overlapping targets. Because of this, the limiting factor in the optimization was not the normal tissue dose. This also explains why the MU reductions were similar when adding in the mean normal tissue dose

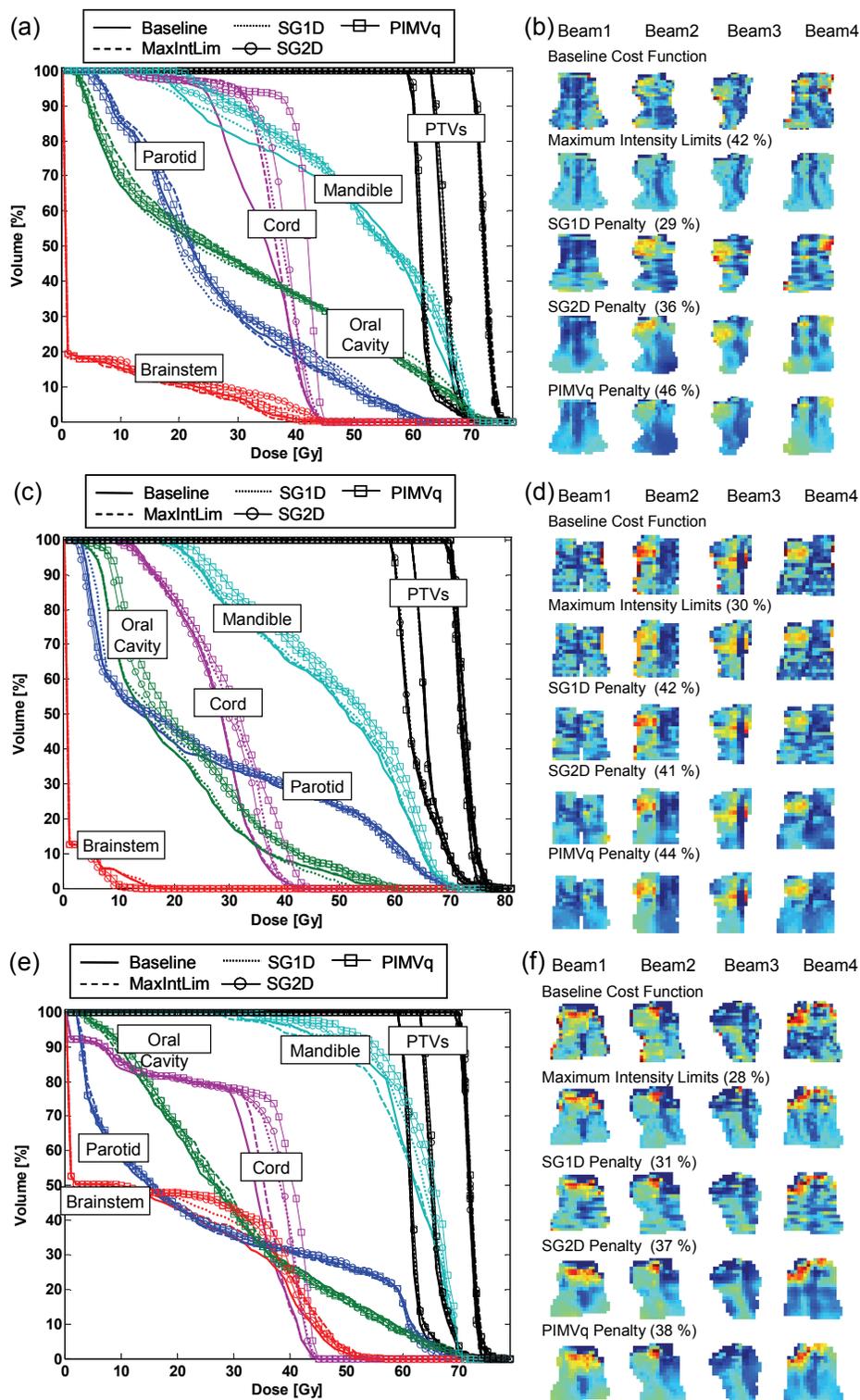


Figure IV-18. DVHs and corresponding beams 1-4 out of 9 shown optimized with each of the beam modulation reduction methods for (a,b) HN 1, (c,d) HN 2, and (e,f) HN 3. SMLC MU reductions are in parentheses.

objectives in Table IV-1. The performance of the maximum intensity limited plans was poor in this case, demonstrating the need for the higher intensities to reach the homogeneity goals in the targets. The PIMV_q penalty, at higher weights, successfully reduced MU with almost no change in plan DVHs or metrics. In this case there was only a 1.7 Gy increase in mean dose to spinal cord at the highest PIMV_q penalty. This plan represented a 52 % reduction in MU as compared to the baseline cost function.

The DVHs and intensity modulated beams for all cases and techniques are shown in Figure IV-18. Each case is shown at similar net cost values. All of the cases have similar results, with the most noticeable effect of modulation reduction in the cord DVH.

Achieving a mean dose of less than 26 Gy in the left parotid (the only one that could be spared) appeared to be the limiting factor in the optimization in *HN3*. This case shows that sometimes limiting the maximum intensity is not desirable. Figure IV-18(f) shows that higher intensities were necessary in areas near the left parotid (the upper areas of the beams) to create gradients in the targets to compensate for reducing the parotid dose. Therefore, maximum intensity limits could not be applied very strictly, or else the plans would become inferior. The remaining techniques were able to reduce the modulation while still meeting the plan objectives. The fact that the smoothing costlets were still utilizable in this complex geometry demonstrates one advantage of including modulation penalties as weighted costlets versus applying beamlet restrictions in an absolute sense. The PIMV_q penalized plan again had the highest MU reductions, followed by the SG2D and SG1D filter penalties. The latter techniques had slightly reduced normal tissue mean doses compared to the PIMV_q penalty, which can be seen in the DVHs in Figure IV-18.

IV.D. Discussion

A high degree of modulation in intensity modulated radiotherapy beams can lead to large increases in treatment time and monitor units (as compared to conventional 3DCRT plans), and planning and quality assurance time. None of these increases in time and effort is desirable. In this work, several techniques for reducing beam modulation were investigated, with the goal of not significantly altering IMRT plan quality or the ability of the optimizer to reach the prescribed inverse plan objectives. These techniques included (i) maximum intensity restrictions on beamlets, (ii) Savitzky-Golay filter-based

modulation penalties in 1-D and 2-D, and (iii) a modulation penalty based on the overall quadratic plan intensity map variation. Each of these techniques produced a fairly continuous range of plans and results at varying degrees of maximum intensity and modulation penalty weights. At lower maximum intensities and higher modulation penalty weights, there was a greater tradeoff between the low priority objectives (typically overall normal tissue dose minimization) and a greater decrease in MU. Relative MU reductions on the order of 13-52 % are possible in clinical cases with these complexity-reduction methods, while maintaining the normal tissue mean doses achieved with a baseline cost function. This provides the user with a choice of the degree of normal tissue dose increase acceptable for a certain increase in plan efficiency. Each of the modulation penalties could be a powerful tool for a dosimetrist, physicist, or physician to manipulate an IMRT plan on a case-by-case basis, depending on the specific goals of the plan.

While it is always preferred to decrease dose wherever possible, the clinical importance of changes in the low dose region is difficult to judge. Considering the additional transmission and leakage dose delivered to the patient during more modulated and complex deliveries, the advantages of using unconstrained intensity limits could be diminished, or more importantly, outweighed, by the increased normal tissue dose due to leakage/transmission. The average transmission is machine-dependent and is approximately 2 % of the total monitor units for our linear accelerator and MLC design (Varian, 2100 EX, Millennium MLC, Palo Alto, CA)¹⁹. These transmission rates could result in an average increase in dose from the constrained to unconstrained plans of approximately 5.0 cGy per fraction for static delivery. As Mohan *et al.* suggest, in complex deliveries it is not uncommon for some points to receive 100% of their dose through indirect means². Thus, the extra effort to reduce normal tissue doses by making minor intensity adjustments that increase the total monitor units are likely unproductive, and may even increase the total dose eventually received when more accurately accounting for transmission and leakage and geometric uncertainties. Hall *et al.* have suggested that this increased leakage radiation may contribute to an increased risk of second malignancies, and a joint publication by the American Society for Therapeutic Radiation Oncology (ASTRO) and the American Association of Physicists in Medicine (AAPM) has also pointed out the compromises that must be made when considering the increases in MU frequently seen in clinical IMRT²⁰⁻²².

Applying maximum intensity limits to beamlet plans is a simple solution to reduce MU in plans that do not require high degrees of freedom to meet the plan objectives. Thus, this method was successful in both the brain and prostate body sites. However, in the more complicated head/neck cases, limitations were encountered when certain objectives were on the border of being violated. In these cases, moderately restricting the maximum intensity interfered with the optimizer's ability to meet the plan objectives, thus requiring a more sophisticated smoothing mechanism to achieve higher MU reductions.

The first modulation penalty was based on a 1D Savitzky-Golay filter (SG1D penalty) applied only in the direction of the MLC travel, while the second penalty was based on a 2D filter (SG2D penalty). The third modulation costlet penalized the sum of squared differences between all neighboring beamlets ($PIMV_q$ penalty). Each of these penalties, when applied with a reasonable weight in a weighted sum cost function, successfully reduced plan modulation, while still allowing the optimizer ample freedom to meet the plan objectives. An advantage of using these penalties inside the cost function is that tradeoffs with normal structure doses are controlled by the weight of the modulation penalty. This would be impossible if a smoothing procedure were applied iteratively or post-optimization. The SG1D penalty has been used previously and shown to be effective compared to applying smoothing operations after each optimization iteration⁴. While this method can effectively reduce modulation and MU in most cases, its behavior can be somewhat unpredictable at increased weights of the penalty. Also, it (i) only decreases modulation in one direction, creating a step-like intensity pattern in the direction perpendicular to leaf travel, which is non-intuitive when compared to patient anatomy, (ii) could introduce unwanted errors in QA and delivery, and (iii) could potentially increase a plan's sensitivity to geometric shifts in the direction perpendicular to MLC travel. These potentially negative features can be useful in cases where smaller structures are present in the field (such as the penile bulb in the prostate and the spinal cord in the head/neck). Still, in the majority of the cases, use of the SG2D penalty will be superior in terms of overall beam modulation reduction and MU reduction, while producing plans with similar DVHs and dose metrics. In most cases, a 2D modulation penalty should be applied to reduce potential problems with QA and delivery.

Finally, the $PIMV_q$ penalty most consistently reduced MU as the weight was increased. This method eliminates the need for a separate computation of the smoothed

filtered beam, as it simply penalizes a direct measure of the overall beam modulation. The behavior of this penalty is very predictable, although a potential problem in this method is that it penalizes all modulation, whereas the SG penalties only penalize deviation from a smoothed version of the plan. In the head/neck plans, where the spinal cord frequently lies in the beam, the beam gradient is reduced as a result of the penalty, and the mean dose to the spinal cord is increased. In cases such as this, the Savitzky-Golay methods may be best-suited, or changes to the cost function may need to be made when using the PIMV-based penalty. Also, it may be advantageous to examine the use of “smarter” smoothing mechanisms that can distinguish between areas that should or should not be smoothed. Such a method has been investigated by Llacer *et al.*, who showed that for a simple 2D test case, a space-variant filter can be more effective at smoothing without compromising PTV coverage than other conventional filtering techniques⁵. Another method based on diffusion principles for 3D inverse planning that allows for spatially adaptive smoothing is discussed in Chapters V and VI.

A competing method for improving plan delivery accuracy and efficiency is the optimization of the size and shape of the segments to be used in IMRT delivery. A fair comparison of direct segment optimization (DSO)^{23,24} to fluence map optimization with smoothing and improved leaf sequencing is warranted to determine which produces the best dosimetric results in complicated cases, while aiming to improve overall planning and delivery efficiency.

Several authors have reported an increase in planning efficiency when incorporating smoothing into the cost function in contrast to smoothing at the end of each iteration^{4,25}. This trend has been observed in this work and is likely due to an increase in curvature of the cost function near the global minimum, reducing the time to fine-tune the plan after convergence.

Another potential advantage to reducing beam complexity is a decrease in delivery time. In fact, delivery time reductions on the order of 25 % in the prostate cases studied here were observed, but a significant reduction in delivery time was not seen in the other body sites. It may be that the effective leaf travel is not substantially reduced until the beam is almost flat. More analysis of the delivery time implications of the PIMVq costlet is shown in Chapter VII.

One disadvantage of incorporating smoothing into the cost function is a difficulty in analyzing the tradeoffs between the smoothness criteria and the target and normal

tissue objectives. However, this problem remains for assessing all weighted-sum cost function plans, and the proposed Pareto surface navigation could be applied to analyze the tradeoffs in question²⁶. It is also possible to use a modulation penalty as a priority stage in a Lexicographic ordering optimization²⁷. This idea is discussed further in Chapter VI. Nevertheless, it currently remains a case-by-case situation to properly incorporate smoothing into a cost function, and this general approach would most likely include iterating through several importance factors for smoothness before reaching the desired solution. The resulting trial-and-error process may offset the decreased time needed for optimization, but the ultimate increase in plan quality based upon improved QA and delivery efficiency should prove beneficial.

References

- ¹ L. Ma, "Smoothing intensity-modulated treatment delivery under hardware constraints," *Med. Phys.* **29**, 2937-2295 (2002).
- ² R. Mohan, M. Arnfield, S. Tong, Q. Wu, and J. Siebers, "The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy," *Med. Phys.* **27**, 1226-1237 (2000).
- ³ X. Sun and P. Xia, "A new smoothing procedure to reduce delivery segments for static MLC-based IMRT planning," *Med. Phys.* **43**, 2785-2794 (2004).
- ⁴ S. Spirou, N. Fournier-Bidoz, J. Yang, C. Chui, and C. Ling, "Smoothing intensity-modulated beam profiles to improve the efficiency of delivery," *Med. Phys.* **28**, 2105-2112 (2001).
- ⁵ J. Llacer, N. Agazaryan, T. D. Solberg, and C. Promberger, "Degeneracy, frequency response and filtering in IMRT optimization," *Phys. Med. Biol.* **49**, 2853-2880 (2004).
- ⁶ B. A. Fraass and D. L. McShan, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F.H. van der Giessen, H.J. van Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 273-276.
- ⁷ B. A. Fraass, D. L. McShan, and K. J. Weeks, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F H van_der_Giessen, H J van_Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 193-196.
- ⁸ B. A. Fraass, D. L. McShan, R. K. TenHaken, and K. M. Hutchins, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F H van_der_Giessen, H J van_Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 521-525.
- ⁹ D. L. McShan and B. A. Fraass, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F H van_der_Giessen, H J van_Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 41-44.
- ¹⁰ D. L. McShan, B. A. Fraass, and A. S. Lichter, "Full integration of the beam's eye view concept into computerized treatment planning," *Int J Radiat Oncol Biol Phys* **18**, 1485-1494 (1990).
- ¹¹ M. L. Kessler, D. L. McShan, M. Epelman, K. A. Vineberg, A. Eisbruch, T. S. Lawrence, and B. A. Fraass, "Costlets: a generalized approach to cost functions for automated optimization of IMRT treatment plans," *Optimization and Engineering* **6**, 421-448 (2005).
- ¹² J. H. Kim, N. Dogan, D. L. McShan, and M. L. Kessler, "An AVS-based system for optimization of conformal radiotherapy treatment plans," *Proceedings of 1995 International Advanced Visual Systems User and Developer Conference*, Boston MA, 1995.
- ¹³ M. M. Coselmon, J. M. Moran, J. Radawski, and B. A. Fraass, "Improving IMRT delivery efficiency by applying intensity limits during inverse planning," *Med. Phys.* **32**, 1234-1245 (2005).
- ¹⁴ A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry* **36**, 1627-1639 (1964).

- ¹⁵ A. Griewank, D. Juedes, H. Mitev, J. Utke, O. Vogel, and A. Walther, "ADOL-C: A Package for the Automatic Differentiation of Algorithms Written in C/C++," *ACM TOMS* **22**, 131-167 (1996).
- ¹⁶ K.-W. Jee, B. A. Fraass, and D. L. McShan, in *Automatic Differentiation: Applications, Theory, and Tools*, edited by H. M. Bucker, G. Corliss, P. Hovland et al. (Springer, 2005).
- ¹⁷ T. R. Mackie, J. W. Scrimger, and J. J. Battista, "A convolution method of calculating dose for 15-MV x rays," *Med. Phys.* **12**, 188-196 (1985).
- ¹⁸ T. R. Bortfeld, D. L. Kahler, T. J. Waldron, and A. L. Boyer, "X-ray field compensation with multileaf collimators," *Int J Radiat Oncol Biol Phys* **28**, 723-730 (1994).
- ¹⁹ T. LoSasso, C. S. Chui, and C. C. Ling, "Physical and dosimetric aspects of a multileaf collimation system used in the dynamic mode for implementing intensity modulated radiotherapy," *Med Phys* **25**, 1919-1927 (1998).
- ²⁰ J. M. Galvin, G. Ezzell, A. Eisbruch, C. Yu, B. Butler, Y. Xiao, I. Rosen, J. Rosenman, M. Sharpe, L. Xing, P. Xia, T. Lomax, D. A. Low, and J. Palta, "Implementing IMRT in clinical practice: a joint document of the American Society for Therapeutic Radiology and Oncology and the American Association of Physicists in Medicine," *Int J Radiat Oncol Biol Phys* **58**, 1616-1634 (2004).
- ²¹ E. J. Hall, "Intensity-modulated radiation therapy, protons, and the risk of second cancers," *Int J Radiat Oncol Biol Phys* **65**, 1-7 (2006).
- ²² E. J. Hall and C. Wu, "Radiation induced second cancers: The impact of 3D-CRT and IMRT," *Int J Radiat Oncol Biol Phys* **56**, 83-88 (2003).
- ²³ B. A. Fraass, D. L. McShan, and M. L. Kessler, "Dose-base conformal field shaping using automated optimization," *Proceedings of XIIIth International Conference on the Use of Computers in Radiotherapy*, Heidelberg, Germany, 2000.
- ²⁴ B. A. Fraass, J. M. Moran, K. A. Vineberg, R. Marsh, and D. L. McShan, "Automated optimization of 3DCRT and multisegment IMRT plans," *Proceedings of ESTRO*, Amsterdam, the Netherlands, 2004.
- ²⁵ M. Alber and F. Nusslin, "Intensity modulated photon beams subject to a minimal surface smoothing constraint," *Phys. Med. Biol.* **45**, N49-N52 (2000).
- ²⁶ D. Craft, T. Halabi, and T. Bortfeld, "Exploration of tradeoffs in intensity-modulated radiotherapy," *Phys Med Biol* **50**, 5857-5868 (2005).
- ²⁷ K.-W. Jee, D. L. McShan, and B. A. Fraass, "Intuitive multicriteria IMRT optimization using a lexicographic approach," *Med. Phys.* **31**, 1715 (2004).

CHAPTER V.

ADAPTIVE DIFFUSION SMOOTHING

Chapter IV presented several different types of smoothing filters and modulation penalties that can be used during the optimization process to promoting smoothing of IMRT fields in an effort to reduce plan complexity. While these methods were successful, there is still a concern that smoothing filters and modulation penalties cannot distinguish between desirable and undesirable modulation and therefore, all modulation gets penalized to the same degree in the cost function. This can lead to penalization of important gradients and result in sub-optimal smoothing and possible degradation of the plan in areas where high intensity gradients are necessary. This chapter will introduce another smoothing method that we have developed using an analogy to various processes described by the diffusion equation. This new method, which we have termed *adaptive diffusion smoothing*, or ADS, uses beamlet-specific diffusion coefficients to automatically customize the amount of smoothing that occurs in different regions of the beam. We will describe the theory and implementation of this method for use in IMRT treatment planning and present several clinical examples of its use. Chapter VI will present several further applications of this method.

V.A. Motivation

As we have shown in previous chapters, considerable effort has been spent trying to reduce the complexity of IMRT fluence patterns while preserving the advantages gained by employing IMRT. Methods that have been considered include beamlet restrictions¹, direct optimization of delivery segments²⁻⁴, and smoothing of intensity modulated beams either during or after optimization⁵⁻¹⁰. As the parameters in each of these methods are relaxed, the unconstrained beamlet solution is approached. The goal is to restrain the parameters enough so that the solution is an acceptable compromise between the ideal beamlet solution and a more efficient conventional 3D conformal solution. Unfortunately,

these previously-described smoothing methods may not be optimal because they cannot distinguish between parts of the beam that should and should not be smoothed.

So far, the most promising smoothing methods are those that penalize or account for modulation inside the inverse planning cost function. Examples of these include the smoothing costlets presented in Chapter IV. There, we showed that significant increases in efficiency are possible while preserving the full dosimetric quality of the original (unsmoothed) IMRT plan. However, the problem remains that the beam is either penalized for being modulated, which is the fundamental feature of IMRT in the first place, or the beam is penalized for not conforming to a certain filtered version of itself, which may not be ideal for the specific beam in question. The latter was found to be true in several cases when using the Savitsky-Golay filter penalty during optimization. The amount of smoothing possible is always limited by the eventual tradeoffs that are made with target or normal tissue coverage, leading us to believe that a more adaptive smoothing procedure that can distinguish between areas that should or should not be smoothed could produce superior results.

A spatially adaptive smoothing method was investigated by Llacer *et al.* for a simple 2D test case, and was shown to compromise PTV coverage to a lesser degree than other conventional filtering techniques¹¹. In their work, beamlets around the PTV were manually selected to receive less filtering in order to preserve the intensity near the target. In the following sections, we present a new *adaptive diffusion smoothing* (ADS) method for preferential smoothing in 3D inverse planning. The ADS method, which makes simple use of the common diffusion equation, allows for preferential smoothing by using variable and automatically-defined diffusion coefficients. This method has the potential to distinguish between important and non-important areas of modulation, facilitating smarter tradeoffs between the cost function and smoothing criteria. In the following, we (i) describe the ADS theory, method and implementation for controlling IMRT beam complexity, (ii) characterize it using a test phantom optimization case, and (iii) demonstrate its utility and potential in clinical IMRT cases.

V.B. Methods

V.B.1. Adaptive diffusion smoothing formulation and implementation

The time-dependent diffusion equation, given in Equation (V.1), is found in many areas of science and engineering. This equation describes the propagation of a material from areas of high concentration to areas of low concentration according to a spatially variant

diffusion coefficient, $D(x, y)$, which depends on the local properties of the medium (this equation is cast in 2-D for application to the beamlet intensity situation).

$$\begin{aligned} \frac{\partial \phi}{\partial t}(x, y, t) &= \nabla \cdot D(x, y) \nabla \phi(x, y, t) \\ &= \frac{\partial}{\partial x} D(x, y) \frac{\partial \phi}{\partial x}(x, y, t) + \frac{\partial}{\partial y} D(x, y) \frac{\partial \phi}{\partial y}(x, y, t) . \end{aligned} \quad (V.1)$$

By replacing the density of diffusing material, $\phi(x, y, t)$, with beamlet intensity, $I(x, y, t)$, we can make an analogy from the diffusion of a material to the smoothing of an IMRT beam. Using this idea, we have designed a procedure, named “adaptive diffusion smoothing” or ADS, in which an intensity modulated beamlet grid is smoothed using Equation V.1

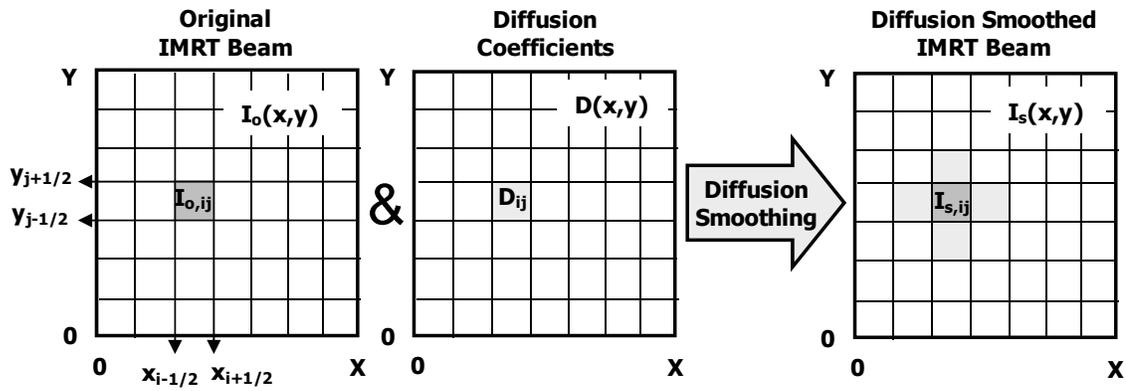


Figure V-1. A simplified illustration of the diffusion smoothing procedure and some of the notation used in this work. We start with an intensity modulation beam, I_o , and a grid of corresponding diffusion coefficients, D , that dictate the degree of the smoothing allowed or applied over each beamlet and then solve our diffusion equation analogy to get the diffusion smoothed beam, I_s .

with a diffusion coefficient $D(x, y)$ that is automatically defined for each beamlet. This allows the smoothing characteristics to adapt to each individual plan and allows for spatially variant smoothing. The diffusion process and some of the discrete notation used below is illustrated in Figure V-1. This figure is simplified for illustration purposes, as it only shows one beamlet, and the diffusion coefficients at the beamlet borders are actually a function of the neighboring diffusion coefficients.

To describe the ADS method, we consider the following time-dependent diffusion problem:

$$\frac{\partial I_s}{\partial t}(x, y, t) = \nabla \cdot D(x, y) \nabla I_s(x, y, t) \quad \text{for} \quad 0 < x < X, 0 < y < Y, \quad (\text{V.2})$$

with boundary conditions:

$$\begin{aligned} \frac{\partial I_s}{\partial x}(0, y, t) = \frac{\partial I_s}{\partial x}(X, y, t) = 0 \quad \text{for} \quad 0 < y < Y, \\ \frac{\partial I_s}{\partial y}(x, 0, t) = \frac{\partial I_s}{\partial y}(x, Y, t) = 0 \quad \text{for} \quad 0 < x < X, \end{aligned} \quad (\text{V.3a})$$

and initial condition:

$$I_s(x, y, 0) = I_0(x, y). \quad (\text{V.3b})$$

Here $I_0(x, y)$ is the unsmoothed intensity map, $I_s(x, y, t)$ is the smoothed intensity map at time t , and $D(x, y) \geq 0$ is the prescribed diffusion coefficient (discussed in detail below). For $t = 0$, $I_s(x, y, 0) = I_0(x, y)$ and no smoothing has taken place. For $D(x, y) > 0$, as t increases, $I_s(x, y, t)$ becomes increasingly smoothed. The combination of the size of the diffusion coefficient and Δt control the amount of smoothing that occurs. Also, integrating Equation V.2 over $0 \leq x \leq X$ and $0 \leq y \leq Y$ and using the boundary conditions expressed in Equation V.3a, we easily obtain for all $t \geq 0$

$$\int_0^Y \int_0^X I_s(x, y, t) dx dy = \int_0^Y \int_0^X I_0(x, y) dx dy. \quad (\text{V.4})$$

Thus, the total intensity of the unsmoothed beam I_0 is automatically preserved.

To estimate the smoothed beam at time $\Delta t > 0$, we integrate Equation V.2 over $0 \leq t \leq \Delta t$ to obtain:

$$I_s(x, y, \Delta t) = I_0(x, y) + \nabla \cdot D \nabla \int_0^{\Delta t} I_s(x, y, t) dt . \quad (\text{V.5})$$

To evaluate the integral term, we consider the implicit and explicit time-differencing approximations. Using the implicit approximation, we obtain the following integral term:

$$\begin{aligned} \int_0^{\Delta t} I_s(x, y, t) dt &\approx \int_0^{\Delta t} I_s(x, y, \Delta t) dt \\ &= \Delta t I_s(x, y, \Delta t) \\ &= \Delta t I_s(x, y) . \end{aligned} \quad (\text{V.6})$$

Using the explicit approximation, we obtain instead

$$\begin{aligned} \int_0^{\Delta t} I_s(x, y, t) dt &\approx \int_0^{\Delta t} I_s(x, y, 0) dt \\ &= \Delta t I_0(x, y) . \end{aligned} \quad (\text{V.7})$$

Introducing the implicit approximation, Equation V.6, into Equation V.5, we get:

$$I_s(x, y) = I_0(x, y) + \nabla \cdot w D(x, y) \nabla I_s(x, y) , \quad (\text{V.8})$$

where $w = \Delta t$. In Equation V.8, the smoothed beamlet intensity depends on the smoothed beam as well as the original beam, making an iteration scheme necessary to solve for I_s . On the other hand, using the explicit approximation, Equation V.7, in Equation V.5, we obtain:

$$\begin{aligned} I_s(x, y) &= I_0(x, y) + \nabla \cdot w D \nabla I_0(x, y) . \\ &\text{where } w = \Delta t \end{aligned} \quad (\text{V.9})$$

Here, I_s depends only on the original beamlet intensities, which results in a quicker and simpler computation of the smoothed beam. In order to apply Equation V.9 to the beamlet geometry, we must consider the discrete situation in which $I(x, y)$, and $D(x, y)$ are represented by a series of beamlet intensities I_{ij} , and diffusion coefficients D_{ij} , as shown

in Figure V-1. The length of each square beamlet is given by $x_{i+1/2} - x_{i-1/2} = h$. First, we apply the gradient operator and write Equation V.9 as

$$I_s(x, y) = I_0(x, y) + \frac{\partial}{\partial x} \left(wD(x, y) \frac{\partial I_0(x, y)}{\partial x} \right) + \frac{\partial}{\partial y} \left(wD(x, y) \frac{\partial I_0(x, y)}{\partial y} \right). \quad (\text{V.10})$$

Next, to simplify the notations, we introduce, $J_x(x, y)$ and $J_y(x, y)$ as

$$\begin{aligned} J_x(x, y) &= wD(x, y) \frac{\partial I_0(x, y)}{\partial x} \quad \text{and} \\ J_y(x, y) &= wD(x, y) \frac{\partial I_0(x, y)}{\partial y}. \end{aligned} \quad (\text{V.11})$$

Then, we write Equation V.10 as the first order system

$$I_s(x, y) = I_0(x, y) + \frac{\partial}{\partial x} J_x(x, y) + \frac{\partial}{\partial y} J_y(x, y) \quad (\text{V.12})$$

with boundary conditions

$$\begin{aligned} J_x(0, y) = J_x(X, y) &= 0, \quad 0 \leq y \leq Y \\ J_y(x, 0) = J_y(x, Y) &= 0, \quad 0 \leq x \leq X \end{aligned} \quad (\text{V.13})$$

Next, we integrate over the cell *at the* (i, j) position:

$$\begin{aligned} \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{i-1/2}}^{y_{i+1/2}} I_s(x, y) dx dy &= \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{i-1/2}}^{y_{i+1/2}} I_0(x, y) dx dy + \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{i-1/2}}^{y_{i+1/2}} \frac{\partial}{\partial x} J_x(x, y) dx dy + \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{i-1/2}}^{y_{i+1/2}} \frac{\partial}{\partial y} J_y(x, y) dx dy \\ \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{i-1/2}}^{y_{i+1/2}} I_s(x, y) dx dy &= \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{i-1/2}}^{y_{i+1/2}} I_0(x, y) dx dy + \int_{y_{i-1/2}}^{y_{i+1/2}} J_x(x_{i+1/2}, y) dy - \int_{y_{i-1/2}}^{y_{i+1/2}} J_x(x_{i-1/2}, y) dy + \dots \\ &\quad + \int_{x_{i-1/2}}^{x_{i+1/2}} J_y(x, y_{i+1/2}) dx - \int_{x_{i-1/2}}^{x_{i+1/2}} J_y(x, y_{i-1/2}) dx \end{aligned} \quad (\text{V.14})$$

Using the discrete notation, we can rewrite the first two terms as

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{i-1/2}}^{y_{i+1/2}} I_s(x,y) dx dy = \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{i-1/2}}^{y_{i+1/2}} I_{s,ij} dx dy = h^2 I_{s,ij} \text{ and} \quad (V.15)$$

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{i-1/2}}^{y_{i+1/2}} I_0(x,y) dx dy = \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{i-1/2}}^{y_{i+1/2}} I_{0,ij} dx dy = h^2 I_{0,ij}$$

and we can also represent the third term as

$$\int_{y_{i-1/2}}^{y_{i+1/2}} J_x(x_{i+1/2}, y) dy \equiv J_{x,i+1/2,j} \quad (V.16)$$

Similar notation applies to the remaining terms in Equation V.14. If $x_{i+1/2}$ lies on an outer boundary of R, then by the boundary conditions given in Equation V.13, $J_{x,i+1/2,j}=0$. Therefore, we can write:

$$J_{x,i+1/2,j} = J_{x,i+1/2,j} = 0, \quad i \leq j \leq J. \quad (V.17)$$

However, if the integration in Equation V.16 is over an interior cell, we must perform the integration in Equation V.16. First we insert Equation V.11 to obtain:

$$J_{x,i+1/2,j} = \int_{y_{i-1/2}}^{y_{i+1/2}} wD(x_{i+1/2}, y) \frac{\partial}{\partial x} I_0(x_{i+1/2}, y) dy. \quad (V.18)$$

This introduces an unknown variable, $I_0(x_{i+1/2}, y)$, which is the original beam intensity on an interior cell edge. To solve Equation V.18, we express it in terms of two one-sided finite differences which are required to be equal to each other:

$$J_{x,i+1/2,j} = \int_{y_{i-1/2}}^{y_{i+1/2}} w D_{i+1,j} \frac{I_{0,i+1,j} - I_{0,i+1/2,j}}{h/2} dy = \int_{y_{i-1/2}}^{y_{i+1/2}} w D_{ij} \frac{I_{0,i+1/2,j} - I_{0,ij}}{h/2} dy \quad (a)$$

$$hw D_{i+1,j} \frac{I_{0,i+1,j} - I_{0,i+1/2,j}}{h/2} = hw D_{ij} \frac{I_{0,i+1/2,j} - I_{0,ij}}{h/2} \quad (b) \quad (V.19)$$

$$D_{i+1,j} (I_{0,i+1,j} - I_{0,i+1/2,j}) = D_{ij} (I_{0,i+1/2,j} - I_{0,ij}) \quad (c)$$

Now, we solve (c) for the unknown value $I_0(x_{i+1/2}, y)$:

$$\begin{aligned} D_{i+1,j} I_{0,i+1,j} - D_{i+1,j} I_{0,i+1/2,j} &= D_{ij} I_{0,i+1/2,j} - D_{ij} I_{0,ij} \\ I_{0,i+1/2,j} &= \frac{D_{i+1,j} I_{0,i+1,j} + D_{ij} I_{0,ij}}{D_{ij} + D_{i+1,j}} \end{aligned} \quad (V.20)$$

Next, we insert Equation V.20 into the left side of Equation V.19(b) to solve for $J_{x,i+1/2,j}$:

$$\begin{aligned} J_{x,i+1/2,j} &= hw D_{i+1,j} \frac{I_{0,i+1,j} - I_{0,i+1/2,j}}{h/2} \\ &= hw D_{i+1,j} \frac{I_{0,i+1,j} - \frac{D_{i+1,j} I_{0,i+1,j} + D_{ij} I_{0,ij}}{D_{ij} + D_{i+1,j}}}{h/2} \\ &= 2w D_{i+1,j} \left(I_{0,i+1,j} - \frac{D_{i+1,j} I_{0,i+1,j} + D_{ij} I_{0,ij}}{D_{ij} + D_{i+1,j}} \right) \\ &= 2w D_{i+1,j} \left(\frac{I_{0,i+1,j} (D_{ij} + D_{i+1,j}) - D_{i+1,j} I_{0,i+1,j} - D_{ij} I_{0,ij}}{D_{ij} + D_{i+1,j}} \right) \\ &= 2w D_{i+1,j} \left(\frac{D_{ij} I_{0,i+1,j} - D_{ij} I_{0,ij}}{D_{ij} + D_{i+1,j}} \right) \\ &= w (I_{0,i+1,j} - I_{0,ij}) \left(\frac{2D_{ij} D_{i+1,j}}{D_{ij} + D_{i+1,j}} \right) \\ J_{x,i+1/2,j} &= w (I_{0,i+1,j} - I_{0,ij}) D_{i+1/2,j}, \text{ where } D_{i+1/2,j} = \left(\frac{2D_{ij} D_{i+1,j}}{D_{ij} + D_{i+1,j}} \right) \end{aligned} \quad (V.21)$$

Using the same methods, we also solve for the remaining J terms in Equation V.14. To summarize the final four terms as they apply to interior cell edges, we have:

$$\begin{aligned}
J_{x,i+1/2,j} &= w(I_{0,i+1,j} - I_{0,ij})D_{i+1/2,j} \\
J_{x,i-1/2,j} &= w(I_{0,ij} - I_{0,i-1,j})D_{i-1/2,j} \\
J_{y,i,j+1/2} &= w(I_{0,i,j+1} - I_{0,ij})D_{i,j+1/2} \\
J_{y,i,j-1/2} &= w(I_{0,ij} - I_{0,i,j-1})D_{i,j-1/2}
\end{aligned} \tag{V.22}$$

If we combine the boundary conditions given in Equation V.13 with the D values above, we can allow Equation V.22 to apply to both interior and boundary cells. Thus, we have:

$$\begin{aligned}
D_{i+1/2,j} &= \begin{cases} \left(\frac{2D_{ij}D_{i+1,j}}{D_{ij} + D_{i+1,j}} \right) & \forall 1 \leq i \leq I-1, 1 \leq j \leq J \\ 0 & \forall i = 0 \text{ or } I, 1 \leq j \leq J \end{cases} \\
D_{i,j+1/2} &= \begin{cases} \left(\frac{2D_{ij}D_{i,j+1}}{D_{ij} + D_{i,j+1}} \right) & \forall 1 \leq j \leq J-1, 1 \leq i \leq I \\ 0 & \forall j = 0 \text{ or } J, 1 \leq i \leq I \end{cases}
\end{aligned} \tag{V.23}$$

Therefore, we combine Equations V.15, 22 and 23 to obtain the discrete version of Equation V.9:

$$\begin{aligned}
I_{s,ij}h^2 &= I_{0,ij}h^2 + J_{x,i+1/2,j} - J_{x,i-1/2,j} + J_{y,i,j+1/2} - J_{y,i,j-1/2} \\
&= I_{0,ij}h^2 + w(I_{0,i+1,j} - I_{0,ij})D_{i+1/2,j} - w(I_{0,ij} - I_{0,i-1,j})D_{i-1/2,j} + \dots \\
&\quad w(I_{0,i,j+1} - I_{0,ij})D_{i,j+1/2} - w(I_{0,ij} - I_{0,i,j-1})D_{i,j-1/2}
\end{aligned} \tag{V.24}$$

Finally, we solve for $I_{s,ij}$:

$$\begin{aligned}
I_{s,ij} &= I_{0,ij} + \frac{w}{h^2} \left[\begin{aligned} &(I_{0,i+1,j} - I_{0,ij})D_{i+1/2,j} - (I_{0,ij} - I_{0,i-1,j})D_{i-1/2,j} + \dots \\ &(I_{0,i,j+1} - I_{0,ij})D_{i,j+1/2} - (I_{0,ij} - I_{0,i,j-1})D_{i,j-1/2} \end{aligned} \right] \\
&= I_{0,ij} + \frac{w}{h^2} \left[\begin{aligned} &D_{i+1/2,j}I_{0,i+1,j} - D_{i+1/2,j}I_{0,ij} - D_{i-1/2,j}I_{0,ij} + D_{i-1/2,j}I_{0,i-1,j} + \dots \\ &D_{i,j+1/2}I_{0,i,j+1} - D_{i,j+1/2}I_{0,ij} - D_{i,j-1/2}I_{0,ij} + D_{i,j-1/2}I_{0,i,j-1} \end{aligned} \right] \quad (V.25) \\
&= I_{0,ij} \left[1 - \frac{w}{h^2} (D_{i+1/2,j} + D_{i-1/2,j} + D_{i,j+1/2} + D_{i,j-1/2}) \right] + \dots \\
&\quad \frac{w}{h^2} (D_{i+1/2,j}I_{0,i+1,j} + D_{i-1/2,j}I_{0,i-1,j} + D_{i,j+1/2}I_{0,i,j+1} + D_{i,j-1/2}I_{0,i,j-1}).
\end{aligned}$$

Equation V.25 represents the classic cell-average spatial discretization¹² of the diffusion problem, using the explicit time differencing approximation. As Equation V.25 shows, the smoothed intensities are based explicitly on the original intensities, making the smoothed intensity pattern easy to calculate. Also, this solution can be applied inside of the optimization procedure as a penalized cost function component to be used with a gradient-based optimization scheme. While it would be possible to use a discretization of the implicit version inside the cost function, the cost function gradient with respect to the original intensities cannot be calculated, making it necessary to employ a non-gradient based (and likely much slower) optimization algorithm, such as simulated annealing. For these reasons, we employ the explicit discretization of Equation V.25 in the remainder of this work.

V.B.2. Diffusion Smoothing Properties

The explicit diffusion smoothing scheme has several interesting features (many of which are shared with the implicit solution), including:

1. If $I_o = \text{uniform intensity}$, then $I_s = I_o$. Thus, ADS does not alter a flat field. This is easily shown by setting all values of I_o in Equation V.25 to be equal values, I_c :

$$I_{s,ij} = I_c \left[\begin{aligned} &1 - \frac{w}{h^2} (D_{i+1/2,j} + D_{i-1/2,j} + D_{i,j+1/2} + D_{i,j-1/2}) + \dots \\ &\frac{w}{h^2} (D_{i+1/2,j} + D_{i-1/2,j} + D_{i,j+1/2} + D_{i,j-1/2}) \end{aligned} \right] = I_c$$

2. If $\frac{w}{h^2} \left(D_{i+\frac{1}{2},j} + D_{i-\frac{1}{2},j} + D_{i,j+\frac{1}{2}} + D_{i,j-\frac{1}{2}} \right) \leq 1$ for all ij , then $I_{o,\min} \leq I_s(x,y) \leq I_{o,\max}$.

Thus, if the first inequality is true, the maximum and minimum intensities of the smoothed beam will lie between the maximum and minimum intensities of the unsmoothed beam. (For the implicit solution, this property is always true.) This property ensures that as long as the original beamlet intensity is positive, the smoothed beamlet intensity will also be positive. This is an important property, since the preservation of a positive beamlet intensity is essential. In addition, this property gives us a guideline for choosing the absolute values of the time-step w and diffusion coefficients. As mentioned previously, the combination of D and w control the amount of smoothing that occurs in the diffusion smoothing process. To achieve maximum smoothing, the inequality above should be close to 1. Since many beamlet intensity values are in the range of 0 to 10 or above, we have chosen to limit D to the range between 0 and 10 as well for the sake of having some intuition in choosing D . Therefore, in order to satisfy the above inequality, and to provide the possibility for a relatively high degree of smoothing, we chose $w = 0.02h^2$, where h is the beamlet dimension (usually 0.5 cm or 1.0 cm) for the remainder of this work. Limiting D between 0 and 10 allows us to keep w fixed while still providing a wide range of smoothing capabilities.

3. $\sum_{\text{all beamlets}} I_o = \sum_{\text{all beamlets}} I_s$, or the smoothed beam preserves the total intensity of

the unsmoothed beam. [The discretization scheme preserves Equation V.4.] This property ensures that the smoothed plan will not be drastically different from the original plan and that the contributing beams will be consistent. In some cases, this feature could be considered restrictive, since the ability to shift dose contributions from one beam to another could be important to achieve the most optimal plan. However, in the iterative ADS penalty smoothing scheme discussed below, a transition of intensity contributions from one beam to another is still possible and will occur if it is beneficial to the cost function. If a large intensity shift is required, this feature could potentially slow the optimization time to convergence.

4. Because of the properties of the diffusion operator, the ADS process preferentially suppresses high frequency components of I_o when D is uniform.

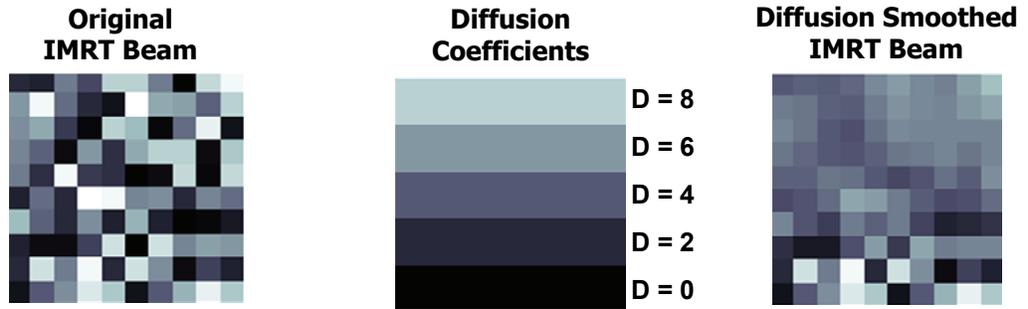


Figure V-2. An example of the increase in smoothing that occurs as the diffusion coefficient value increases from the bottom to the top of the field.

This is an important feature, because, unlike other smoothing techniques such as polynomial fitting, ADS ensures that the smoothed beam is actually a less modulated version of the original beam. No unwanted high frequency components can arise as artifacts of the ADS smoothing process. This can be explicitly proven for the original diffusion problem in Equation VII.1, but can be more easily inferred by examining Equation V.25. If we examine our diffusion smoothing operator, L , which operates on a function f , we have:

$$I_{s,ij} = Lf(x_i, y_j) = f(x_i, y_j) \left[1 - 4 \frac{wD}{h^2} \right] + \dots$$

$$\frac{wD}{h^2} \left[f(x_{i+1}, y_j) + f(x_{i-1}, y_j) + f(x_i, y_{j+1}) + f(x_i, y_{j-1}) \right]$$

The least oscillatory motion possible on the grid would be when $f(x,y)$ equals some constant c over the beam. In that situation, we have $Lf(x_i, y_j) = f(x_i, y_j) = c$. The other extreme on the grid would be when $f(x_i, y_j)$ oscillates between -1 and 1 over each grid point, or $f(x_i, y_j) = c(-1)^{i+j}$. In this situation, we have $Lf(x_i, y_j) = c \left[1 - 8 \frac{wD}{h^2} \right] (-1)^{i+j} = \left[1 - 8 \frac{wD}{h^2} \right] f(x_i, y_j)$ where $4wD/h^2 <$

1. Thus, when we apply the diffusion smoothing operator to a beam, or function f , we will obtain a less extreme function.

5. If $D_{ij}=0$ in beamlet ij , then $I_{s,ij}=I_{o,ij}$; if D_{ij} is large, $I_{s,ij}$ is strongly smoothed. This gives us a basis on which to define D according to the amount of smoothing desired in a certain beamlet. Figure V-2 illustrates the smoothing that occurs

when starting from a random intensity beamlet pattern and using a stepwise diffusion coefficient. The increase in smoothing can be clearly seen as the diffusion coefficient increases from the bottom to the top of the field.

6. The explicitly smoothed $I_{s,ij}$ is determined by the original plan I_o only at the ij beamlet and its four nearest neighbors. The implicit solution smoothes over the entire beam. Thus, in addition to being simpler to calculate, the explicit solution also leads to more local smoothing, which is most likely more desirable for IMRT beams. However, if a more globally smoothed beam is desired, exploration of the implicit method would certainly be warranted. (Alternatively, one could apply the explicit diffusion smoothing procedure multiple times.)

V.B.3. Diffusion Coefficients

We believe that the full power of the diffusion smoothing procedure lies in the definition of D , the diffusion coefficient. D can be defined in a multitude of ways, the only constraint being $D_{ij} \geq 0$ for all beamlets. We know from the above properties that if $D_{ij} = 0$ for a certain beamlet then the smoothed intensity of that beamlet will equal the original intensity. Conversely, if D_{ij} is large for a beamlet, then there will be a large amount of smoothing between the original and smoothed beamlet intensities. This gives the user a high degree of control in the amount of smoothing applied over the field and allows for spatially variable amounts of smoothing.

Smoothing procedures that are applied as a cost function penalty appear to have the fewest drawbacks in terms of plan degradation compared to those applied outside the cost function^{6,8,13}. Therefore, we chose to characterize adaptive diffusion smoothing for use inside the cost function. From the previous paragraph, we can specify D so that modulation is penalized strongly or weakly in given areas of the beam. For example, if we would like to keep sharp edges at the target boundary, we can choose D to have low or zero components near the edges of the beamlet intensity map. Similarly, if we do not wish to penalize a large intensity gradient over a target/normal tissue interface, we can choose D to be small over those beamlets. Because there are practically no restrictions on the definition of the coefficient values, there are many possible ways to prescribe D . As a first characterization of the ADS method, we will consider two logical ways to define the diffusion coefficients:

V.B.3.1. Uniform

In IMRT cases with simple geometries, there is generally a large solution space of acceptable plans, and the cost function is not sensitive to small changes in the beamlet intensity patterns. In these situations, the simplest way to define D is as a uniform value across the field. Although this definition of D does not exploit the full potential of defining individual diffusion coefficients (and thus does not distinguish between desirable and undesirable modulations), it provides a good starting point to test the diffusion smoothing scheme and assess whether customized diffusion smoothing coefficients are necessary for simple plans.

V.B.3.2. Gradient

The ultimate goal of smoothing an IMRT intensity pattern should be to maximally smooth the field with the minimum change in cost function value, which translates to the minimum negative effect on the inverse plan dose prescription. In a gradient-based optimization method, the partial derivatives of the cost function (CF) with respect each of the beamlet intensities (I_{ij}), $\partial CF/\partial I_{ij}$, must be calculated at every iteration. Each of these partial derivatives describes how important each beamlet value is to minimizing the cost function value. Beamlets with large absolute values of the gradient have the property that altering those beamlets would have a large effect on the cost function, and vice versa. We have observed that these gradients, at convergence, can vary by several orders of magnitude, with a small percentage of the gradients having very high values, many having very small values and a few having moderate values. With this information, we can define D to be a function of these gradient values to achieve little to no smoothing in high gradient beamlets, moderate smoothing in moderate gradient beamlets, and a high degree of smoothing in low gradient beamlets. We have studied a variety of possible formulations for D , including making D_{ij} inversely proportional to $|\partial CF/\partial I_{ij}|^n$. In order to achieve the desired distribution of D , we have concluded that the following function is a robust and tunable formulation for D that can be used in most, if not all, IMRT cases:

$$D_{ij} = \frac{10}{1 + c \left(\frac{|\partial CF/\partial I_{0,ij}|}{s} \right)^n} . \quad (\text{V.26})$$

Here, s is a gradient scaling factor that is equal to the median value of $|\partial CF/\partial_{ij}|$. The median value was chosen over the mean because of the tendency for some gradients to have extremely high values and bias the mean towards the higher end. The parameters c and n are tunable parameters that allow D to be further customized, depending on the individual case. It is likely that similar plans will have similar optimal values of c and n . The parameter c serves to shift the transition from high to low D and n controls the steepness of the fall-off from high to low D . Figure V-3 shows the behavior of Equation V.26 at several different c and n values for the converged gradients in a prostate cancer IMRT case. For equal c values, all curves will intersect when the gradient is equal to the median value. The large range of the gradient values over several orders of magnitude should also be noted.

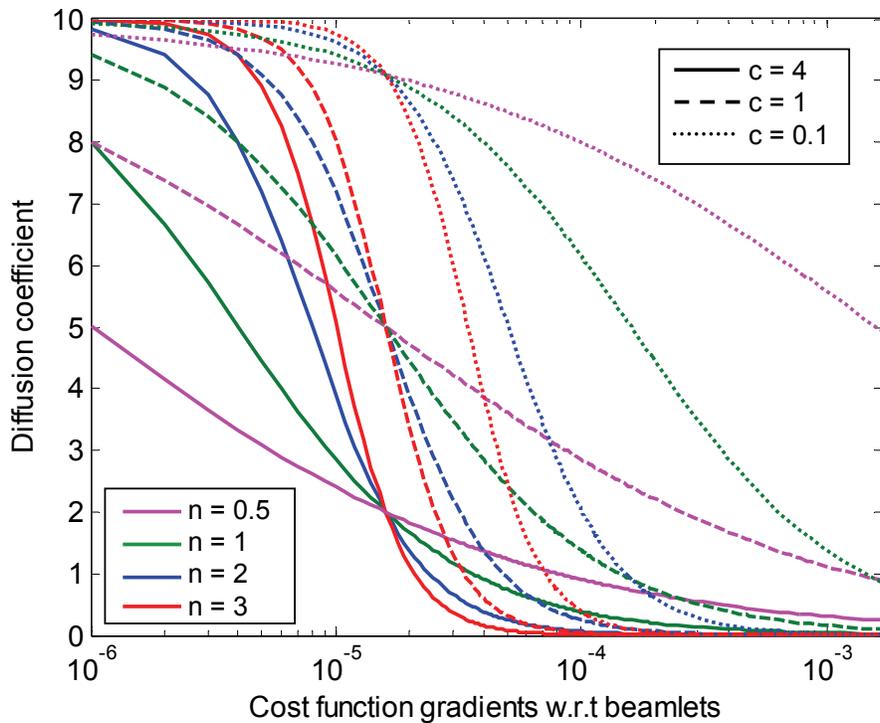


Figure V-3. The diffusion coefficient as a function of the cost function gradients with respect to the beamlets at different c and n values according to Equation V.26.

In theory, Equation V.26 implies that the highest amount of smoothing will occur in beamlets that have a small effect on the dose prescription goals, and only minor

smoothing will occur in beamlets that have a large effect on the cost function. Thus, the use of a variable diffusion coefficient defined in this way yields a method that can distinguish between desirable and undesirable modulations.

As an illustration of each of the above diffusion coefficient definitions and the ADS method itself, Figure V-4(a) shows a standard optimized intensity modulated beam from a brain cancer treatment example with the PTV and several critical structures outlined. Figure V-4(b) shows the diffusion smoothed version of that beam using uniform diffusion coefficients across the entire field. The beam was “diffusion smoothed” over five iterations to accentuate the smoothing for illustration purposes. Figures V-4(c)-(e) demonstrate the gradient-based diffusion smoothing process. Figure V-4(c) shows the $|\partial CF/\partial I_{ij}|$ values, which are highest in areas of the beam that project onto the critical normal tissues. Figure V-4(d) shows the gradient-based diffusion coefficients that are

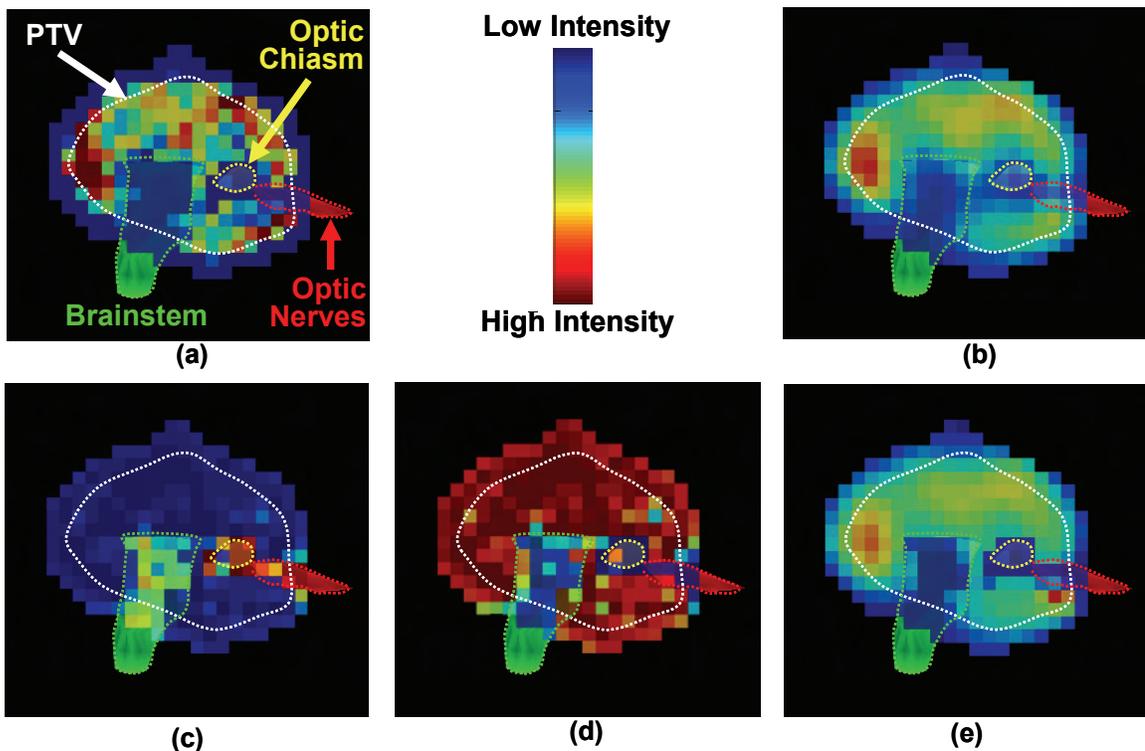


Figure V-4. (a) An optimized brain IMRT beam, (b) the diffusion smoothed beams when using uniform coefficients, (c) absolute value of the cost function gradients with respect to the beamlets, $|\partial OF/\partial I_{ij}|$, (d) adaptive diffusion smoothing coefficients that are defined as a function of (c), and (e) diffusion smoothed beams when using the gradient-based diffusion coefficients in (d). Beams are shown in the beam’s eye view in relation to the PTV, brainstem, optic nerves, and optic chiasm. Blue represents low intensity beamlets (gradients, or diffusion coefficients) while red represents high intensity.

calculated using Equation V.26 with $c = 1$ and $n = 2$. Finally Figure V-4(e) shows the diffusion smoothed beam calculated with the diffusion coefficients in shown in (d). This beam is also shown after five diffusion smoothing calculations to highlight the differences between the two methods. We see that when D is uniform, the degree of smoothing is the same over the entire field, smoothing out the modulation near the organs at risk. Conversely, when D is defined using the cost function gradients with respect to the beamlets, the least amount of smoothing occurs in areas where modulation is necessary to meet the plan objectives. In this case, these areas occur around in the overlap regions of the PTV, brainstem and optic structures. It should be noted that the smoothing that occurs with the gradient-based diffusion coefficients is not always intuitive, as in this beam, since the gradient values depend not only on the normal structure locations but also on the interplay between each of the beams.

V.B.4. ADS Penalty

Due to the fact that even small changes in the beamlet weights can cause large changes in the cost function, smoothing is usually most successful when applied inside the cost function. Thus, for the remainder of this work, the adaptive diffusion smoothing procedure is used inside the IMRT inverse cost function. To do this, we first calculate the diffusion smoothed beamlet intensities and then calculate the deviation between the original beamlet intensities and the smoothed intensities. This deviation is then penalized as a part of a weighted sum cost function with weight, ρ . Specifically, the following adaptive diffusion smoothing penalty is added to the total cost:

$$\text{ADS Penalty} = \rho \times \sum_{\text{all } ij} \sum (I_{o,ij} - I_{s,ij})^2 . \quad (\text{V.27})$$

Here $I_{s,ij}$ is calculated using Equation V.25, and D and w must abide by property 2. As stated previously, D was chosen to fall between 0 and 10 and $w = 0.02h^2$, where h is the beamlet dimension. The individual D values are variables between 0 and 10, depending on the type of adaptive diffusion smoothing coefficients chosen. We note that this penalty is similar to the modulation penalty used in Chapter IV for the Savitzky-Golay filtered beams. In Chapter IV, the smoothed beam was equal to the SG filtered beam, and here, the smoothed beam is equal to the diffusion smoothed beam.

Table V-1. Test case plan objectives and plan degradation unit (PDU) scale for comparisons to baseline plan

Structure	Goal	Approximate loss in plan quality (relative to baseline plan) equal to a plan degradation unit (PDU) of 1.0
PTV	% vol > 57 Gy = 100	0.1 Gy increase to 100 % (same as 1 Gy to 1 %)
	% vol > 63 Gy = 0	0.1 Gy increase to 100 %
OAR1	% vol > 35 Gy = 15	0.1 Gy increase to 85 %
	% vol > 63 Gy = 0	0.5 Gy increase to 100 %
	Minimize Mean Dose	1 Gy increase
OAR2	% vol > 20 Gy = 0	0.1 Gy to 100 %
	% vol > 63 Gy = 0	0.5 Gy to 100 %
	Minimize Mean Dose	1 Gy increase
Normal	% vol > 63 Gy = 1	1 Gy increase to 99 %
Tissue	Minimize Mean Dose	2 Gy increase

V.B.5. Characterization in CT Phantom and Clinical Examples

To characterize the method and coefficient choices, we applied adaptive diffusion smoothing to a test case with a central spherical target surrounded by two normal structures. The plan consisted of three 6 MV beams with 0.5 cm by 0.5 cm beamlets covering the PTV. The baseline inverse plan objectives are shown in Table V-1. The point density in the structures was adequate to properly sample the region and minimize any point-based artifacts in the beamlet intensities. Beamlet intensity optimization was first performed using the baseline cost function with UMOpt.

After optimization with the baseline cost function, the ADS penalty was added to the cost function at a given weight and the plan was re-optimized. This was repeated with increasing ADS penalty weights to study the consequences of increasing the importance of the ADS smoothness costlet. In these studies, the weight of the ADS penalty was systematically varied to analyze the range of plans and tradeoffs possible for both the uniform and gradient-based diffusion coefficients.

In addition to the comparisons made between baseline IMRT and ADS plans, we also optimized plans using the PIMV_q modulation penalty¹⁴, which was introduced in Chapter IV. To summarize, this method penalizes the *quadratic plan intensity map variation* defined by

$$PIMV_q = \sum_{n=1}^{N_b} \left(\sum_{j=1}^{J-1} \sum_{k=1}^{K-1} \left[(b_{jk} - b_{j,k+1})^2 + (b_{jk} - b_{j+1,k})^2 + \frac{1}{2}(b_{jk} - b_{j+1,k+1})^2 + \frac{1}{2}(b_{jk} - b_{j+1,k-1})^2 \right] \right). \quad (V.28)$$

Here N_b is the number of beams in a plan, J is the maximum number of beamlets in the direction parallel to the motion of the multileaf collimator (MLC), K is the maximum number of beamlets in the direction perpendicular to the motion of the MLC, and b_{jk} is the intensity of the beamlet at the (j,k) grid position. The $PIMV_q$ penalty has been shown to be a simple, yet viable smoothing costlet, and we apply it here to determine whether

Table V-2. Prostate plan objectives and plan degradation unit (PDU) scale for comparisons to baseline plan

Structure	Goal	Approximate loss in plan quality (relative to baseline plan) equal to a plan degradation unit (PDU) of 1.0
Prostate + 5 mm	% vol > 78 Gy = 100	0.1 Gy decrease to 100 % (same as 1 Gy to 1 %)
	% vol > 88 Gy = 0	0.1 Gy increase to 100 %
Rectum	% vol > 45 Gy = 20	0.1 Gy increase to 80 %
	% vol > 85 Gy = 0	0.1 Gy increase to 100 %
	Minimize Mean Dose	1 Gy increase
Bladder	< 30 % gets > 45 Gy	0.1 Gy increase to 70 %
	% vol > 85 Gy = 0	0.1 Gy increase to 100 %
	Minimize Mean Dose	1 Gy increase
Penile Bulb	Mean < 35	1 Gy increase
	% vol > 85 Gy = 0	0.1 Gy increase to 100 %
	Minimize Mean Dose	1 Gy increase
Femurs	Mean < 30	1 Gy increase
	< 10 % gets > 40	1 Gy increase to 90 %
	% vol > 45 Gy = 0	1 Gy increase to 100 %
	Minimize Mean Dose	1 Gy increase
Normal Tissue	Max < 88 Gy	2 Gy
	Minimize Mean Dose	2 Gy increase

there is an advantage to using the more sophisticated ADS scheme. The ADS penalty was evaluated by comparing the results of the ADS penalized plans to the baseline IMRT plan and the PIMVq plan for each case.

After characterizing the ADS procedure and penalty in the simple test case, we employed the same scheme to clinical examples in the prostate and head/neck. Both cases had seven equi-spaced 6 MV beams; the baseline planning objectives are shown

Table V-3. Head/neck plan objectives and plan degradation unit (PDU) scale for comparisons to baseline plan

Structure	Goal	Approximate loss in plan quality (relative to baseline plan) equal to a plan degradation unit (PDU) of 1.0
PTV70	% vol > 69.3 Gy = 100	0.15 Gy decrease to 100 % (same as 1 Gy to 1 %)
	% vol > 77 Gy = 0	0.2 Gy increase to 100 %
PTV64	% vol > 63.4 Gy = 100	0.15 Gy decrease to 100 %
	% vol > 70.4 Gy = 0	0.2 Gy increase to 100 %
PTV60	% vol > 59.4 Gy = 100	0.15 Gy decrease to 100 %
	% vol > 66 Gy = 0	0.2 Gy increase to 100 %
Cord	Max < 50 Gy	0.1Gy increase
	Minimize Mean Dose	1 Gy increase
Brainstem	Max < 54 Gy	0.1 Gy increase
	Minimize Mean Dose	1 Gy increase
Mandible	Max < 73.5 Gy	0.2 Gy increase
	Minimize Mean Dose	1 Gy increase
Parotids	Max < 77 Gy	0.2 Gy increase
	Mean < 26 Gy	0.25 Gy increase
	Minimize Mean Dose	1 Gy increase
Esophagus	Max < 50 Gy	1 Gy increase
Larynx	Minimize Mean Dose	0.5 Gy increase
	Max < 77 Gy	1 Gy increase
Oral Cavity	Mean < 49 Gy	0.5 Gy increase
	Minimize Mean Dose	1 Gy increase
Normal	Max < 73.5 Gy	1 Gy increase
Tissue	Minimize Mean Dose	2 Gy increase

in Tables V-2 and V-3 and reflect in-house IMRT protocols. The normal tissue goals for the prostate are conservatively based on several published toxicity recommendations from RTOG 9406^{15,16}. As in the rest of the studies, all cases were planned for a 6 MV linear accelerator (Varian Medical Systems, 21EX) with 120 leaf MLC (0.5 and 1.0 cm leaf widths). Dose calculations for the inverse planning system were performed by a convolution/superposition algorithm originally based on the work of Mackie *et al.*¹⁷, but optimized for beamlet calculations. Treatment planning was performed with our in-house 3-D treatment planning and IMRT optimization software packages, UMPlan and UMOpt¹⁸⁻²⁰. Leaf sequencing for static MLC (SMLC) delivery was performed with an in-house-developed leaf sequencer based on a method published by Bortfeld *et al.*²¹.

Plan comparisons (without and with varying amounts of ADS smoothing) are described by examining dose-volume histograms, relevant dose metrics, IMRT beam complexity, and delivery efficiency (MUs required). However, comparisons of different plans, especially when somewhat different optimization schemes are used, can be difficult, especially since values of the total cost function or individual costlets do not have any specific clinical relevance that can be used to compare the importance of the tradeoffs that are used to achieve the final “optimal” plans. Therefore, we describe here a method for judging the quality of inverse plan compromises (or tradeoffs) which we call “plan degradation units,” or PDU. The goal of the PDU construct is to describe a consistent unit of “tradeoff”, since different kinds of compromises are typically made among the many goals involved in a clinical inverse treatment plan.

To facilitate these comparisons, we develop a PDU scale for each case along with the baseline objectives in Tables V-1, 2, and 3. This scale is developed along with the design of the baseline cost function: we assign a concession or sacrifice value for each cost function goal which corresponds to 1 PDU, and then we calculate the costlet weight that would correspond to this value. Each concession is meant to correspond to a consistent level of plan degradation. For example, we could assign a PDU of 1.0 to be a minimum PTV dose of 59.9 Gy instead of 60 Gy, while also defining 1 PDU as 99 % of the PTV receiving 60 Gy and 1 % of the PTV receiving 59 Gy. For a less important objective, a PDU of 1.0 may correspond to larger dose concession, such as allowing the unspecified mean normal tissue dose to increase by 2 Gy. Inspection of the PDU scales shown in Tables V-1 through 3 shows that we have chosen PDU scales that are conservative. In other words, the PDU value is meant to be very sensitive to small changes in the dosimetric goals. This was done purposefully, to show that a significant

amount of smoothing is possible in IMRT plans without sacrificing clinical plan quality. In Tables V-1, 2 and 3, the plan quality metric sacrifices (relative to the achieved baseline plan objectives in Gy that correspond to a PDU of 1.0) are given.

V.C. Results

V.C.1. Phantom Study

The adaptive diffusion smoothing procedure was implemented into our 3D treatment optimization system infrastructure as a penalty or costlet to be used inside the cost function of an inverse IMRT plan. A test case was used to test the ADS implementation and characterize the two different methods for setting diffusion coefficients. A baseline plan was established for this case by optimizing beamlet weights according to the minimization of the cost functions in Table V-1. Then, to study the impact of including the ADS penalty at varying weights in the cost function, we re-optimized the baseline plan while systematically increasing the ADS penalty weights.

Results of the phantom comparisons are shown in Figure V-5. Figure V-5(a) shows the simple 3-field beam arrangement and anatomy. Figure V-5(b) illustrates the tradeoff between MU and dosimetric plan quality as the modulation penalty weights are increased for the PIMV_q, ADS-Uniform and ADS-Gradient smoothing penalties. This figure plots the relative MU required to deliver the plan as a function of plan degradation (PDUs) with respect to the baseline plan. Table V-1 contains the PDU scale for the phantom case. A PDU of 1 corresponds to a loss in plan quality relative to the baseline plan and could be equivalent to, for example, a 0.1 Gy decrease in the minimum dose to the target, or a 2 Gy increase in the mean dose to the normal tissue, or a lesser combination of the two. As stated previously, we have purposefully made the PDU scale sensitive to changes in plan quality, to show that a large reduction in MU is possible with small losses in plan quality.

In Figures V-5(b) and (c), the plans optimized with each different method to similar relative MU values are compared (denoted by the circled plans in Figure V-4(b)). Figure V-5(c) shows the DVHs for each of these plans compared to the baseline plan, and Figure V-5(d) illustrates the effect of the various smoothing penalties on the beamlet intensity maps. For plans requiring approximately the same MU, the ADS-Gradient penalty yielded the highest dosimetric quality plan. With a PDU of 2.0, the difference between the baseline plan and ADS-Gradient plan is very small, demonstrating the conservative PDU scale. From the beam's eye view in Figure V-5(d), we see a large

reduction in overall modulation when using each of the smoothing methods, although the complexity reduction is slightly different for each technique.

The large amounts of smoothing possible in this case may be indicative of the simple geometry and cost function. In fact, the optimal c and n values in Equation V.26 for this case were 0.1 and 2, respectively. While $n = 2$ results in a reasonable fall-off from high D to low D , the low c value of a shifts this fall-off so that it occurs at a fairly high value of D , meaning that the majority of the beamlets were maximally smoothed.

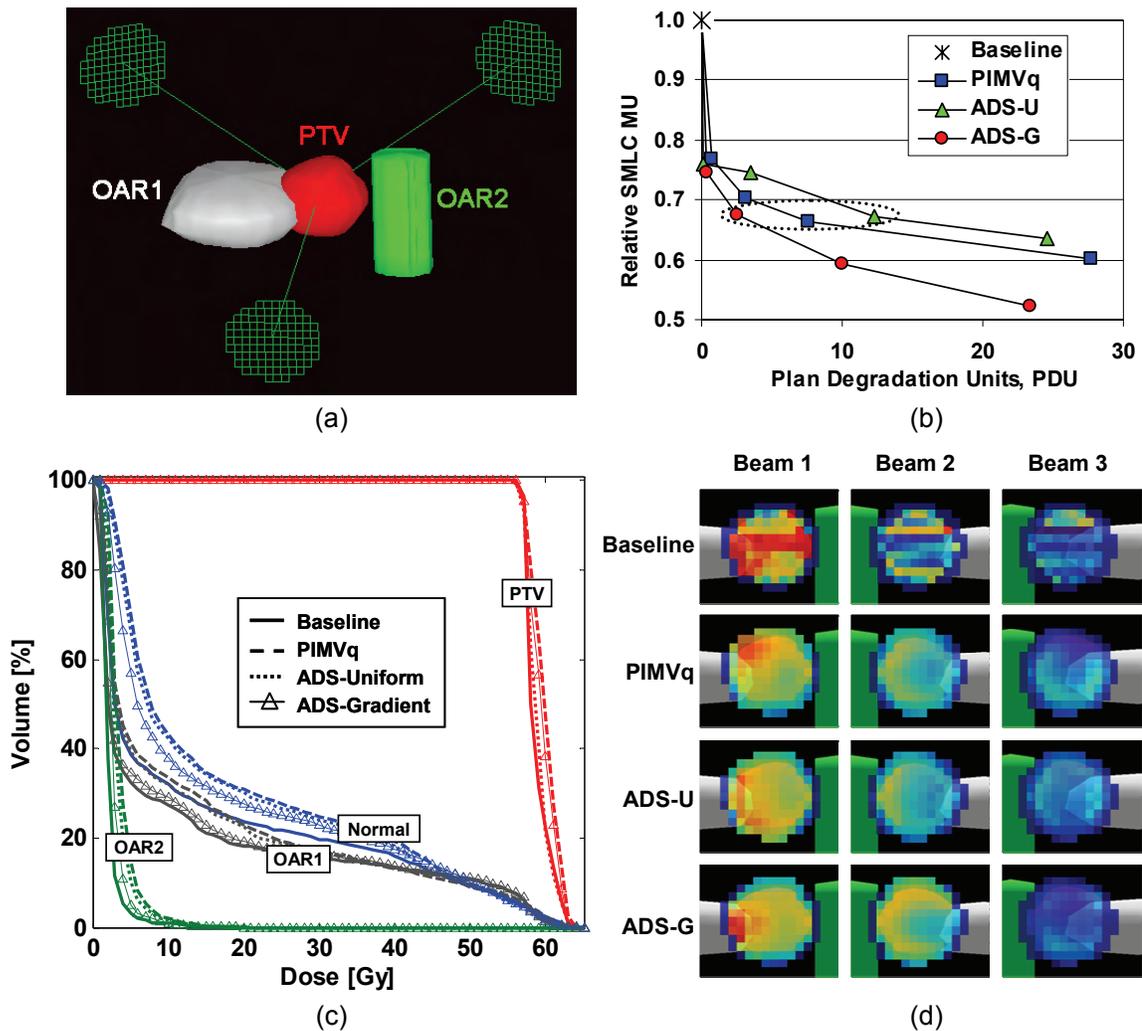


Figure V-5(a) The geometry and beam arrangement, (b) relative MU (to the baseline plan) as a function of plan degradation units, (c) dose-volume histograms for the circled plans in (b), and (d) the corresponding intensity modulated beams in the beam's eye view for the phantom case 1 for standard IMRT and the three different modulation penalties. The outline of the two OARs is shown by the dotted lines. Red corresponds to high intensity beamlets and blue to low intensity beamlets.

As expected, the steep fluence gradients and modulation near normal tissue interfaces and overlap regions appear to be preserved to a greater extent in the ADS-Gradient beams than in the ADS-Uniform beams. On the other hand, the PIMV_q penalty promotes an overall reduction of the intensity variation and high intensities in the fields and leads to the distribution of intensity more evenly across the fields. When applying the truly adaptive gradient-based diffusion coefficients in ADS, there appears to be a higher degree of smoothing in areas that are not in tissue overlap regions. This is expected because the cost function gradients with respect to the vector of beamlet intensities ($\partial CF/\partial \mathbf{I}$) are likely to be more sensitive in these interface regions. Because of this additional smoothing, the ADS-Gradient beams may result in lower MU than the ADS-Uniform and PIMV_q penalties. While the latter two penalties do penalize the modulation more across the entire field, they are still able to preserve gradients in important areas where smoothing has a large negative effect on the cost function. This is a consequence of using the penalty inside the cost function. We would expect to see greater differences, such as those in Figure V-4, if we smoothed outside the cost function. This would be even more pronounced if the PIMV_q penalty were applied post-optimization.

These initial results demonstrate that diffusion smoothing is a promising method for controlling modulation in IMRT fields, with an advantage in using the adaptive gradient-based coefficients.

V.C.2. Clinical Examples

To gain more insight into the merits of each of the two different coefficient definition methods, we tested each method for the more complicated clinical geometries of the prostate and head/neck. These cases were optimized with the cost functions shown in Tables V-2 and V-3, and the ADS penalties (using both diffusion coefficient definition methods). As in the phantom case, the baseline cost function was first optimized and then smoothing penalties were added at increasing weights inside the cost function. The weights were systematically increased to observe the tradeoffs made with the baseline goals.

V.C.2.1. Prostate

Figure V-5(a) shows results of the prostate baseline optimization along with the potential for reducing MU through use of the ADS and PIMV_q penalties, as a function of plan

degradation units. The addition of any of the smoothing penalties reduced the MU substantially, although a greater benefit is observed with the ADS-Gradient penalty. The maximum reductions in MU were around 26 - 36 %, although this would be much higher if more degradation were allowed in the plan objectives. It can be seen that the ADS-Gradient method achieves its maximum result with smaller tradeoffs (PDUs) than the other penalties. This is because the ADS-Gradient penalty does not highly penalize the important areas of modulation. Therefore, this penalty will be small for high quality plans, while the other methods can still create large penalties because they are penalizing overall modulation.

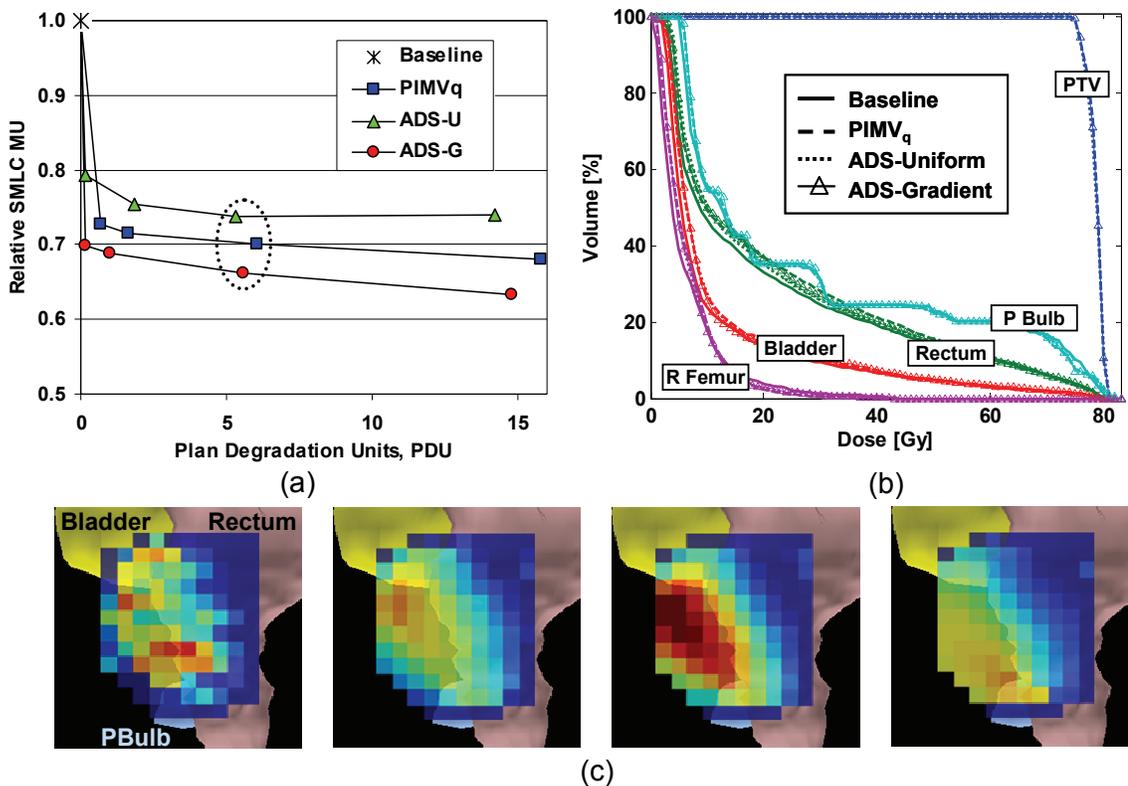


Figure V-6(a) MU (relative to the baseline plan) as a function of plan degradation units, **(b)** dose-volume histograms for the circled plans in (a), and **(c)** the corresponding intensity modulated beams in the beam's eye view for the prostate case for baseline IMRT and the three different modulation penalties. From left to right is Baseline, PIMVq, ADS-Uniform, and ADS-Gradient. The bladder, rectum, and penile bulb are shown in the beam's eye view.

In Figure V-5, we showed that in plans with similar MU requirements, the ADS-Gradient plan had the best dosimetric quality. Conversely, in Figure V-6, we compare plans with similar dosimetric quality. Figure V-6(b) shows the DVHs for the baseline

plan and the other plans at a PDU level of approximately 6. Because of the number of structures in the cost function, this level of plan degradation is very slight, as seen by inspection of the (nearly identical) DVHs. In Figure V-6(c), we present an example typical beam intensity distribution for each optimization method, along with the rectum, bladder, and penile bulb in the beam's eye view. Small differences are observed between the ADS-Uniform and ADS-Gradient beams: the ADS-Gradient beams are more uniform in areas where there are no normal tissue overlap regions and allow for quicker fall-off between high and low intensity regions. These features combine throughout all of the beams to produce more smoothing and MU reduction at the same level of plan quality. The smoothing that occurs is a function of the interplay between the cost function and many beamlets from seven directions. Therefore, we must be careful not to draw significant conclusions from the display of a single beam. An inspection of all seven beams, however, demonstrates that the PIMV_q beams display an overall flattening of intensity across the beams with the intensity contributions from each beam becoming more uniform. On the other hand, the ADS-Gradient plans show a shift in intensity contributions to the beams that intersect the fewest organs at risk. From this point-of-view, the ADS-Gradient beams are much more intuitive than the PIMV_q beams.

In comparison to the optimal c and n values in Equation V.26 for the phantom compared to this prostate example, the n value remained at 2 while the c value increased to an optimum of 0.5. Again, the relative small c value means that the majority of the beamlets are being maximally smoothed, but not quite to the same degree as the phantom case. A preliminary observation is that the c value is an indication of the difficulty of the geometry and/or cost function.

V.C.2.2. Head/Neck

The inverse plan objectives used in the head/neck example were very strict and closely reflect our current clinical standard. Despite the strict cost function and seemingly small amount of solution space to work with, the application of both the ADS and PIMV_q penalties resulted in a substantial reduction of modulation and MU (Figure V-7). This demonstrates that, even in complicated cases, there may still be a large range of plans that can achieve similar results, and that some of those plans may be more desirable in terms of plan efficiency. The use of the ADS (and PIMV_q) smoothing costlets enabled us to find a more efficient plan without sacrificing the quality achieved with the baseline plan.

In contrast to the phantom and prostate cases, there was no substantial difference in MU reduction between the different smoothing costlets (Figure V-7(a)). The ADS-Gradient and PIMV_q penalties showed nearly identical results in terms of MU reduction and the ADS-Uniform penalty was slightly worse. Figure V-7(c) demonstrates typical qualitative differences between beams smoothed with each of the different methods. These differences are very small, which may point to the fact the head/neck plan solution space with the additional objectives of modulation reduction is fairly small. The ADS beams both appear to do a slightly better job at sparing the larynx in the beam shown, although the DVHs are nearly identical for all methods. The optimal c and n values in Equation V.26 may also indicate the complexity of the case itself. The resulting c and n values were $c = 5$ and $n = 0.25$, which is quite different from the previous two cases. Here, the optimal diffusion coefficients fall off very slowly from high

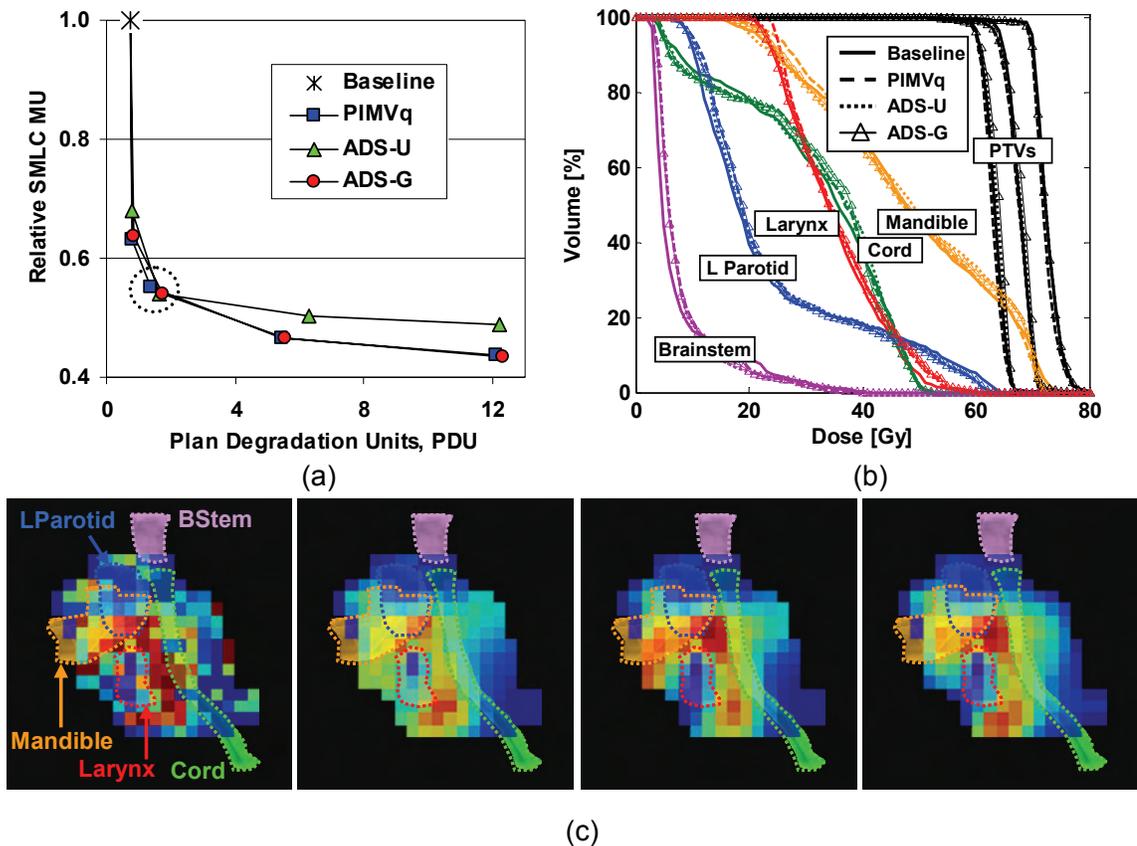
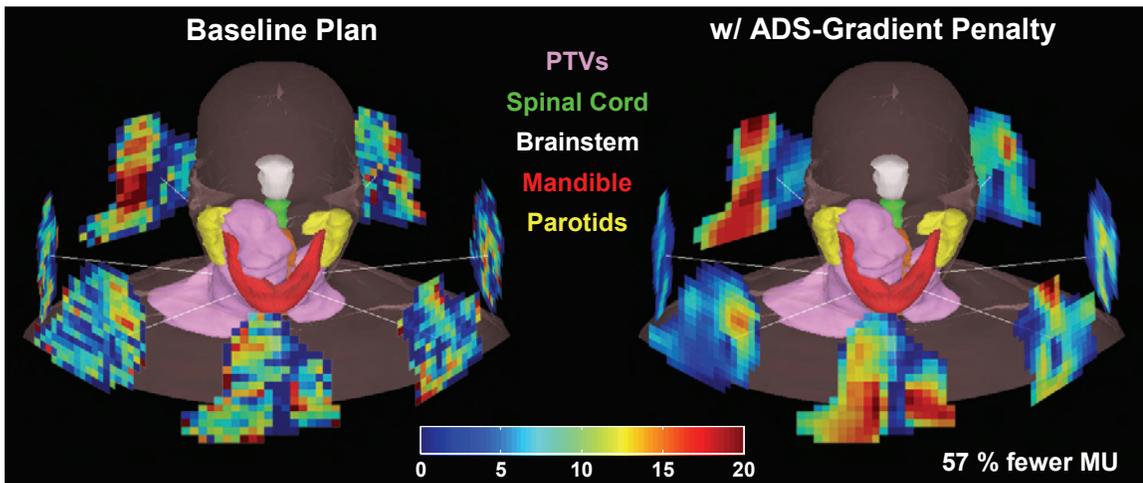


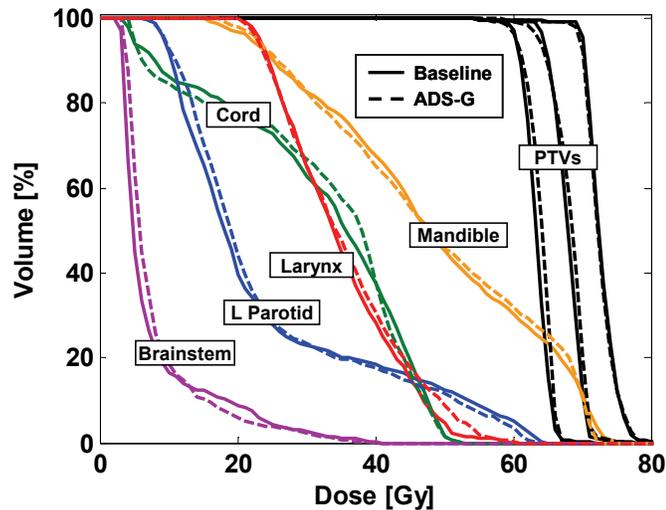
Figure V-7(a) Relative SMLC monitor units are shown as a function of the plan degradation baseline cost for the head/neck plan. **(b)** DVHs and **(c)** a typical intensity modulated beam are shown for the circled plan in (A) for each of the optimization methods. From left to right is Baseline, ADS-Gradient, ADS-Uniform, and PIMVq. Several normal structures are outlined in the beam's eye view.

to low intensity and are focused on a smaller and lower range of values, due to the combination of the low n and high c values.

To demonstrate the ability of the ADS-Gradient penalty to preserve intensity gradients as well as to smooth at high penalty weights, the final (most penalized) ADS-Gradient plan is shown in Figure V-8. 7. Figure V-8(a) illustrates the difference between the baseline and ADS-smoothed plans via a 3D visualization of the two plans, with several of the important regions of interest displayed. This ADS-Gradient plan is noticeably smoother and can be delivered with 57 % fewer MU than for the baseline



(a)



(b)

Figure V-8 (a) Head/neck geometry and intensity modulated beams for the baseline plan and the last (most smoothed) ADS-Gradient plan shown in Figure 5. The ADS-gradient plan can be delivered with 57 % fewer SMLC monitor units with minimal effect on the baseline plan quality, as shown in the DVHs in (b).

plan. The loss in dosimetric quality is minimal, which can be seen in the DVH comparison of the two plans in Figure V-8(b).

V.D. Discussion

In summary, in the phantom and prostate, we are able to reduce MU by approximately 30 - 40 % with no loss in the plan quality when using ADS, and MU reductions greater than 40 % are attainable with only very small concessions in the baseline plan. While all smoothing penalties performed well, use of the adaptive gradient-based diffusion coefficients in the ADS-Gradient penalty was able to reduce MU by around 10 % more at higher penalty weights in the phantom and prostate cases. This advantage may be due to the fact that the ADS-Gradient penalty more appropriately penalizes the less important modulation more than the more important modulation. Therefore, it can preserve – with minimum penalty – the essential modulation in the plan while smoothing large regions of the beam.

In the prostate, this results in more uniform areas in the beam that require fewer MU to deliver. Although a large amount of smoothing was achieved in the head/neck case with the ADS-Gradient method, significant differences in MU reduction were not seen in this example compared to the $PIMV_q$ penalty. Superior improvement in delivery efficiency with the more sophisticated ADS-Gradient method may not have been observed in this case because the cost function gradients were much higher and fluctuated more than in the other sites, due to the large number of clinical goals, and the comparative difficulty of the cost function. In addition, we expect that some uncertainties in the gradients (due to the point sampling, for example) will exist and may lead to undesirable variation in the diffusion coefficients.

This preliminary evaluation of adaptive diffusion smoothing with spatially variant diffusion coefficients (the ADS-Gradient penalty) revealed that it has great potential as a tool to reduce IMRT beam complexity in regions where the complexity is not necessary to produce a quality plan. The ADS penalty behaves in such a way that increasing penalty weights results in a smoother beam. This is a very important property of ADS that is not held by many other smoothing filters, such as the Savitzky-Golay filter that we used in Chapter IV. At higher weights, the SG penalty introduced modulation into the field, which made it undesirable and unpredictable for use in a clinical setting.

We have shown that the use of the ADS penalty does not have to lead to a reduction in dosimetric plan quality and can significantly reduce modulation and MU. In the

phantom and prostate examples tested in this work, the use of uniform smoothing methods such as the ADS-Uniform penalty or PIMVq penalty was adequate to reduce beam complexity. However, the gradient-based diffusion coefficients in the ADS-Gradient penalty more successfully smoothed in areas where modulation is not essential to meeting the plan objectives.

The case studies provided some guidance on the optimal c and n values that can be used in Equation V.26 to define the gradient diffusion coefficients in different geometries. Having these parameters is an advantage because it allows us to further customize the method to individual treatment sites, but it can also be a disadvantage if this customization is required for each individual patient. The geometries shown in this chapter demonstrated a large range of optimal c and n values, and tuning these parameters could be time-consuming if it had to be done for each case. Further testing in each of these sites suggests that c and n can be fixed for similar geometries.

In this chapter (and in this dissertation as a whole), we have not studied smoothing outside of the plan optimization (e.g., within a leaf sequencer), other than in the initial example (Figure V-3), which was used to demonstrate how the smoothing (by itself) works. However, it would be possible to study the application of an ADS algorithm for post-optimization smoothing within the leaf sequencing step. In this kind of study, we would expect the diffusion coefficient definition to have a much greater impact on the final solution. Thus, for centers employing post-optimization smoothing or filtering inside the leaf sequencing process, gradient-based diffusion smoothing could be an attractive option.

One concern with smoothing IMRT plans is the effect it will have on the geometric sensitivity of the plan. This complicated question can only be fully answered through the simulation of a large number of treatment courses for a variety of sites. While such an involved study is beyond the scope of this dissertation, we present some preliminary testing of the impact of smoothing costlets on geometric sensitivity in a phantom case in Chapter VII. In addition, we have observed from preliminary testing on the prostate case that there may be an advantage to using the ADS-Gradient method to improve target coverage in the face of setup uncertainty.

One of the most exciting results of this study is the large amount of smoothing possible without affecting the quality of the plan. In Chapter VI, we retrospectively analyze a series of clinical breast cancer IMRT plans to learn how much smoothing would have been possible with the use of the ADS-Gradient penalty. Chapter VI also

combines ADS with the Lexicographic Ordering²² method to quantify the tradeoffs between the modulation penalties and the plan objectives. This will allow physicians to make educated choices between smoothing and the plan objectives.

On the other hand, for weighted-sum cost functions, the development of a more quantitative way to show comparisons between plans using the new idea of plan degradation units (PDU) has been successful. Many current optimization algorithms rely on the use of conventional weighted-sum cost functions and require a large number of trial-and-error iterations to choose the proper weights for the individual objectives. The addition of a modulation penalty can affect the other objectives in different ways, and evaluating the overall cost after the addition of the modulation penalty can be difficult. The plan degradation unit scale puts a value on different degrees of plan degradation and makes it more intuitive to evaluate the tradeoffs made when including a modulation penalty. Instead of simply providing a “cost” with no obvious clinical relevance, the PDU value gives a more reliable gauge of the change in plan quality and makes it easier to choose an acceptable weight for the modulation penalty. The adoption of this methodology has been useful for judging and designing cost functions as well as applying modulation penalties.

Chapter I introduced several alternate methods that are used to reduce MU and modulation in IMRT, such as direct aperture, or direct segment optimization^{2,3}. Adaptive diffusion smoothing is unique because of the ability to customize the diffusion coefficients. For example, specialized diffusion coefficients could be used to manipulate beamlet intensities to reduce plan sensitivity to setup errors²³, organ motion, and even undesirable delivery artifacts such as tongue and groove under-dosage. Custom diffusion coefficients may also have a place in adaptive radiation therapy to ensure that large intensity gradients do not occur in areas that may require corrections in an adaptive scheme. This should make it easier to apply feedback during the treatment course, to make fractional changes in the intensity patterns required to correct or change the dose prescription. Some of these ideas are discussed in Chapter VII. Thus, the increase of delivery efficiency may be just one of the possible applications of the ADS method.

V.E. Conclusion

The diffusion equation has been used in a procedure that preferentially smoothes IMRT plans, using a diffusion coefficient matrix that allows the degree of smoothing to adapt to

each individual plan. This procedure was used to define an “adaptive diffusion smoothing” (ADS) penalty, applied inside an inverse planning cost function to promote overall smoothing and monitor unit reduction. Two methods for definition of the diffusion coefficients — to promote uniform smoothing and smoothing based on the beamlet gradients (partial derivatives of the cost function with respect to the beamlet intensity) — were applied and tested on a CT phantom and two clinical examples. Without compromising the baseline cost function, MU reductions on the order of 30 % and 40 % were obtained with the ADS penalties. Compared to the ADS-Uniform penalty, the ADS-Gradient penalty was better able to preserve intensity gradients and modulation in important areas of the IMRT fields, leading to an advantage in reducing MU in the phantom and prostate cases. This was possible because the gradient-based diffusion coefficients preferentially induce smoothing in the beam where it does not interfere with meeting the dose prescription objectives. All smoothing penalties were equally successful in the head/neck example. Overall, the ADS procedure and penalty is a promising tool for smoothing the unnecessary modulation in IMRT plans, and it may well have additional important uses due to the possibility to customize the diffusion coefficients for other specific purposes.

References

- ¹ M. M. Coselmon, J. M. Moran, J. Radawski, and B. A. Fraass, "Improving IMRT delivery efficiency by applying intensity limits during inverse planning," *Med. Phys.* **32**, 1234-1245 (2005).
- ² B. A. Fraass and D. L. McShan, "Optimization of Conformal and IMRT Plans with Direct Segment Optimization," Proceedings of XVth International Conference on the Use of Computers in Radiation Therapy, Toronto, ON, Canada, 2007.
- ³ D. M. Shepard, M. A. Earl, X. A. Li, S. Naqvi, and C. Yu, "Direct aperture optimization: A turnkey solution for step-and-shoot IMRT," *Med. Phys.* **29**, 1007-1018 (2002).
- ⁴ J. L. Bedford and S. Webb, "Constrained segment shapes in direct-aperture optimization for step-and-shoot IMRT," *Med Phys* **33**, 944-958 (2006).
- ⁵ J. Markman, D. A. Low, A. W. Beavis, and J. O. Deasy, "Beyond bixels: Generalizing the optimization parameters for intensity modulated radiation therapy," *Med. Phys.* **29**, 2298-2304 (2002).
- ⁶ R. Mohan, M. Arnfield, S. Tong, Q. Wu, and J. Siebers, "The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy," *Med. Phys.* **27**, 1226-1237 (2000).
- ⁷ S. Spirou, N. Fournier-Bidoz, J. Yang, C. Chui, and C. Ling, "Smoothing intensity-modulated beam profiles to improve the efficiency of delivery," *Med. Phys.* **28**, 2105-2112 (2001).
- ⁸ X. Sun and P. Xia, "A new smoothing procedure to reduce delivery segments for static MLC-based IMRT planning," *Med. Phys.* **43**, 2785-2794 (2004).
- ⁹ S. Webb, D. J. Convery, and P. M. Evens, "Inverse planning with constraints to generate smoothed intensity-modulated beams," *Phys. Med. Biol.* **43**, 2785-2794 (1998).
- ¹⁰ M. Alber and F. Nusslin, "Intensity modulated photon beams subject to a minimal surface smoothing constraint," *Phys. Med. Biol.* **45**, N49-N52 (2000).
- ¹¹ J. Llacer, N. Agazaryan, T. D. Solberg, and C. Promberger, "Degeneracy, frequency response and filtering in IMRT optimization," *Phys. Med. Biol.* **49**, 2853-2880 (2004).
- ¹² R. J. J. Stamm'ler and M. J. Abbate, *Methods of Steady-State Reactor Physics in Nuclear Design*. (Academic Press, London, 1983).
- ¹³ L. Ma, "Smoothing intensity-modulated treatment delivery under hardware constraints," *Med. Phys.* **29**, 2937-2295 (2002).
- ¹⁴ M. M. Matuszak, E. W. Larsen, and B. A. Fraass, "Reduction of IMRT beam complexity through the use of beam modulation penalties in the objective function," *Med. Phys.* **34**, 507-520 (2006).
- ¹⁵ M. Roach, K. Winter, J. M. Michalski, J. D. Cox, J. A. Purdy, W. Bosch, X. Lin, and W. S. Shipley, "Penile bulb dose and impotence after three-dimensional conformal radiotherapy for prostate cancer on RTOG 9406: Findings from a prospective, multi-institutional, phase I/II dose-escalation study," *Int J Radiat Oncol Biol Phys* **60**, 1351-1356 (2004).

- ¹⁶ J. M. Michalski, K. Winter, J. A. Purdy, M. Parliament, H. Wong, C. A. Perez, M. Roach, W. Bosch, and J. D. Cox, "Toxicity after three-dimensional radiotherapy for prostate cancer on RTOG 9406 level V," *Int J Radiat Oncol Biol Phys* **62**, 706-716 (2005).
- ¹⁷ T. R. Mackie, J. W. Scrimger, and J. J. Battista, "A convolution method of calculating dose for 15-MV x rays," *Med. Phys.* **12**, 188-196 (1985).
- ¹⁸ B. A. Fraass and D. L. McShan, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F.H. van der Giessen, H.J. van Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 273-276.
- ¹⁹ J. H. Kim, N. Dogan, D. L. McShan, and M. L. Kessler, "An AVS-based system for optimization of conformal radiotherapy treatment plans," Proceedings of 1995 International Advanced Visual Systems User and Developer Conference, Boston MA, 1995.
- ²⁰ M. L. Kessler, D. L. McShan, M. Epelman, K. A. Vineberg, A. Eisbruch, T. S. Lawrence, and B. A. Fraass, "Costlets: a generalized approach to cost functions for automated optimization of IMRT treatment plans," *Optimization and Engineering* **6**, 421-448 (2005).
- ²¹ T. R. Bortfeld, D. L. Kahler, T. J. Waldron, and A. L. Boyer, "X-ray field compensation with multileaf collimators," *Int J Radiat Oncol Biol Phys* **28**, 723-730 (1994).
- ²² K.-W. Jee, D. L. McShan, and B. A. Fraass, "Lexicographic Ordering: Intuitive Multicriteria Optimization for IMRT," *Phys Med Biol* **52**, 1845-1861 (2007).
- ²³ M. M. Matuszak, E. W. Larsen, R. K. Ten Haken, and B. A. Fraass, "Uniform radiobiological targeting with adaptive diffusion smoothing during IMRT optimization," Proceedings of XVth International Conference on the Use of Computers in Radiation Therapy, Toronto, Ontario, Canada, 2007.

CHAPTER VI.

APPLICATIONS OF ADAPTIVE DIFFUSION SMOOTHING

Chapter V introduced a diffusion-based procedure that can be used to intelligently smooth IMRT beams. This adaptive diffusion smoothing (ADS) procedure is able to perform spatially-varying smoothing based on diffusion coefficients that are defined for each individual beamlet. The application of ADS to improve the delivery efficiency of IMRT plans was studied in the previous chapter. However, the versatility of the diffusion coefficient paradigm suggests that ADS may have other uses in IMRT. To illustrate these uses, this chapter will discuss three applications of ADS, including (i) the combination of ADS smoothing with Lexicographic Ordering, a multi-criteria optimization method, (ii) a retrospective clinical study to determine the effect that ADS smoothing could have on patients being treated with an accelerated partial breast IMRT technique, and (iii) the use of ADS in promoting smoothness over critical tumor regions that may be more susceptible to risks associated with heterogeneous fluence distributions.

VI.A. ADS applied within Lexicographic Ordering

VI.A.1. Introduction

Chapter V showed the ADS penalty to be a useful tool in obtaining optimally smoothed IMRT plans that are of high dosimetric quality and can be delivered more efficiently than standard, unsmoothed IMRT solutions. However, choosing the proper weight at which to apply the ADS penalty inside a weighted-sum cost function may require many iterations and may impact the plan in such a way that the original target and normal tissue objective weights are altered. This trial and error process is time-consuming and does not ensure that one will reach the optimal tradeoff between beam complexity and the other plan objectives. This sub-optimal planning strategy makes it difficult to include a smoothness objective into clinical IMRT protocols.

The use of an intuitive method for multicriteria IMRT optimization called lexicographic ordering (LO) was introduced at our institution¹ (and others^{1,2}) as a way to articulate planning goals through the definition of priority levels. In many clinical protocols, including those at the University of Michigan, planning goals are grouped by their importance into distinct levels. The LO method allows for optimization of a plan according to these priority levels, with each prior level dominating the subsequent level. Thus, it is not necessary to design a weighted-sum cost function to facilitate tradeoffs between the different objectives, and more importantly, sacrifices are not allowed in more important goals to improve the less important goals.

The combination of multicriteria optimization and smoothing objectives has been discussed by Wilkens *et al.*, who suggested that fluence smoothness could be promoted as a final priority in a prioritized optimization method². Their work used a Laplacian filter to generate a measure of the “roughness” of the fields and then minimized this value as a final priority in the optimization. This method captures and tries to minimize the high frequency components of the field, and their results showed a reduction in the field Laplacian after application of the smoothness step. However, this method does not distinguish between undesirable and desirable modulation, and the effect on plan delivery efficiency was not reported. Another study by Craft *et al.* explored the direct tradeoff between monitor units and individual plan objectives using a Pareto optimality scheme³. Their work supports the notion that IMRT plan delivery efficiency can be significantly improved with relatively small concessions in the plan objectives.

This section studies the combination of adaptive diffusion smoothing modulation penalties with the intuitive multicriteria lexicographic ordering method to (i) determine the amount of delivery efficiency improvement possible in several clinical examples by minimizing the ADS penalty as the final priority in LO, and (ii) demonstrate how to facilitate controlled tradeoffs between the ADS penalty and the dosimetric objectives through the retrospective relaxation of previously achieved objectives during LO optimization.

VI.A.2. Methods

VI.A.2.1. Adaptive Diffusion Smoothing

The adaptive diffusion smoothing theory and implementation is discussed in detail in Chapter V. To summarize, this method is based on the diffusion equation and preferentially smoothes IMRT fields according to diffusion coefficients that are defined

for each beamlet. These coefficients dictate the amount of smoothing that is applied in each beamlet and they can be defined in many different ways. A smoothed IMRT beamlet, $I_{s,ij}$, is calculated from an original IMRT beam, I_o , and a matrix of diffusion coefficients, \mathbf{D} , according to:

$$I_{s,ij} = I_{o,ij} \left[1 - \frac{w}{h^2} \left(D_{i+\frac{1}{2},j} + D_{i-\frac{1}{2},j} + D_{i,j+\frac{1}{2}} + D_{i,j-\frac{1}{2}} \right) \right] + \frac{w}{h^2} \left(D_{i+\frac{1}{2},j} I_{o,i+1,j} + D_{i-\frac{1}{2},j} I_{o,i-1,j} + D_{i,j+\frac{1}{2}} I_{o,i,j+1} + D_{i,j-\frac{1}{2}} I_{o,i,j-1} \right), \quad (\text{VI.1})$$

where

$$D_{i+\frac{1}{2},j} = \frac{2D_{ij}D_{i+1,j}}{D_{ij} + D_{i+1,j}}, \quad 0 \text{ on boundaries,} \quad (\text{VI.2})$$

$$D_{i,j+\frac{1}{2}} = \frac{2D_{ij}D_{i,j+1}}{D_{ij} + D_{i,j+1}}, \quad 0 \text{ on boundaries.}$$

In Equation VI.1, w is the time-step of the diffusion process, and h is the dimension of one side of a square beamlet. These equations reveal that the amount of smoothing is proportional to the size of the diffusion coefficient. Another property of diffusion smoothing is that:

$$\text{If } \frac{w}{h^2} \left(D_{i+\frac{1}{2},j} + D_{i-\frac{1}{2},j} + D_{i,j+\frac{1}{2}} + D_{i,j-\frac{1}{2}} \right) \leq 1 \text{ for all } ij, \quad (\text{VI.3})$$

$$\text{then } I_{o,\min} \leq I_s(x,y) \leq I_{o,\max}$$

If this inequality is satisfied, the maximum and minimum intensities of the smoothed beam will lie between the maximum and minimum intensities of the unsmoothed beam. This property also ensures that as long as the original beam is positive, the smoothed beam will also be positive. As in our previous work, we have chosen to limit D to the range between 0 and 10 for all beamlets, and thus to satisfy the above inequality, we chose $w = 0.02h^2$.

The adaptive diffusion method has an advantage over conventional smoothing filters because of the ability of the user to customize the diffusion coefficients in such a way that the desirable modulation is not penalized to the same degree as the

undesirable, or unnecessary modulation. As discussed in Chapter II, the existence of areas in the beam that are less important to meeting the cost function goals has been proven by Alber *et al.* in their study of the degeneracy of IMRT problems⁴. They found that the number of beamlets that have significant effects on the cost function can be quite small compared to the number of beamlets in an IMRT plan, and this can lead to unnecessarily noisy IMRT beams. This degeneracy has also been observed in work by other authors⁵⁻⁷. Therefore, the beamlet intensity distribution can become noisy because the cost function is not strongly affected by the intensity of these beamlets, and smoothing these “unimportant” beamlets can reduce the modulation and MU required to treat the plan with equivalent dosimetric quality.

The ADS method addresses this issue by defining each diffusion coefficient to be a function of the cost function gradient with respect to that individual beamlet,

$$\frac{\partial f}{\partial I_{ij}} , \quad (VI.4)$$

where f is the cost function and I_{ij} is the beamlet at the position (i,j) in the beamlet grid. After a study of a variety of different formulations for the diffusion coefficients, we have chosen the following formulation from Chapter V to be a robust and tunable formulation for the diffusion coefficients as a function of the cost function gradients:

$$D_{ij} = \frac{10}{1 + c \left(\frac{|\partial f / \partial I_{0,j}|}{s} \right)^n} . \quad (VI.5)$$

Here, s is a gradient scaling factor that is equal to the median value of $|\partial f / \partial I_0|$. The median value was chosen instead of the mean value so that the scaling factor was not biased by extremely high values of the gradient. The parameters c and n are tunable parameters that allow D to be further customized depending on the individual case. Our studies have shown that plans in the same general class will have similar optimal values of c and n . The parameter c serves to shift the transition from high to low D and n controls the steepness of the fall-off from high to low D .

Equation VI.5 promotes the highest amount of smoothing in beamlets that have a small effect on the cost function, and only minor smoothing will occur in beamlets that have a large effect on the cost function. Our use of this variable diffusion coefficient definition has yielded a way to distinguish between desirable and undesirable modulation in the smoothing procedure. Even though this method results in much less plan degradation after smoothing than a uniform smoothing procedure, there can still be some loss in plan quality after the smoothing procedure. Therefore, an ADS penalty is used inside the cost function as a costlet which penalizes the difference between the beam chosen by the optimizer and the corresponding diffusion smoothed beam at each iteration of the optimization:

$$\text{ADS Penalty} = \sum_{\text{all } ij} \sum (I_{o,ij} - I_{s,ij})^2 . \quad (\text{VI.6})$$

Here $I_{s,ij}$ is calculated using Equation VI.2, and a weighting factor could be used to scale the ADS penalty for use in a weighted-sum cost function or inside a sub-problem in lexicographic ordering.

VI.A.2.2. Lexicographic Ordering

The Lexicographic Ordering (LO) method divides a large optimization problem with many objectives into a series of smaller optimization problems that can be solved sequentially based on their priority in the problem. Our implementation of this method for use in planning intensity-modulated radiotherapy treatments is discussed in detail by Jee *et al.*¹, and a summary of the method is presented here.

The initial step in the LO method is to rank the objectives according to their significance in the optimization problem, and then group objectives of similar importance into priority levels from 1 (highest priority) to N (lowest priority). After prioritization, an optimization search algorithm is applied to solve each level of the optimization problem. After the first level priority objective function, $f_1(\mathbf{I}_1)$, is minimized, and the achieved objectives are converted into inequality constraints with boundary values, $f_1(\mathbf{I}_1^*)$, that are set by the prior attained solutions, $\mathbf{I}_1^* = \text{argmin } f_1(\mathbf{I})$. As the LO method proceeds from priority level 1 to N, it solves each optimization problem according to:

$$\begin{aligned}
& \min f_i(\mathbf{I}) \\
& \text{subject to } f_j(\mathbf{I}) \leq f_j(\mathbf{I}_j^*) \\
& \text{where } i = 1, 2, \dots, N \text{ and } j = 1, 2, \dots, i-1 \text{ if } i > 1
\end{aligned} \tag{VI.7}$$

As the method progresses, the number of constraints increases, thereby gradually reducing the feasible space until the final priority level is reached.

The search algorithm employed in this work is a large-scale nonlinear constrained optimization algorithm called Sequential Quadratic Programming (SQP)^{8,9}. This method models non-linear-constrained optimization problems by quadratic programming sub-problems and searches with directions obtained from the minimization of sub-problems. The SQP implementation used in this study, which employs a BFGS reduced-Hessian algorithm^{10,11} and an automatic differentiation (AD) algorithm called ADOL-C^{12,13}, has been discussed in detail by Jee *et al.*¹²

VI.A.2.3. Treatment Planning and Case Studies

All treatment planning and optimization in this work was performed using the in-house software packages, UMPlan and UMOpt^{14,15}. UMOpt supports the use of a variety of objectives in the inverse planning cost function including dose, dose-volume, biological, and probability-based goals¹⁶. The ADS method is applied in UMOpt as an objective in which the square of the difference between the original and ADS beams is minimized. Lexicographic ordering has been implemented as an optimization method in UMOpt that utilizes the SQP search algorithm. All plans are optimized for delivery by a linear accelerator with 120 multi-leaf collimator and sequenced for step and shoot delivery by a method based on work by Bortfeld *et al.*¹⁷

To evaluate the application of an adaptive diffusion smoothing (ADS) costlet as the final priority in an IMRT optimization problem solved using the Lexicographic Ordering (LO) method, several cases have been optimized and evaluated by these methods. First, a modified version of the National Cancer Institute's (NCI) IMRT benchmark geometry¹⁸ (also used in Chapter II) shown in Figure VI.1 was used to characterize this combined method to evaluate (i) the potential to reduce modulation as a final optimization priority without compromising the preceding objectives and (ii) the possibility to make controlled tradeoffs by exploiting the advantages of the LO planning method. In the latter, we designed tradeoff scenarios that will be used to evaluate the

amount of smoothing possible when allowing for difference concessions in the optimized plan.

Table VI-1 lists the prioritized objective function and tradeoff scenarios used in the phantom. Shown are the original plan objectives and the relaxations that were allowed in those goals to promote smoothing with ADS. For example, tradeoff scenario (TS) "0" corresponded to simply making the minimization of the ADS penalty the final LO priority. The remaining tradeoff scenarios A-G include relaxing objectives by the given amounts before applying the ADS penalty and re-optimizing. The relaxations denoted as N-N were considered non-negotiable objectives; they were not to be relaxed to promote smoothing. After each of the tradeoff scenarios were optimized with the ADS penalty, the reduction in plan MU was analyzed to determine which scenario appeared to have the best cost-to-benefit ratio.

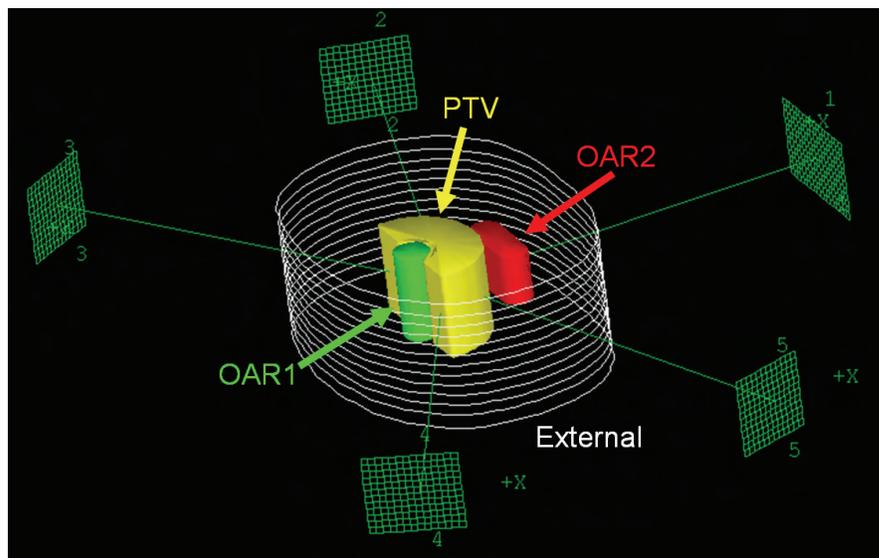


Figure VI-1. The modified IMRT benchmark geometry and beam arrangement used to study the initial characterization of adaptive diffusion smoothing and lexicographic ordering.

The motivation for applying ADS in the LO method is to not interfere with any high priority objectives and to not worsen any objectives that are already past their intended limits. Therefore, in clinical practice, we would apply several tradeoff rules to maintain plan integrity: (1) If a structure is also past its limits, no relaxation will occur unless explicitly called for, and (2) If a structure has a mean dose objective, but ends up below that objective due to the overall dose minimization step, then relaxation of the

mean dose can occur as given in the tradeoff scenario, but only up to the original limit. These rules are intuitive but are written here because future automation of the tradeoff scenario planning will require explicit instructions on each of the allowed tradeoffs.

In order to evaluate the potential clinical application of using ADS with LO to promote controlled smoothing tradeoffs, we applied this method to prostate and head/neck clinical examples. These sites were chosen because of their prevalence as IMRT sites and different levels of geometric and cost function complexity. Both clinical examples were optimized according to the University of Michigan's current IMRT standard for that treatment site. The prioritized objective functions and clinical tradeoff scenarios used for each site are given in Tables VI-2 and 3. As in the phantom, the ADS penalty was applied as the final priority in the lexicographic method to determine the delivery efficiency improvement possible with no sacrifice in the baseline IMRT plan. The clinical tradeoff scenarios in Tables VI-2 and 3 were then optimized to determine the amount of smoothing possible with the given concessions in the objectives. The clinical tradeoff scenarios for each site were designed by a radiation oncologist, based on her clinical knowledge of which cost function objectives could be relaxed in an effort to improve delivery efficiency. The least important objectives are relaxed first, followed by more important objectives. Certain objectives, such as the maximum dose to the spinal cord, were considered non-negotiable, and were not to be relaxed. Therefore, the tradeoff scenarios only included realistic tradeoffs that would be allowed to promote delivery efficiency and improved delivery accuracy. After each of the tradeoff scenarios were optimized and the treatment delivery efficiency gains were obtained, the results were presented to the radiation oncologist to choose which tradeoff scenario would be applied in clinical practice.

VI.A.3. Results

VI.A.3.1 Phantom

The presentation of the phantom results uses the tradeoff scenario identifier given in Table VI-1. Tradeoff scenario "0" corresponds to the smoothing possible with no concession in plan quality. The percent monitor unit reductions relative to the baseline plan are shown in Figure VI-2. Here, the MU reduction possible with equivalent baseline plan quality was 18.4 %. Allowing the OAR2 mean dose to return to its 25 Gy limit was able to reduce the MU by 28.7 % while holding all of the other baseline objectives constrained. Allowing further relaxations in TSH increase the MU reduction to 41.7 %.

Table VI-1. Prioritized Objectives and Clinical Tradeoff Scenarios (Allowed Relaxations to Promote Beam Smoothing and Delivery Efficiency) for the Phantom Prioritized Cost Function

Priority	Structure	Goal	Tradeoff Scenarios (Allowed Relaxations Shown)												
			A	B	C	D	E	F	G	H	H				
1	OAR1	% vol > 36 Gy = 5	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	PTV	Max < 72 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	OAR1	Max < 72 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	OAR2	Max < 72 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	Normal	Max < 72 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
2	PTV	Maximize vol > 60 Gy	0	0	0	0	2 % Vol	0	0	0	0	0	0	2 % Vol	
3	OAR2	Mean < 25 Gy	0	0	0	1 Gy	0	0	0	0	1 Gy	0	1 Gy		
4	OAR1	Minimize Mean	0	0	2 Gy	0	0	0	0	2 Gy	2 Gy	0	2 Gy		
	OAR2	Minimize Mean	0	0	Up to 25 Gy	0	0	0	0	Up to 25 Gy	Up to 25 Gy	0	Up to 25 Gy		
	Normal	Minimize Mean	0	2 Gy	0	0	0	0	0	2 Gy	2 Gy	0	2 Gy		

Table VI-2. Prioritized Objectives and Clinical Tradeoff Scenarios (Allowed Relaxations to Promote Beam Smoothing and Delivery Efficiency) for Prostate

Priority	Structure	Goal	Tradeoff Scenarios (Allowed Relaxations Shown)												
			A	B	C	D	E	F	G	H	I	J	K		
1	PTV	Min 77.7 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
		Max 83 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
2	Rectum	% vol > 70 Gy = 25 %	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
		% vol > 50 Gy = 50 %	0	0	0	0	0	0	0	5%	0	0	0	5%	
3	Rectum	Max < 77.7 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	Bladder	Max < 77.7 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	Femora	Max < 50 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	Penile Bulb	Max < 77.7 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
4	Normal Tissue-within 2 cm of PTV	Max < 80 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
		Max < 50 Gy	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
5	Rectum	Minimize Mean Dose	0	0	0	0	0	0	0	0	0	1 Gy	0	0	1 Gy
	Penile Bulb	Minimize Mean Dose	0	0	0	0	0	0	0	1 Gy	0	0	0	0	1 Gy
	Femora	Minimize Mean Dose	0	0	0	0	0	0	2 Gy	0	0	0	0	0	2 Gy
	Bladder	Minimize Mean Dose	0	0	0	2 Gy	4 Gy	0	0	0	0	0	0	4 Gy	4 Gy
	Normal Tissue	Minimize Mean Dose	0	2 Gy	4 Gy	0	0	0	0	0	0	0	0	4 Gy	4 Gy

Table VI-3. Prioritized Objectives and Clinical Tradeoff Scenarios (Allowed Relaxations to Promote Beam Smoothing and Delivery Efficiency) for Head/Neck

Priority Structure	Prioritized Objective Function	Tradeoff Scenarios (Allowed Relaxations Shown)													
		0	A	B	C	D	E	F	G	H	I	J	K		
1	Cord	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	Max < 45 Gy														
	Cord + 5 mm	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	Max < 50 Gy														
2	Brainstem + 5 mm	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	Max < 54 Gy														
	PTV70	0	0	0	0	0	0	+10 %	0	0	0	+10 %	0	+10 %	+10 %
	64 Gy (-1 % / +7 %)	0	0	0	0	0	0	+10 %	0	0	0	+10 %	0	+10 %	+10 %
3	PTV60	0	0	0	0	0	0	+10 %	0	0	0	+10 %	0	+10 %	+10 %
	60 Gy (-1 % / +7 %)	0	0	0	0	0	0	+10 %	0	0	0	+10 %	0	+10 %	+10 %
	Mandible	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	Max 70 Gy														
4	Contralateral Parotid	0	0	0	0	0	0	0	0	0	0	0	0	0	2 Gy
	Mean < 24 Gy														2 Gy
	Ipsilateral Parotid	0	0	0	0	0	0	0	0	0	0	0	0	0	2 Gy
	Mean < 24 Gy														2 Gy
	Contralateral Submandibular	0	0	0	0	0	0	0	0	0	0	0	0	0	2 Gy
	Mean < 24 Gy														2 Gy
5	Ipsilateral Submandibular	0	0	0	0	0	0	0	0	0	0	0	0	0	2 Gy
	Mean < 24 Gy														2 Gy
	Larynx	0	0	0	0	0	0	0	0	0	0	0	0	0	2 Gy
	Mean < 50 Gy														2 Gy
6	Pharyngeal Constrictors	0	0	0	0	0	0	0	0	0	0	0	0	0	2 Gy
	Mean < 50 Gy														2 Gy
	Esophagus	0	0	0	0	0	0	0	0	0	0	0	0	0	2 Gy
	Mean < 45 Gy														2 Gy
7	Non-involved Oral Cavity	0	0	0	0	0	0	0	0	0	0	0	0	0	2 Gy
	Max 70 Gy														2 Gy
8	Normal Tissue	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N	N-N
	Max 70 Gy														N-N
9	All Normal Structures	0	0	0	0	0	0	0	0	0	0	0	0	0	2 Gy
	Minimize Mean Dose														2 Gy
10	Normal Tissue	0	2 Gy	4 Gy	0	0	0	0	0	0	0	0	0	0	4 Gy
	Minimize Mean Dose														4 Gy

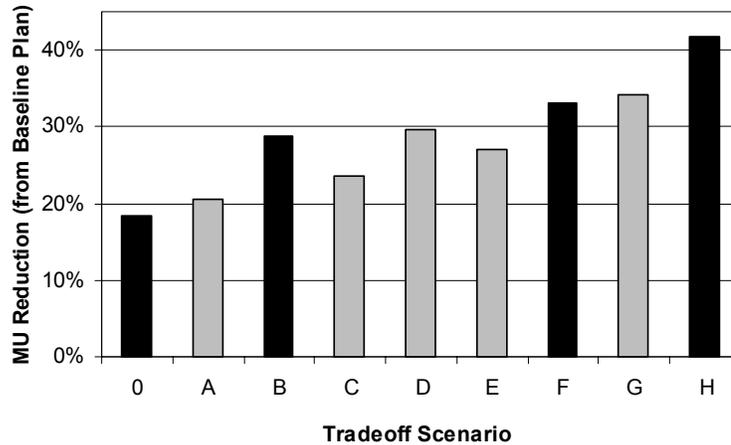


Figure VI-2. The percent reduction in monitor units possible in each of the tradeoff scenarios (corresponding to Table VI-1) relative to the baseline geometry for the phantom case. The DVHs and beams for the tradeoff scenarios denoted in black are shown in Figure VI-3.

To illustrate the differences between plans at various tradeoff scenarios, Figure VI-3 shows the dose-volume histograms and intensity modulated beamlet patterns for the baseline plan and the tradeoff scenarios shaded in black in Figure VI-2. The smoothing of the beams is apparent when going from the Baseline plan to TS0, and is even more pronounced in the remaining tradeoff scenarios with a noticeable change in the beamlets covering the overlap between OAR2 and the PTV. The large MU reduction possible by relaxing the OAR2 mean dose is achieved by this slight change in the beamlet patterns; this was most noticeable in beams 3 and 4.

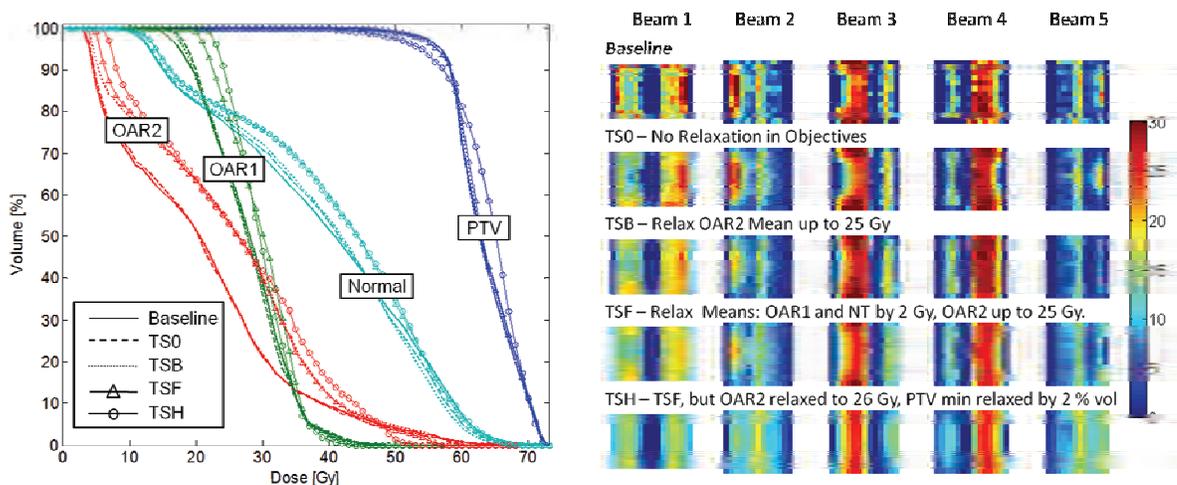


Figure VI-3. The optimized dose-volume histograms and beams for the baseline plan and tradeoff scenarios 0, B, F and H for the phantom case shown in Figure VI-1.

VI.A.3.2 Prostate

The percent monitor unit reductions relative to the baseline plan that were possible in each of the tradeoff scenarios are shown in Figure VI-4. The tradeoff scenarios correspond to the relaxations given in Table VI-2 for the prostate cost function. Tradeoff scenario 0 demonstrated that a 23 % reduction in MU was possible for this patient with no loss in clinical quality. The tradeoff scenarios that allowed an increase in mean dose

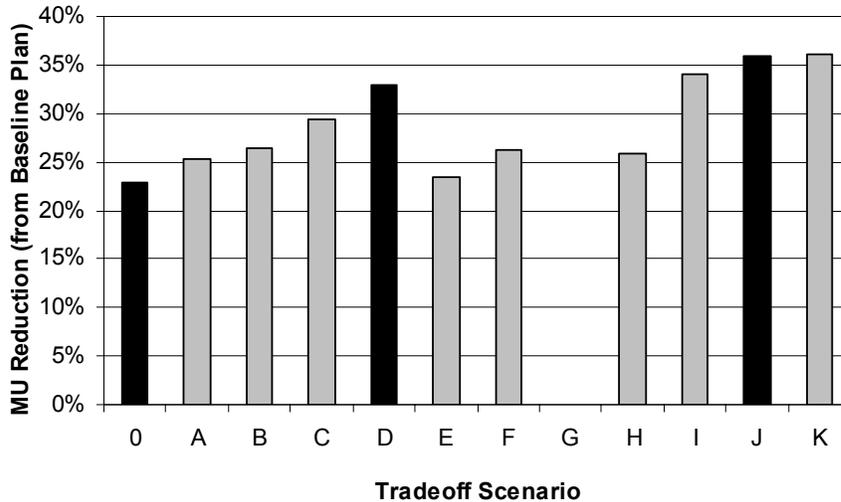


Figure VI-4. The percent reduction in monitor units possible in each of the tradeoff scenarios (corresponding to Table VI-1) relative to the baseline geometry for the prostate case. Tradeoff scenario G was not applicable because the rectum dose-volume constraint was considerably below the objective. The DVHs and beams for the tradeoff scenarios denoted in black are shown in Figures VI-5 and VI-6.

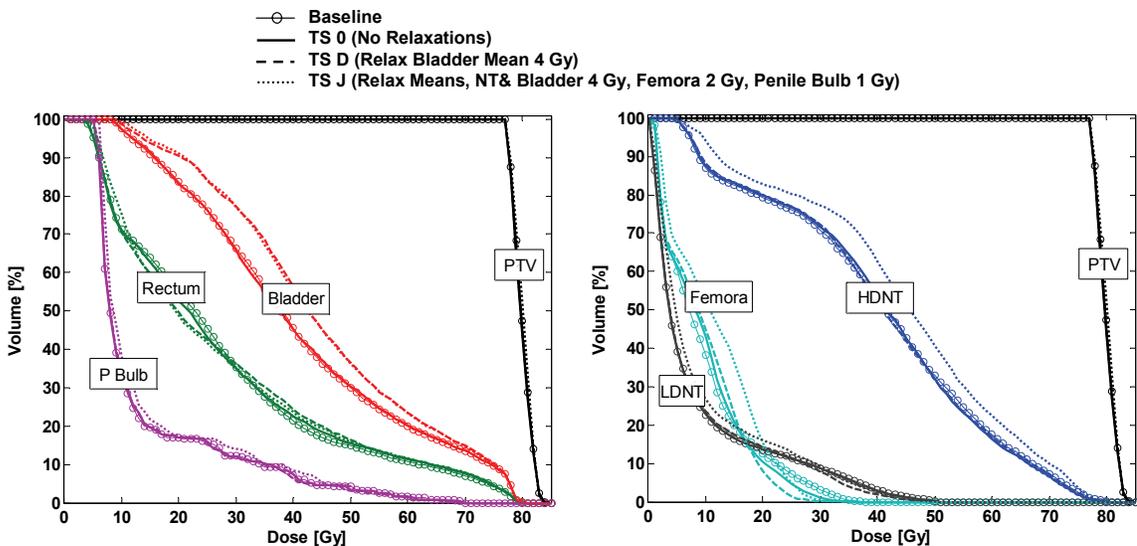


Figure VI-5. The optimized dose-volume histograms for the baseline plan and tradeoff scenarios 0, D, and J. The DVHs are shown in two panels for viewing purposes.

to the bladder of 4 Gy required 33 - 36 % fewer MU to deliver than the baseline. To illustrate the changes in the plan DVHs and beams as the various objectives are relaxed, we have plotted the baseline plan, and tradeoff scenarios “0”, “D”, and “J” in Figures 5 and 6. In TS0, where the baseline objectives are constrained before the use of ADS, the DVHs look very similar with only slight shape changes in the normal tissues.

This shows that slightly different solutions are being found that maintain the same normal tissue mean or maximum doses achieved in the baseline plan. The ADS beams

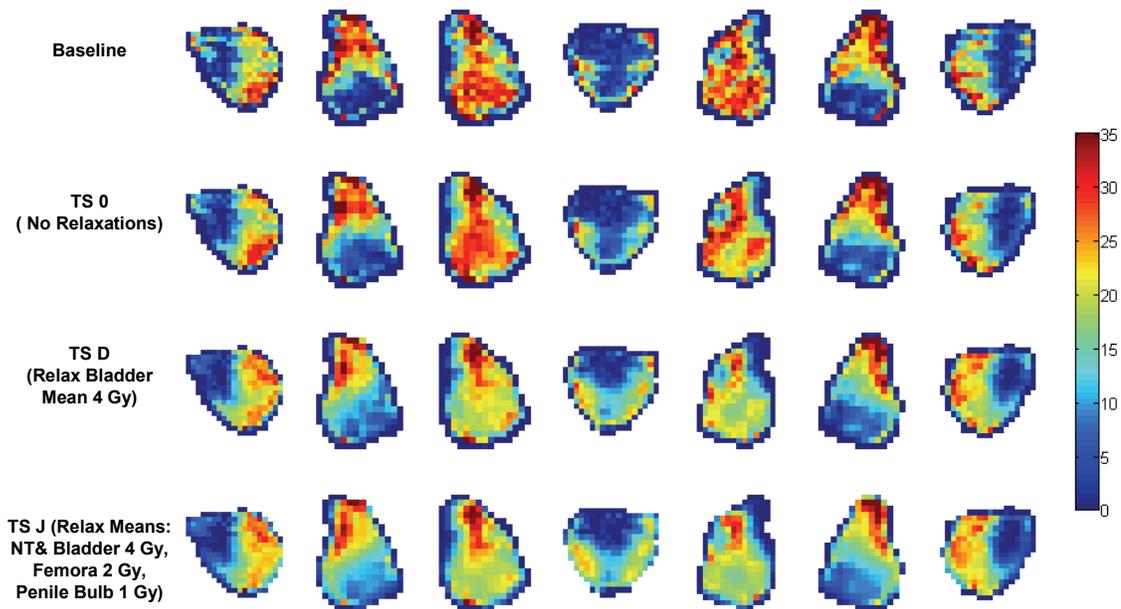


Figure VI-6. The optimized beams for the baseline plan and tradeoff scenarios 0, D, and J. The corresponding DVHs are shown in Figure VI-5.

are noticeably smoother with a reduction in the random appearing “noise” in each of the fields compared to the baseline fields. As the objectives are relaxed in TSD and TSJ, the beams become progressively smoother, and the DVHs clearly show that only the chosen relaxed objectives are affected in each tradeoff scenario.

VI.A.3.3 Head/Neck

The percent monitor unit reductions relative to the baseline plan that were possible in each of the tradeoff scenarios are shown in Figure VI-7. The tradeoff scenarios correspond to the relaxations given in Table V-3 for the head/neck cost function. Similar to the prostate results, a 21.6 % reduction in MU was possible for this patient with no concessions in the baseline plan quality, with the addition of the ADS penalty and

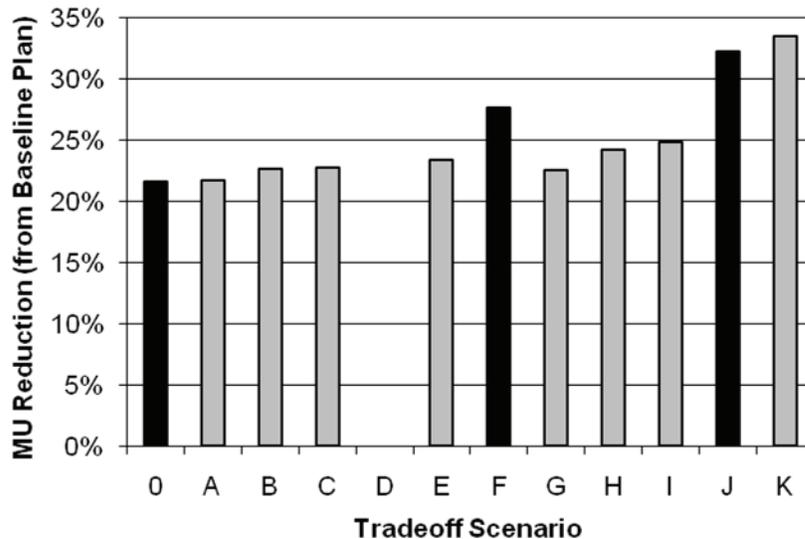


Figure VI-7. The percent reduction in monitor units possible in each of the tradeoff scenarios (corresponding to Table VI-3) relative to the baseline geometry for the head/neck case. Tradeoff scenario D was not applicable because the oral cavity mean dose was already past its intended limit. The DVHs and beams for the tradeoff scenarios denoted in black are shown in Figures VI-8 and VI-9.

Relaxations in the lower priority objectives in the tradeoff scenarios did not result in further large reductions of MU until the PTV maximum dose was relaxed from + 7 % to +10 % of the prescription dose. This resulted in an MU reduction of 27.6 %, and the final composite tradeoffs that included the PTV maximum dose relaxations had MU reductions of up to 33.5 %. The apparent lack of freedom to reduce MU demonstrates the complicated nature of IMRT planning for the head/neck and shows that the solution space can be greatly reduced when adding constraints into the plan using LO.

A comparison of the DVHs for the baseline plan with TS0, TSF, and TSJ is shown in Figure VI-8. An interesting feature of the baseline vs. ADS plans is a difference in shape of the normal tissue DVHs. In the ADS plans, the normal tissue DVHs show that higher volumes of normal tissue are treated to low doses, but smaller volumes received high doses. The beam comparisons are given in Figure VI-9 and demonstrate the improvement that can be made in the plan smoothness by simply using the ADS penalty. The ADS beams are not only smoother, but have fewer isolated high intensity beamlets as well as less “noise”. These features improve further as the PTV maximum doses are relaxed.

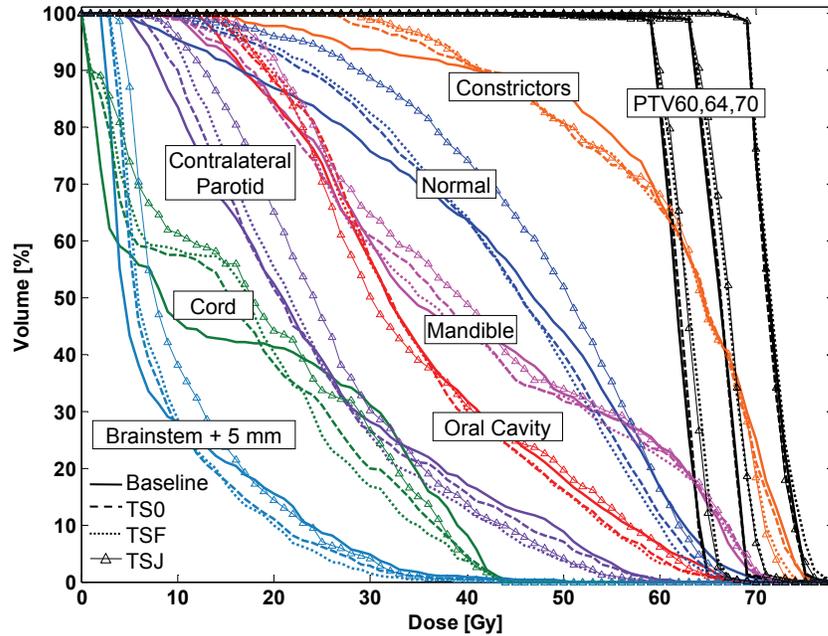


Figure VI-8. The optimized head/neck dose-volume histograms for the baseline plan and tradeoff scenarios 0, F, and J. Not all structures are shown for easier viewing although those not pictured showed similar trends.

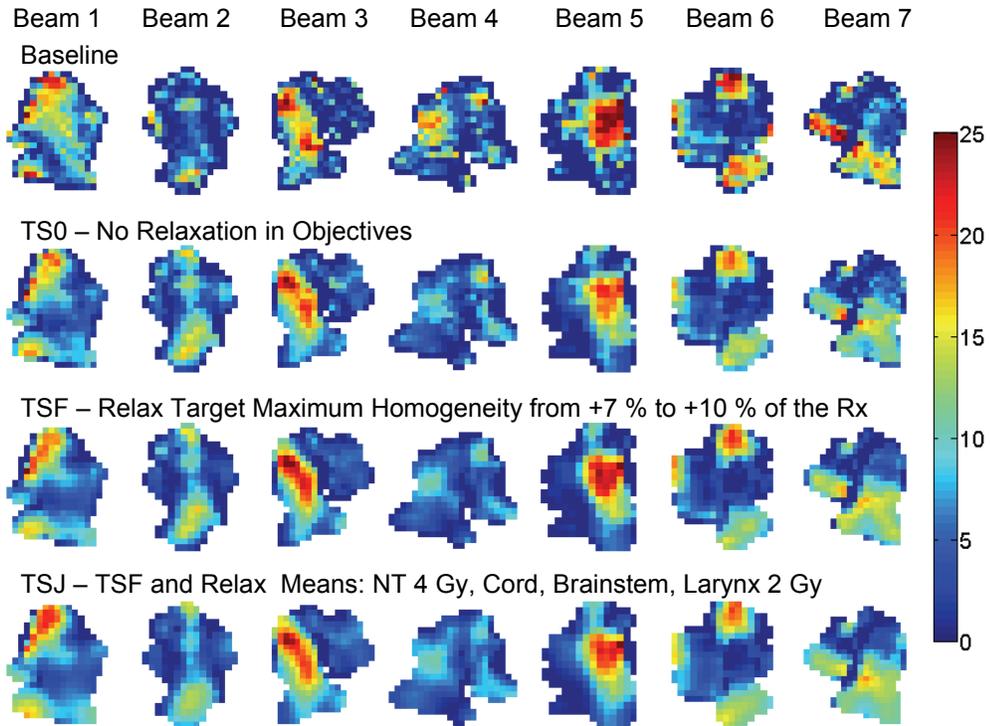


Figure VI-9. The optimized head/neck beams for the baseline plan and tradeoff scenarios 0, F, and J. The corresponding DVHs are shown in Figure VI-8.

VI.A.4. Discussion

This section work demonstrates that a significant amount of complexity reduction is possible without compromising the dosimetric results of the original IMRT plan. However, if one wishes to smooth further, a debate can be made about which planning objectives should be relaxed to allow for smoother beams that can delivered more efficiently and accurately. The answer is that the site and size of tradeoffs allowed are a function of the specific case geometry, patient characteristics, and physician preferences. Therefore, efficient tools are necessary to be able to calculate and present the tradeoffs in question. Generation of tradeoff choices is similar to the generation and navigation of Pareto solutions done with weighted sum cost functions. Recent work by Craft *et al.* has shown that Pareto-optimal plan libraries that make tradeoffs between monitor units and plan objectives can be constructed³. However, with LO, the quantitative changes allowed in each structure can be specified before optimization instead of calculated afterwards based on the optimization results. Therefore, LO may potentially save time and resources because the number of tradeoff instances can be minimized based on the preferences of the physician. A similar reduction in solution space could be achieved by using multiple constraints in the generation of the Pareto solutions. Still, one of the main positive aspects of LO is the use of priority levels to obtain the baseline line without iterating through different importance factors or creating multi-dimensional Pareto surfaces that must be navigated. If an IMRT protocol is not defined well by priority levels and the initial plan is optimized using a weighted sum cost function, the LO method can still be used after obtaining the weighted sum solution to generate tradeoff scenarios with smoothing or other objectives.

The process we have laid out in this work is a starting point for including smoothness criteria into clinical IMRT protocols in an intelligent and robust way. The design of lexicographic cost functions and tradeoff scenarios with the aid of a physician takes a minimal amount of time and should be able to be used for all cases in a protocol. If the tradeoff scenarios are agreed on by several physicians, then they could cover all physician preferences and allow a large, but manageable library of solutions to choose from. Without the use of LO, individual importance factors for both the plan objectives and smoothing criteria would have to be iteratively searched to find the optimal combination of weights for each individual case and geometry, making it difficult to say which tradeoffs would or would not be allowed to promote smoothing in the context of a weighted sum cost function. Another advantage of this method is that the important

question ‘How much can the plan be smoothed without sacrificing plan quality?’ can be posed and answered. Using LO to answer this question means that even if none of the baseline objectives are negotiable, we can achieve a certain amount of smoothing. The examples shown in this work and previous experience with this method suggest that this amount of “free” smoothing can result in approximately 15 - 25 % reduction in MU for the vast majority of IMRT cases.

VI.B. Minimizing Beam Modulation with ADS: A Retrospective Study in Partial Breast IMRT

The clinical significance of improving IMRT delivery efficiency can come from a variety of sources, including improved delivery accuracy, shorter delivery times, less accelerator wear-and-tear, and decreased leakage dose to the patient. The latter can become an even more important factor when risks of developing a secondary cancer are higher for a certain patient population. For example, many lower-risk breast cancer patients are young women with long life expectancies after treatment, so keeping any extraneous dose to a minimum is an important goal. The young age of many breast cancer patients also makes accelerated breast conserving therapy a choice that is rising in popularity, due to both the breast conservation and quicker treatment course. In part due to this increase in convenience for patients, accelerated partial breast irradiation (APBI) is under investigation at many institutions for early-stage breast cancer treatment. A previous study at our institution has shown that a strict IMRT protocol can improve target coverage and reduce normal tissue dose compared to 3D conformal therapy¹⁹. Despite these advantages, treatment providers must also be aware of the increased delivery time and monitor units required for IMRT. Unfortunately, the increased number of monitor units delivered by APBI IMRT treatments can increase the leakage dose to the patient and could possibly increase the risk of induced secondary cancers^{20,21}. The purpose of the following study was to evaluate the use of the adaptive diffusion smoothing (ADS) method during optimization to improve delivery efficiency while preserving target coverage and normal tissue sparing. By reducing the MU (and thus leakage dose) for accelerated partial breast IMRT, the secondary cancer risk may be lessened while also improving the efficiency and accuracy of IMRT QA and delivery.

VI.B.1. Materials/Methods

To evaluate the impact of applying ADS during optimization of APBI cases, IMRT plans were generated retrospectively for 10 patients treated on an IRB-approved protocol at the University of Michigan. Under this protocol, 38.5 Gy is delivered in 10 fractions with strict homogeneity requirements and conservative normal tissue dose objectives. The protocol objectives are given in Table VI-4. After optimization with the protocol objectives to establish a baseline plan, three plans were optimized with ADS. First, the baseline objectives were constrained and ADS was applied to determine the delivery

Table VI-4. Accelerated Partial Breast IMRT Plan Objectives

Structure	Objective
Left Anterior Descending Coronary Artery	< 1 % Vol > 3 Gy
Heart	< 1 % Vol > 3 Gy
CTV	Min Dose > 38.5 Gy < 1 % Vol > 40.4 Gy
PTV	> 95 % Vol > 38.5 Gy < 1 % Vol > 40.4 Gy
Ipsilateral Lung	< 10 % Vol > 5 Gy
Contralateral Breast	< 1 % Vol > 5 Gy
All Normal Tissue	Minimize Overall Dose

efficiency improvement possible with no loss in plan quality. Next, the minimum dose objectives to the CTV and PTV were relaxed by 3 % to 37.4 Gy and 5 % to 36.6 Gy to determine the further smoothing possible with small concessions in strict homogeneity requirements. No relaxation was made in the maximum dose because of the already high doses being given in the hypo-fractionation protocol. The controlled target relaxation was done using a gradient search algorithm that allows constraints, sequential quadratic programming⁸. This method is discussed in more detail in the previous section with respect to its use in lexicographic ordering. The plan metrics, beam complexity, and delivery efficiency for all plans were analyzed.

VI.B.2. Results

The static delivery MU/fraction required for all patients and plan techniques are shown in Figure VI-10, and the means and standard deviations for the MU are shown in Table VI-5 relative to the baseline plan. The ADS plans required substantially fewer MU in all cases, and relaxing the target minimum dose further reduced MU by varying degrees

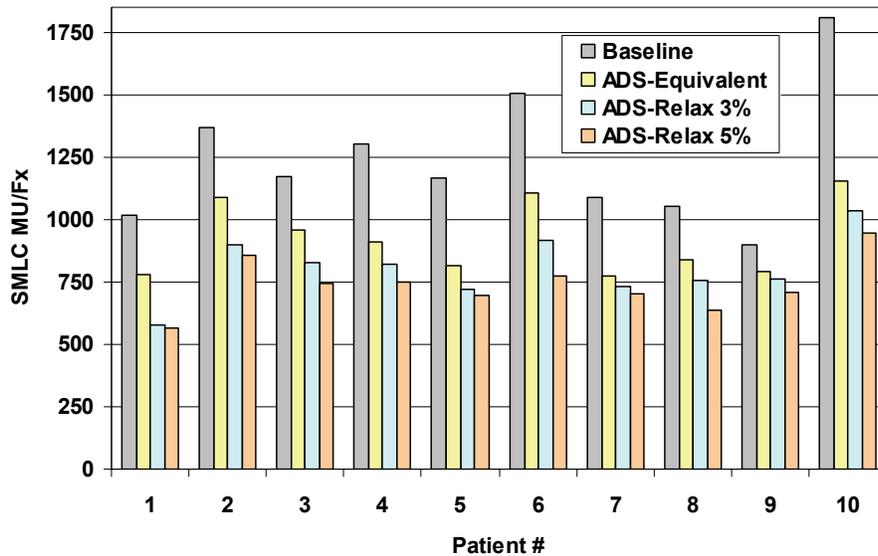
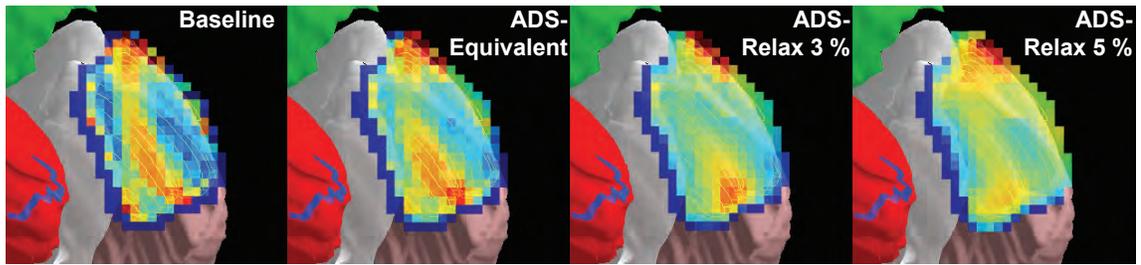


Figure VI-10. Required MU/Fraction for static delivery for each of optimization methods for each of the 10 breast patients.

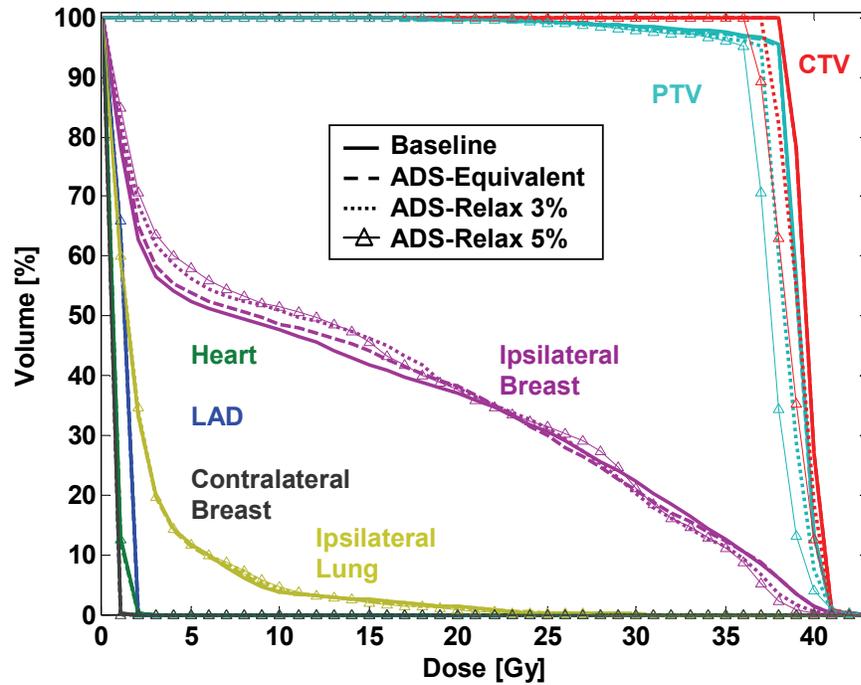
depending on the patient. The MU reduction possible when applying ADS was 24.6 +/- 7.1 % with no loss in plan quality, 34.2 +/- 8.3 % with a 3 % relaxation in the minimum target dose, and 39.4 +/- 7.7 % with a 5 % relaxation in the minimum target dose. All differences observed in the other plan metrics were negligible, with the exception of the

Table VI-5. Relative mean monitor unit requirements and target dose objectives for accelerated partial breast IMRT with and without an adaptive diffusion smoothing penalty used during optimization.

Plan Description	Minimum Target Dose Objective	Maximum Target Dose Objective	Relative Number of MU	
			Mean	σ
Baseline	38.5 Gy	40.4 Gy	1.000	0.000
ADS Equivalent	38.5 Gy	40.4 Gy	0.754	0.071
ADS Relax 3 %	37.4 Gy	40.4 Gy	0.658	0.083
ADS Relax 5 %	36.6 Gy	40.4 Gy	0.606	0.077



(a)



(b)

Figure VI-11. (a) A typical intensity modulated beam in the beam's eye view for the Baseline, ADS-Equivalent, ADS-Relax 3% and ADS-Relax 5% cases with the breasts (brown and green), lung (white), heart (red), and LAD (blue) shown. The beam is covering the CTV and the PTV (wireframe). (b) the corresponding DVHs of each of the plans.

minimum target dose in the final two ADS plans.

Figure VI-11 shows, for a typical case, the change in DVHs and one of the intensity modulated beams in each of the four scenarios tested. The improvement in smoothness with ADS is apparent and comes with no significant changes to any of the other structures, except for a slight change in the shape of the normal breast DVH. In Figure VI-11(A), the Baseline plan required 1503 MU/Fx due to the large target size and small beamlet dimensions. The ADS plans required 1109, 914, and 773 MU/Fx for the

Equivalent, Relax 3%, and Relax 5% plans, respectively. The reduction in MU after the initial addition of the ADS penalty may be due in part to the reduction of isolated high intensity beamlets scattered in the field. It is important to note that the intensity gradients around the edge of the field where the beam intersects the ipsilateral lung are preserved in all ADS plans. It is also interesting to observe the high gradient area near the top right of the field, which may be due to a lack of build-up dose near the patient surface.

VI.B.3. Summary

Concerns over the potentially significant leakage dose associated with the high number of MU in accelerated partial breast IMRT treatment has led to an investigation of the use of the adaptive diffusion smoothing penalty during APBI IMRT optimization. In each of the ten patients studied, the use of the ADS penalty was able to significantly reduce MU without affect the dosimetric quality of the tightly constrained plans. The mean reduction in MU when using ADS was 24.6 +/- 7.1 % with no sacrifice in plan quality. Larger MU reductions were possible with slight concessions in target homogeneity.

The use of adaptive diffusion smoothing during optimization of accelerated partial breast IMRT cases allows the full dosimetric benefit of intensity modulation to be realized without the substantial loss in delivery efficiency that may be observed with conventional IMRT, and clinical use of adaptive diffusion smoothing during optimization should result in reduced leakage dose to patients and shorter, more accurate treatment deliveries.

VI.C. Use of ADS in Radiobiological Targeting

Radiotherapy is entering an era of image-guidance and adaptation. Two integral components in this new treatment paradigm are functional and biological imaging. As these two entities become more prevalent and more reliable, radiobiologic targets and avoidance structures will become increasingly common. With this likely increase in geometric complexity, intensity-modulated radiation therapy (IMRT) will be one of the preferred treatment modalities. Because of its ability to produce very heterogeneous dose distributions, inverse-planned IMRT presents a unique opportunity in the targeting and avoidance of radiobiologically-defined regions. Critical target regions may represent aggressive, rapidly proliferating, or radioresistant disease, and IMRT can allow for dose escalation to these volumes while still sparing normal tissue. Several recent studies

have shown that inverse-planned IMRT has the ability to achieve the heterogeneous dose and fluence patterns necessary to treat targets defined by various radiobiological imaging studies²²⁻²⁵. However, these studies also point out that using functional imaging data is not a trivial matter, and that systematic and random errors in defining the radiobiological regions can result in worse tumor control probabilities (TCPs) when using heterogeneous rather than uniform dose distributions²². If this potential underdosing is combined with geometric uncertainty from setup errors, the risk of creating cold spots in critical target regions becomes even greater²⁶. Also, the more complex IMRT beams become, the longer and more difficult the quality assurance and eventual treatment delivery become. Increases in treatment time can potentially increase dosimetric errors due to increased intrafraction motion, which may impact tumor control due to intrafraction tumor repair^{27,28}. Both of these issues were discussed in Chapter II.

In the following section, we report our study of the hypothesis that more homogeneous fluence patterns over critical radiobiologic targets will reduce sensitivity to geometric uncertainties and thus decrease the amount of underdosing that occurs in the presence of daily setup errors. This section demonstrates the feasibility of using adaptive diffusion smoothing during optimization to promote smoothness over the beamlets that cover key radiobiologic target regions. This could provide the modulation control necessary to ensure that radiobiologic target volumes receive the intended dose when imaging and geometric uncertainties are present.

VI.C.1. Methods

The adaptive diffusion smoothing process used in this section is outlined in Chapter V. The major difference in the application of ADS smoothing to this problem is that instead of being defined based on the cost function gradients, the ADS diffusion coefficients are defined based upon the radiobiological parameters of the geometry in each beamlet projection. Specifically, the diffusion coefficients are defined to be large in beamlets that cover radiobiological targets, thus promoting smoothness and increasing the probability of having homogeneous fluence patterns over key areas.

As in the previous chapter, the ADS smoothed beam is not automatically applied during the optimization process, but the difference between the ADS beam and the original beam is calculated to construct a penalty which behaves as another optimization goal.

VI.C.1.1 Radiobiologically Defined Diffusion Coefficients

In order to automatically define the diffusion coefficients to promote smoothing over high-risk tumor subvolumes, we must use a metric that can relate the beamlet position to the underlying tumor geometry. One way to delineate the risk-level of the tumor geometry is through a dose evaluation metric called generalized equivalent uniform dose (gEUD) ²⁹. This metric takes a heterogeneous dose distribution from a target or normal tissue structure and represents it by the uniform dose value that would have the same radiobiological effect. The generalized EUD is given by:

$$EUD = \left(\frac{1}{N} \sum_i D_i^a \right)^{\frac{1}{a}} \quad (II.6)$$

where D_i is the dose in the i^{th} voxel of the structure, N is the total number of voxels, and a is a structure-specific parameter. The parameter a is positive for normal tissues and negative for tumors. Its value depends on the properties of the tissue, with $a = -5$

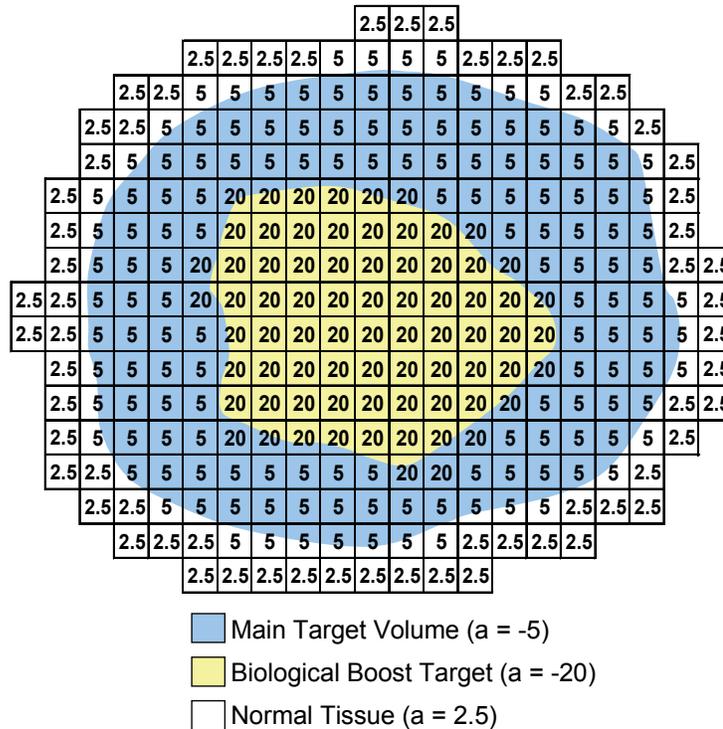


Figure VI-12. Example of how the $|a|$ values are distributed in the beamlets to be used in the diffusion coefficient calculations. The higher $|a|$ structures take priority.

representing a responsive tumor and $a = -20$ representing a resistant tumor. For parallel-behaving normal tissues, a is near 1, and for serial-behaving normal structures, such as spinal cord, a is approximately 8. For this study, we assigned a values to each of the tumor sub-volumes based upon their risk-level and the normal tissues were given a values of 2.5. After assigning a values, the beamlets that projected onto the tumor sub-volumes from each beam's eye view were identified. Then, "beamlet $|a|$ values" were assigned based upon the highest structure $|a|$ values encountered in the projection.

Figure VI-12 shows an example of these effective beamlet $|a|$ values for a beam that covers a tumor with a high-risk sub-volume that will be "boosted" to a higher dose. After effective $|a|$ values are assigned to each beamlet, the diffusion coefficients are calculated using the effective $|a|$ values. In the previous work on ADS presented in Chapter V, the diffusion coefficients were required to be between 0 and 10. To conform to that standard and to promote (i) a high amount of smoothing in the biological target, (ii) moderate smoothing over the main PTV, and (iii) only little smoothing over the rest of the field so as to not require heavy tradeoffs with the other plan objectives, the diffusion coefficients have been defined as

$$D_{ij} = \frac{|a|^{1.5}}{9} \quad (\text{VI.8})$$

This definition of D , chosen based on prior experience and after several trials, means that the beamlets with higher projected $|a|$ values are subject to a higher degree of smoothing than those with smaller $|a|$ values. Thus, ADS will promote smoother, more homogeneous fluence patterns over the areas that are more sensitive to cold spots and underdosing. Minor changes could easily be made to the definition of D to promote relatively more or less smoothing in the different regions. In addition, the formulation provided here could easily be adapted to other functional imaging applications by using different definitions of D .

VI.C.1.2. Test Case

To test the implementation of radiobiologically defined diffusion coefficients in the context of ADS, the scheme was applied in a test case with a main PTV ($a = -5$) and two

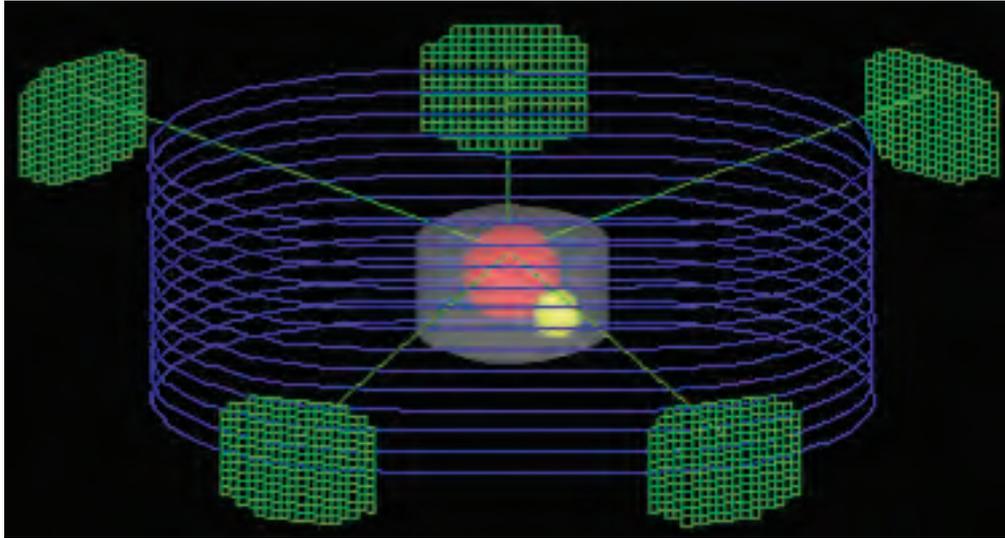


Figure VI-13. The test case geometry and beams. The PTV is gray ($a = -5$), and the boost targets are red ($a = -15$) and yellow ($a = -20$).

smaller boost volumes ($a = -15$ and -20). The case geometry and beam arrangement are shown in Figure VI-13. The normal tissue ($a = 2.5$) used in the optimization was a 2 cm expansion around the PTV. After assigning the diffusion coefficients as discussed above, optimization was performed without ADS, and then with the ADS penalty applied at increasing weights in a weighted sum cost function. The baseline cost function used is described in Table VI-6. Random setup errors were simulated by shifting the plan isocenter 3 mm in each direction, and plans were recalculated at the new positions and then analyzed. Comparisons were made between baseline and ADS plans looking at target gEUD before and after the shifts, plan DVHs, fluence patterns, and SMLC monitor units. In this work, all plans used 6 MV beams with 5 mm x 5 mm beamlets, and optimization was performed with UMOpt using the quasi-Newton search algorithm.

Table VI-6. Test Case Objectives

Structure	Objective
PTV 50	Min Dose > 35 Gy
	< 10 % Vol > 65 Gy
	gEUD (a=-5) > 50 Gy
Boost 70	Min Dose > 55 Gy
	Max Dose < 85 Gy
	gEUD (a=-15) > 70 Gy
Boost 90	Min Dose > 75 Gy
	Max Dose < 105 Gy
	gEUD (a=-20) > 90 Gy
Normal Tissue	Max Dose < 65 Gy
	Mean Dose < 20 Gy
	Min Dose

VI.C.1.3. Clinical Pancreas Example

To demonstrate the clinical potential of this method, ADS was applied in a pancreas example. The preferred treatment for pancreatic cancer is resection, but tumor invasion into the vasculature can inhibit a full resection. One technique that is under investigation to sterilize the tumor-vasculature volume and make the patient eligible for surgical treatment is to boost the dose to the volume with IMRT. In a gEUD-based prescription, this surgical boost volume is prescribed to a higher dose than the rest of the PTV and given an a value of -15 to -20. Sterilization of this target is critical for surgical treatment, and thus, maintaining a high EUD in the face of geometric uncertainties is very important to reaching a desired clinical outcome. To test the hypothesis that ADS will be able to achieve better results than standard IMRT in this situation, we have optimized a pancreas example with both standard and ADS IMRT. The baseline objective function used in the example is described in Table VI-7. The normal tissue objectives are the highest priority in the objective function. Also, the overall dose to all normal tissue is minimized.

Diffusion coefficients were defined according to Equation VI-8 after taking the beamlet $|a|$ projections. The a values given to the PTV, surgical boost, and normal tissue were -5, -20, and 2.5, respectively. Baseline optimization was performed and then the ADS penalty was applied at increasing weights. Plans were sequenced for

Table VI-7. Pancreas Case Objectives

Structure	Objective
Duodenum	Max Dose < 60 Gy < 33 % Vol > 45 Gy
Stomach	Max Dose < 54 Gy < 2 % Vol > 50 Gy < 25 % Vol > 45 Gy
Small Intestine	Max Dose < 54 Gy < 2 % Vol > 50 Gy < 25 % Vol > 45 Gy
Liver	Minimize NTCP
Kidneys	Max Dose < 20 Gy < 10 % Vol > 18 Gy
Cord	Max Dose < 45 Gy
PTV	Min Dose > 45 Gy gEUD ($a=-5$) > 60 Gy < 15 % Vol > 75 Gy
Surgical Boost	Min Dose > 65 Gy gEUD ($a=-20$) > 80 Gy Max Dose < 95 Gy

static delivery, and the required MU were compared. The sensitivity of the surgical boost gEUD to simulated setup errors was analyzed for both techniques using the same procedure as the test case.

VI.C.2. Results

VI.C.2.1. Test Case

The test case was first optimized with the baseline cost function and then with the ADS penalty applied at three increasing weights. The DVHs for these plans are shown in Figure VI-14. Their similarity demonstrates that ADS did not interfere with the baseline objectives. Large changes were observed in the optimized fluence patterns, and Figure VI-15 shows, for a typical beam, the change in intensity modulation when using ADS. The baseline fluence pattern is very heterogeneous, while the ADS beams are smoother

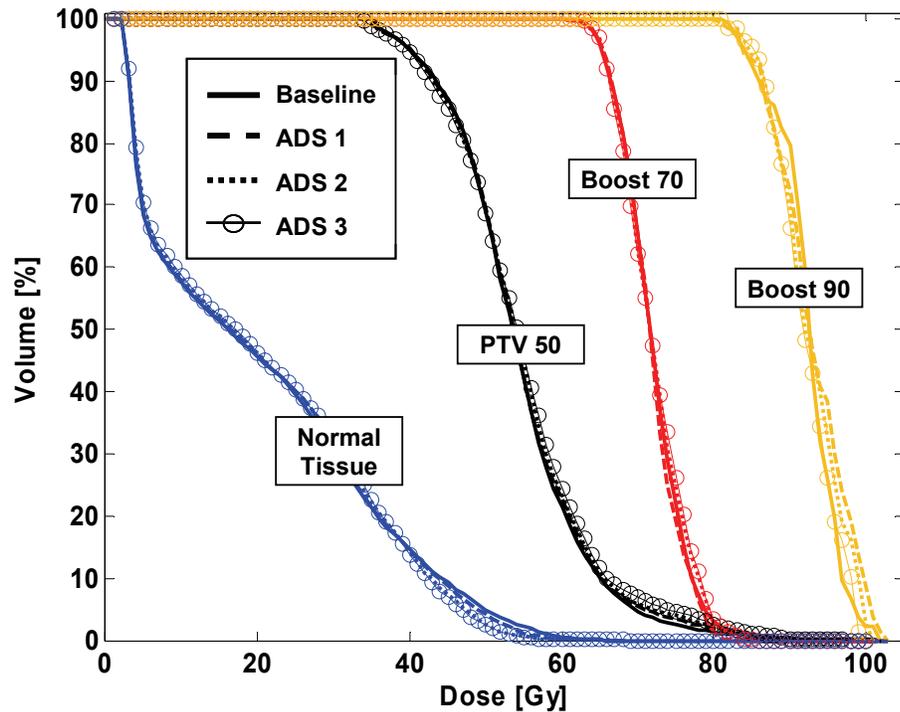


Figure VI-14. Dose-volume histograms for the targets and normal tissues for the baseline and ADS IMRT plans. ADS 1, 2, and 3 denote the relative ADS penalty weight.

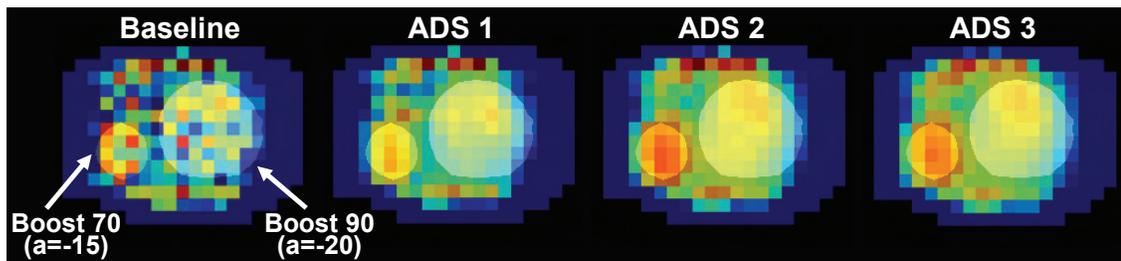


Figure VI-15. A typical beam in the test case as the ADS penalty is increased. The two boost targets are shown. Blue denotes cold beamlets and red denotes hot beamlets.

and more structured. The beamlets intersecting the boost volumes become more uniform as the ADS penalty increases. Also, the ADS plans required 37 - 61 % fewer MU for static delivery. The effect of these more homogeneous beam regions was increased robustness in preserving gEUD in the boost targets after small geometric shifts.

Figure VI-16 shows the gEUD for each target before and after the 3.0 mm isocenter shifts in each direction. The gEUD is relatively unchanged after the y (posterior) shift. However, after shifts in the x (lateral) and z (superior) directions, the gEUD for all targets

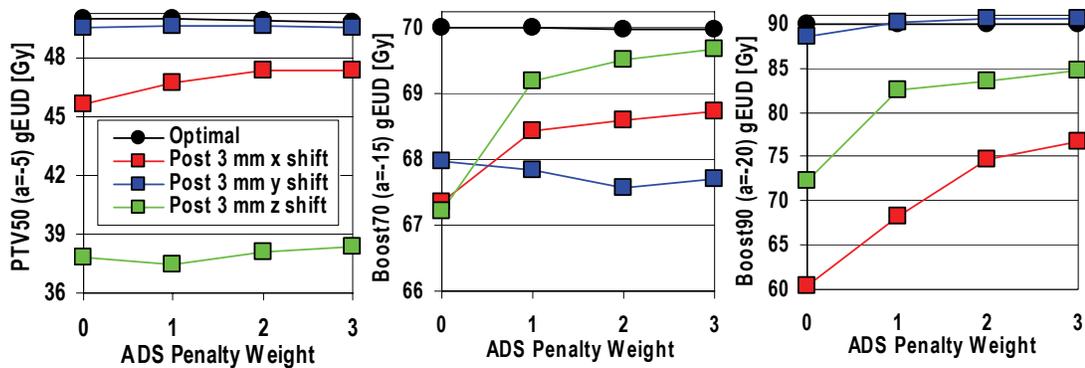


Figure VI-16. The gEUD for PTV50, and the two boost targets (70 and 90) before and after 3 mm isocenter shifts in the x (lateral), y (posterior), and z (superior) directions.

was degraded considerably. This degradation was smaller in the ADS plans, and the post-shift gEUD increased as a function of increasing ADS penalty. The most drastic gEUD losses were observed in Boost90 after x and z shifts, and the use of ADS decreased these losses by 16.4 and 12.5 Gy in the x and z directions, respectively. The use of ADS did not significantly improve the gEUD in PTV50 after the z shift. This was likely due to several high intensity beamlets at the superior regions of the fields moving away from the target. These intense beamlets were also present in the ADS beams because the smoothing in that area was not very strong. The shifts did not affect the mean normal tissue dose, but the maximum post-shift normal tissue dose decreased with increasing ADS penalty.

VI.C.2.2. Clinical Pancreas Example

The pancreas example was optimized with the baseline cost function, and then the ADS penalty was added at increasing weights. Similar to the test case, the ADS penalty did not interfere with the baseline objectives and only slightly changed the optimized DVHs. Figure VI-17 compares DVHs for a moderately weighted ADS plan and baseline IMRT. A positive side effect of ADS smoothing was a reduction in dose to the spinal cord. This was likely due to a decrease in scatter reaching the cord from high intensity beamlets in the standard IMRT cases, which are removed with the ADS smoothing. We note that if the overall dose minimization to the cord was more highly weighted in the cost function, this effect may not have been observed. Both plans are shown in Figure VI-18, and an increase in beam homogeneity is observed in the regions covering the surgical boost in the ADS plan compared to the baseline plan. After 3 mm shifts in each direction, the surgical boost gEUD was degraded, and Table VI-8 gives the pre- and post-shift gEUD in the boost volume for the baseline and ADS plans. In the lateral and superior directions, where setup errors are most common, the improvement in gEUD when using ADS is equal to or larger than one fraction (1.8 Gy). The ADS plan also requires 34 % fewer MU for delivery. The normal tissue objectives were less affected by the shifts than

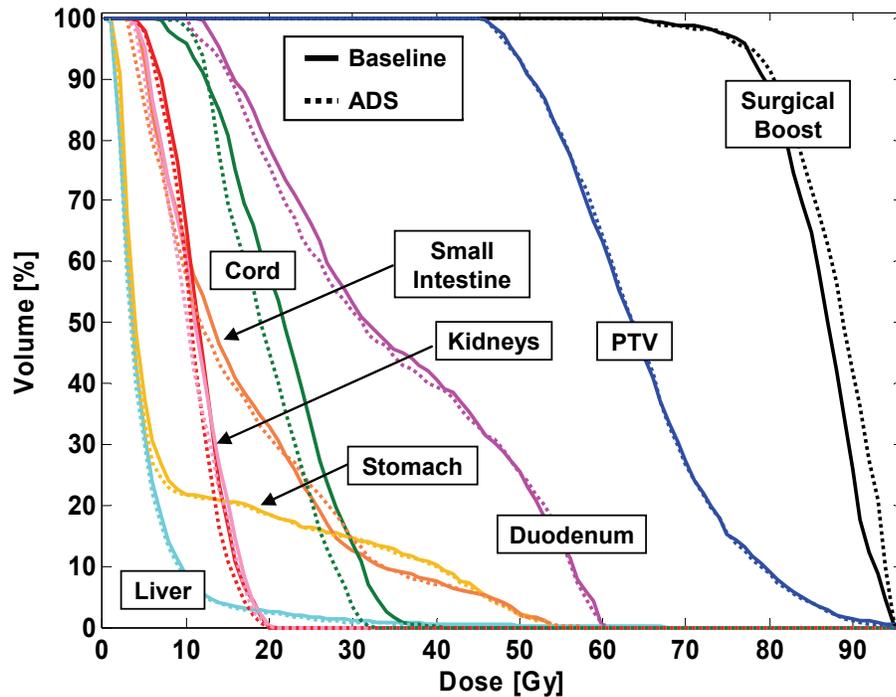


Figure VI-17. Pancreas case DVHs for baseline and ADS IMRT plans.

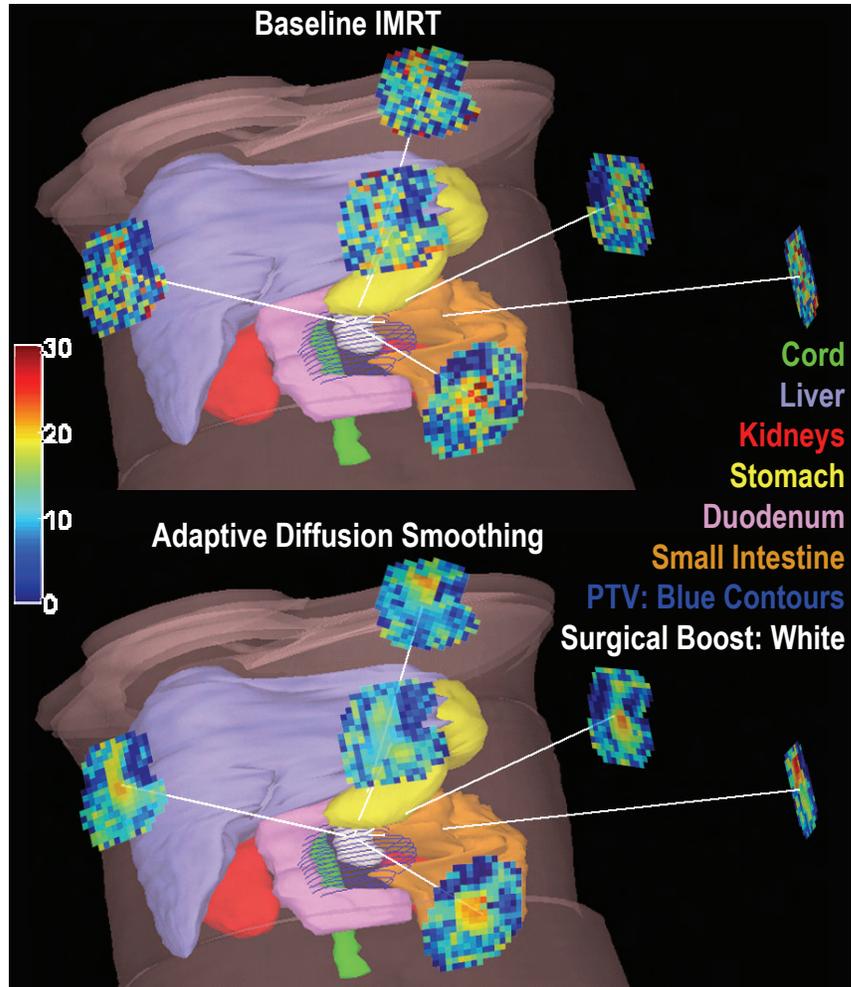


Figure VI-18. 3-D view of baseline vs. ADS IMRT plans in the pancreas.

the targets, and in most cases, were better preserved when using ADS. However, these differences may not be clinically significant.

The effect of ADS in the test case was likely higher than in the pancreas case due to the simpler geometry and axial beams. Still, the gains in delivery efficiency and gEUD in the complex pancreas example were clinically significant and did not risk normal tissue.

Table VI-8. Pancreas surgical boost gEUD [Gy] before and after 3 mm isocenter shifts

Technique	Optimal gEUD before shift	gEUD after 3 mm x (lateral) shift	gEUD after 3 mm y (posterior) shift	gEUD after 3 mm z (superior) shift
Standard IMRT	80.00	75.97	69.95	75.68
ADS IMRT	80.00	79.44	71.37	77.47
Difference		3.5 Gy	1.4 Gy	1.8 Gy

VI.C.3. Summary

The use of IMRT for plans with multiple radiobiologic targets requiring different dose prescriptions is a seemingly good strategy. However, the inherent complexity in IMRT fields can lead to decreased treatment efficiency and potential instability in the face of geometric uncertainties. The concept that modulation can be manipulated to increase delivery efficiency and robustness of plan metrics to geometric uncertainties, while still preserving the merits of IMRT, is worth pursuing. An optimization scheme has been introduced in which smoothness of fluence patterns is promoted in regions over high-risk targets that are susceptible to large losses in gEUD in the presence of small setup errors. The use of adaptive diffusion smoothing with radiobiological-defined diffusion coefficients allowed the preferential smoothing of beamlets that covered critical target regions. The a value of targets in a gEUD formulation was used to define the diffusion coefficients for use in ADS. A test case demonstrated that ADS could alter the fluence patterns—without significantly affecting the plan objectives or DVHs—in such a way that the target gEUD values were more robust to small geometric shifts in the plan. In addition, the use of ADS improved delivery efficiency. ADS was also applied to optimization of a pancreatic case involving a high-risk boost volume. The gEUD in this volume was higher (better) for the ADS plan when compared to standard plans when both were subject to 3 mm simulated setup errors. The improvements in gEUD were 3.5, 1.4, and 1.8 Gy after lateral, posterior, and superior shifts, respectively. We conclude that ADS is a promising method for manipulating IMRT beams to potentially improve outcome in radiobiologic targets.

VI.D. Future Applications of ADS

The potential utility of the adaptive diffusion smoothing method goes far beyond the limited applications shown in this chapter. The power to customize the diffusion coefficients to different situations makes ADS an attractive option for dealing with a variety of issues in intensity-modulated radiation therapy. The use of ADS for smoothing IMRT fields during optimization is a practical and inexpensive—in terms of time and dosimetric quality—way to improve delivery efficiency and reduce the many consequences that accompany beam complexity. The application of ADS to radiobiological-based optimization is an example of how ADS can improve IMRT treatment planning. Future projects in this area could include the extension of the

method to account for smoothing over normal tissues as well as targets. The generalized EUD formulation is designed to predict the EUD for both types of tissues. Therefore, a change in the diffusion coefficient definition to incorporate the use of both positive and negative a values could be a starting point for studying the effect of smoothing over different types of normal tissues (i.e. serial or parallel organs).

One area of concern in IMRT is the possible degradation in plan quality that results from internal patient motion, such as breathing or cardiac motion. If the probable direction and amplitude of motion can be predicted, then diffusion coefficients could be designed to promote smoothing in the direction of motion. This may degrade plans in the short-term, but could make them more robust to the predicted breathing motion. In addition, the dosimetric errors arising from the motion should be smaller than in the unsmoothed case because the intensity gradients will be also be smaller. A similar approach could be taken to incorporate random setup errors into the optimization process.

Radiotherapy often relies on many different forms of imaging to delineate tumor and normal tissue structures. To accommodate the use of imaging data from different modalities, each set of data must be registered with a reference dataset (usually the treatment planning CT scan). The registration process is not error-free and can introduce uncertainty into the contours that are transferred between different data sets. If the location and approximate magnitude of this uncertainty can be predicted, then it may be possible to set the diffusion coefficients to be a function of that uncertainty in an effort to blur the beamlet intensity over those areas. For example, if the tumor boundary is not clear, then it may be advantageous to smooth or blur the beam intensity over the interface of the target instead of having a sharp fall-off at the edge of the delineated tumor. These ideas could be extended to general registration errors as well so that large intensity gradients do not fall over areas that have been poorly registered.

The direct optimization of the segments shapes and weights that are delivered in IMRT is an alternate method that can be used to promote delivery efficiency in IMRT. However, the optimized segments in direct aperture optimization (DAO)³⁰ can still be quite irregular. A possible reason for this is the fact that the fluence map optimization is usually run for several iterations and then sequenced to generate the initial segments. If ADS was applied in this initial step of DAO, it might be possible to start with equivalent, but more regular segments. We show in Chapter VII that segments created from ADS plans are generally much more regular, with larger gaps between the leaves than

conventional IMRT segments. Therefore, the use of ADS in creating initial DAO segments could result in more accurate plans with less tongue-and-groove effect.

Adaptive diffusion smoothing is a unique method that allows for a high level of adaptation and customization. The method can be adapted to a variety of different problems, and inside each of those applications, it can be further individualized for different patient characteristics such as tumor biology and breathing motion. Thus, adaptive diffusion smoothing is a truly adaptive method that may have many future applications in IMRT and even possibly in other areas of radiotherapy.

References

- ¹ K.-W. Jee, D. L. McShan, and B. A. Fraass, "Lexicographic Ordering: Intuitive Multicriteria Optimization for IMRT," *Phys Med Biol* **52**, 1845-1861 (2007).
- ² J. J. Wilkens, J. R. Alaly, K. Zakarian, W. L. Thorstad, and J. O. Deasy, "IMRT treatment planning based on prioritizing prescription goals," *Phys Med Biol* **52**, 1675-1692 (2007).
- ³ D. Craft, P. Suss, and T. Bortfeld, "The tradeoff between treatment plan quality and required number of monitor units in intensity-modulated radiotherapy," *Int J Radiat Oncol Biol Phys* **67**, 1596-1605 (2007).
- ⁴ M. Alber, G. Meedt, F. Nusslin, and R. Reemtsen, "On the degeneracy of the IMRT optimization problem," *Med. Phys.* **29**, 2584-2589 (2002).
- ⁵ J. Llacer, N. Agazaryan, T. D. Solberg, and C. Promberger, "Degeneracy, frequency response and filtering in IMRT optimization," *Phys. Med. Biol.* **49**, 2853-2880 (2004).
- ⁶ J. Llacer, J. O. Deasy, T. R. Portfeld, T. D. Solberg, and C. Promberger, "Absence of multiple local minima effects in intensity modulated optimization with dose-volume constraints," *Phys Med Biol* **48**, 183-210 (2003).
- ⁷ Q. Wu and R. Mohan, "Multiple local minima in IMRT optimization based on dose-volume criteria," *Med Phys* **29**, 1514-1527 (2002).
- ⁸ P. E. Gill, W. Murray, and M. A. Saunders, "SNOPT: An SQP algorithm for large-scale constrained optimization," *SIAM J Optimization* **12**, 979-1006 (2002).
- ⁹ J. Nocedal and S. Wright, *Numerical Optimization*. (Springer-Verlag New York, Inc., 1999).
- ¹⁰ D. F. Shannon, "Conditioning of quasi-Newton methods for function minimization " *Math. Comput.* **24**, 647-656 (1970).
- ¹¹ C. G. Broyden, "The convergence of a class of double rank minimization algorithms " *J. Inst. Math. Appls* **6**, 76-90 (1970).
- ¹² A. Griewank, D. Juedes, H. Mitev, J. Utke, O. Vogel, and A. Walther, "ADOL-C: A Package for the Automatic Differentiation of Algorithms Written in C/C++," *ACM TOMS* **22**, 131-167 (1996).
- ¹³ K.-W. Jee, B. A. Fraass, and D. L. McShan, in *Automatic Differentiation: Applications, Theory, and Tools*, edited by H. M. Bucker, G. Corliss, P. Hovland et al. (Springer, 2005).
- ¹⁴ B. A. Fraass and D. L. McShan, in *The Use of Computers in Radiation Therapy*, edited by I A D Bruinvis, F.H. van der Giessen, H.J. van Kleffens et al. (Elsevier Science Publishers BV, North-Holland, 1987), pp. 273-276.
- ¹⁵ J. H. Kim, N. Dogan, D. L. McShan, and M. L. Kessler, "An AVS-based system for optimization of conformal radiotherapy treatment plans," *Proceedings of 1995 International Advanced Visual Systems User and Developer Conference*, Boston MA, 1995.
- ¹⁶ M. L. Kessler, D. L. McShan, M. Epelman, K. A. Vineberg, A. Eisbruch, T. S. Lawrence, and B. A. Fraass, "Costlets: a generalized approach to cost functions for

- automated optimization of IMRT treatment plans," *Optimization and Engineering* **6**, 421–448 (2005).
- ¹⁷ T. R. Bortfeld, D. L. Kahler, T. J. Waldron, and A. L. Boyer, "X-ray field compensation with multileaf collimators," *Int J Radiat Oncol Biol Phys* **28**, 723-730 (1994).
- ¹⁸ "IMRT Benchmark," Quality Assurance Review Center Website (2002).
- ¹⁹ M. A. Ben-David, R. B. Marsh, J. M. Moran, J. M. Balter, K. A. Griffith, J. A. Hayman, and L. J. Pierce, "Partial Breast Irradiation (PBI) Using IMRT Compared With 3D Conformal PBI and Whole Breast Radiotherapy for Early Stage Breast Cancer: The Difference Is in the Normal Tissue," *Int J Radiat Oncol Biol Phys* **66**, S182-S183 (2006).
- ²⁰ E. J. Hall, "Intensity-modulated radiation therapy, protons, and the risk of second cancers," *Int J Radiat Oncol Biol Phys* **65**, 1-7 (2006).
- ²¹ S. F. Kry, M. Salehpour, D. S. Followill, M. Stovall, D. A. Kuban, R. A. White, and Rosen, II, "The calculated risk of fatal secondary malignancies from intensity-modulated radiation therapy," *Int J Radiat Oncol Biol Phys* **62**, 1195-1203 (2005).
- ²² E. Malinen, Å. Søvik, D. Hristov, Ø. S. Bruland, and D. R. Olsen, "Adapting radiotherapy to hypoxic tumors," *Phys. Med. Biol.* **51**, 4903-4921 (2006).
- ²³ B. Vanderstraeten, W. De Gerssem, W. Duthoy, W. De Neve, and H. Thierens, "Implementation of biologically conformal radiation therapy (BCRT) in an algorithmic segmentation-based inverse planning approach," *Phys Med Biol* **51**, N277-N286 (2006).
- ²⁴ L. Xing, c. Cotrutz, S. Hunjan, A. L. Boyer, E. Adalsteinsson, and D. Spielman, "Inverse planning for functional image-guided intensity-modulated radiation therapy," *Phys Med Biol* **47**, 3567-3578 (2002).
- ²⁵ K. S. C. Chao, W. R. Bosch, S. Mutic, L. J. S, F. Dehdashti, M. M. A, D. J. R, C. A. Perez, J. A. Purdy, and M. J. Welch, "A novel approach to overcome hypoxic tumor resistance: Cu-ATSM-guide intensity-modulated radiation therapy," *Int J Radiat Oncol Biol Phys* **49**, 1171-1182 (2001).
- ²⁶ W. A. Tomé and J. F. Fowler, "On cold spots in tumor subvolumes," *Med Phys* **29**, 1590-1598 (2002).
- ²⁷ J. Z. Wang and X. A. Li, "Impact of tumor repopulation on radiotherapy planning," *Int J Radiat Oncol Biol Phys* **61**, 220-227 (2005).
- ²⁸ J. Z. Wang, X. A. Li, W. D. D'Souza, and R. D. Stewart, "Impact of prolonged fraction delivery times on tumor control: A note of caution for intensity-modulated radiation therapy (IMRT)," *Int J Radiat Oncol Biol Phys* **57**, 543-552 (2003).
- ²⁹ A. Niemierko, "A generalized concept of equivalent uniform dose (EUD)," *Med Phys* **26**, 1100 (1999).
- ³⁰ D. M. Shepard, M. A. Earl, X. A. Li, S. Naqvi, and C. Yu, "Direct aperture optimization: A turnkey solution for step-and-shoot IMRT," *Med. Phys.* **29**, 1007-1018 (2002).

CHAPTER VII.

POTENTIAL CLINICAL IMPACT OF CONTROLLING IMRT BEAM MODULATION

Chapters I and II demonstrated that intensity modulated radiotherapy fluence distributions can be highly complex and that from this complexity arises a variety of concerns associated with the efficiency and reliability of IMRT treatment delivery. Chapters III-VI presented a several approaches to control the complexity of IMRT fields during the optimization process. These techniques can produce IMRT plans that are much more efficient to deliver than their unsmoothed standard IMRT counterparts. However, the true clinical impact of controlling beam complexity has not yet been verified. The purpose of this chapter is to examine some of the areas of potential clinical impact and determine whether there are real advantages to smoothing that can be observed and verified in practice, not just in theory. The areas of clinical significance that this chapter covers are treatment planning consistency and efficiency, delivery accuracy, sensitivity to geometric uncertainty, and the potential impact of smoothing on adaptive radiotherapy.

VII.A. Treatment Planning Robustness and Efficiency

Applying a modulation penalty in the cost function has the potential to affect the treatment planning process in ways that have not yet been discussed. Two issues that need examination are (i) the robustness of the smoothed plans to point sampling effects and (ii) the efficiency of the optimization itself.

VII.A.1. Robustness to Point Sampling

Chapter II discussed how the discrete sampling of points in the optimization regions of interest could be a source of unwanted fluctuations and isolated high intensity peaks in the beamlet intensity patterns. While these features had minimal effect on the

dosimetric quality of the plan when evaluated at different point samplings, they reduce the confidence that the specific beamlet intensity pattern which is used for treatment is the best pattern to treat the patient. The addition of modulation penalties inside the cost function has been shown to reduce small and high frequency fluctuation in the intensity patterns. Therefore, it may be hypothesized that the beams optimized with modulation penalties are less dependent on the specific point sampling in the field, so long as the area is properly sampled. To go a step further, plans with smoother beams may be less susceptible to undersampling artifacts. To test these hypotheses, a phantom case was optimized with and without modulation penalties when using (i) several different random point sampling instances, and (ii) progressively fewer optimization points in the regions of interest.

VII.A.1.1. Random Point Sampling

The modified IMRT benchmark phantom introduced in Chapter II was used to study the robustness of three different modulation penalties to different random point samplings of the regions of interest. The phantom is shown in Figure VII-1. Five different instances of the point sampling were generated with an average voxel size of 0.027 cm^3 in each structure, which is our typical clinical recommendation. Each instance was optimized four times:

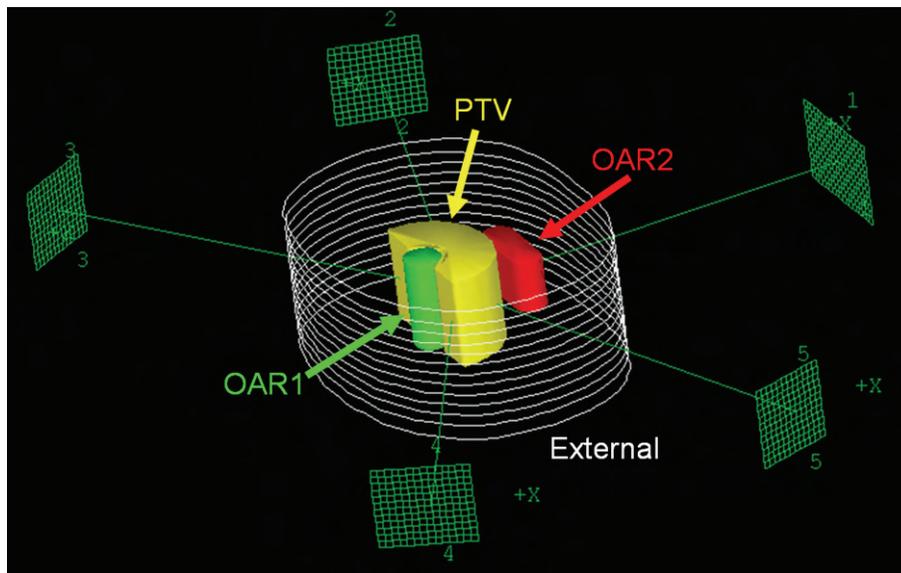


Figure VII-1. The modified IMRT benchmark phantom designed for IMRT validation and testing (reproduced from Chapter II).

1. Baseline optimization with no smoothing using the cost function shown in Table VII-1.
2. The achieved baseline objectives were constrained and the plan was reoptimized using the plan intensity map variation (PIMV) penalty introduced in Chapter IV.
3. The achieved baseline objectives were constrained and the plan was reoptimized using the ADS penalty with uniform diffusion coefficients as introduced in Chapter V.
4. The achieved baseline objectives were constrained and the plan was reoptimized using the ADS penalty with cost function gradient-based diffusion coefficients as also introduced in Chapter V.

Table VII-1. The cost function applied to modified IMRT benchmark phantom

Structure	Objective	Costlet	Weight
PTV	Min > 60 Gy	DVH_GE(60,100)	100
	Max < 66 Gy	DVH_LE(66,0)	100
OAR1	Max < 52 Gy	DVH_LE(52,0)	100
	Mean < 0 Gy	MEAN.LE.0	0.0001
OAR2	Mean < 30 Gy	MEAN.LE.30	10
	Max < 66 Gy	DVH_LE(66,0)	1
	Mean < 0 Gy	MEAN.LE.0	0.0001
Normal	Max < 66 Gy	DVH_LE(66,0)	1
	Mean < 0 Gy	MEAN.LE.0	0.0001

Corresponding optimization trials from each point sampling instance (named A, B, C, D, and E) were compared to determine the similarity between each of the optimized beams for each different point sampling instance. The correlation coefficient (also discussed in Chapter II, Equation II.2) was used as the measure of beam similarity. In addition, optimization results from point samplings B-E were analyzed using the point sampling in A to verify that plan quality was not compromised by the smoothing process in the different point samplings. Chapter II showed that no significant changes in the DVHs were observed when analyzing the baseline optimization results over different point sampling instances; it is important to show that the same is true for plans that use smoothing penalties.

The smoothing costlet weights were chosen to make a minimal impact on the cost function and also to give nearly identical cost values (not including the smoothing

costlet value) between the different methods so they could be fairly compared. The same weights were then used in each different point instance so that the cost function used would be identical. To demonstrate the clinical equivalence of all plans at the 5 different point samplings, we show the DVHs for all point sampling instances for each technique in Figure VII-2. Even subtle changes, such as the OAR1 DVH shift observed in point sampling E, are constant across the different methods. The MU reduction averaged $46.0 \pm 3.4 \%$ for all of the smoothing costlets in each of the point sampling versus the baseline plan. No clear advantage in MU reduction at this low weighting factor was observed for any of the smoothing costlets.

Although the DVHs are very similar, the optimized beamlet intensity patterns can vary by a large amount due to the difference in the locations of the sampled dose evaluation points. In Figure VII-3, we show one optimized beam from each point

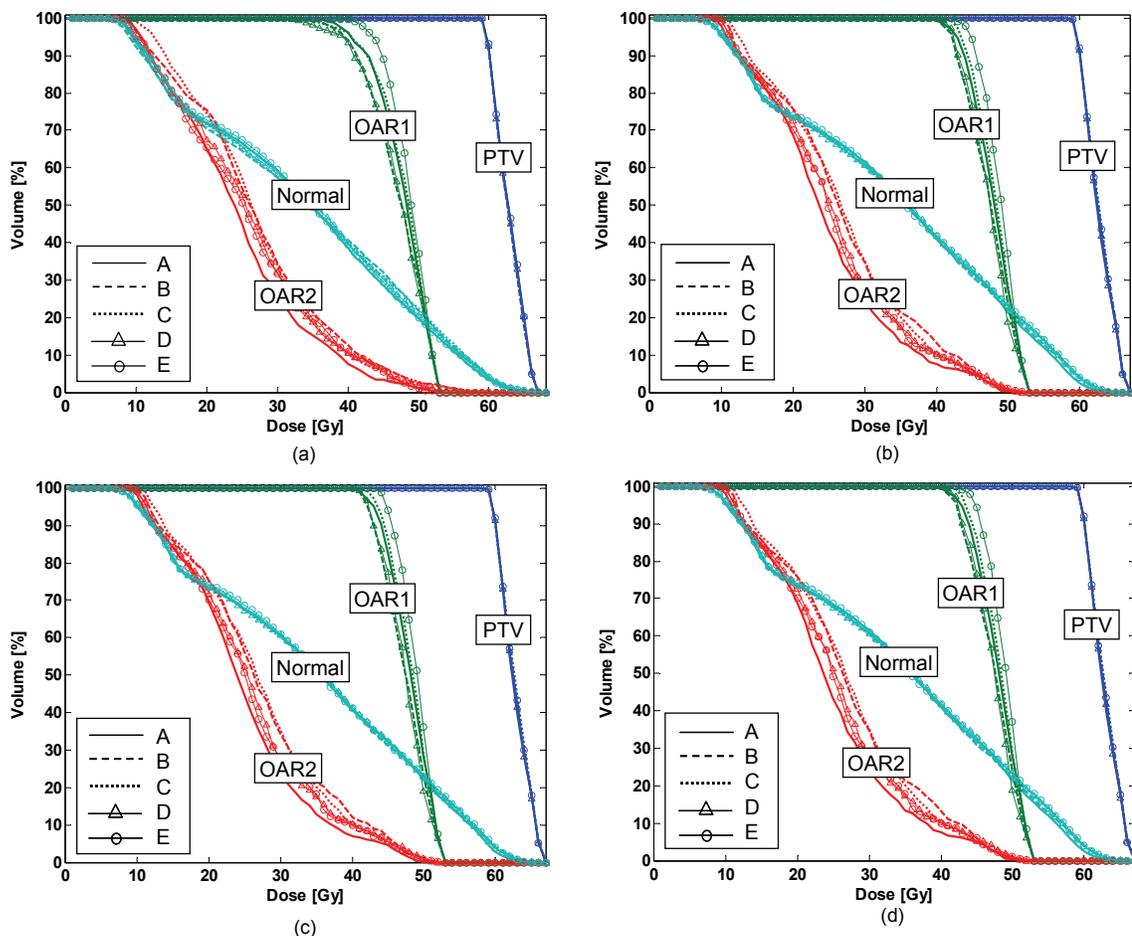


Figure VII-2. The group of (a) Baseline, (b) PIMV, (c) ADS-U, and (d) ADS-G DVHs optimized for each different point sampling A-E.

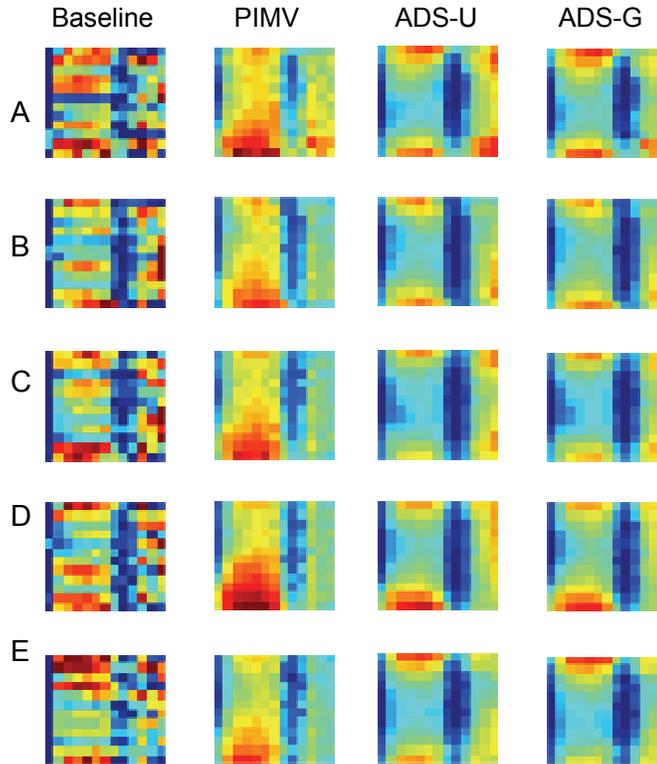


Figure VII-3. Beam 3 as optimized with each different technique in each of the different random point sampling A-E

sampling instance for each of the 4 different techniques. The intensity changes in the baseline plan are much more noticeable and of higher intensity than the differences in the smoothed beams. This is shown more quantitatively in Figure VII-4 by the mean

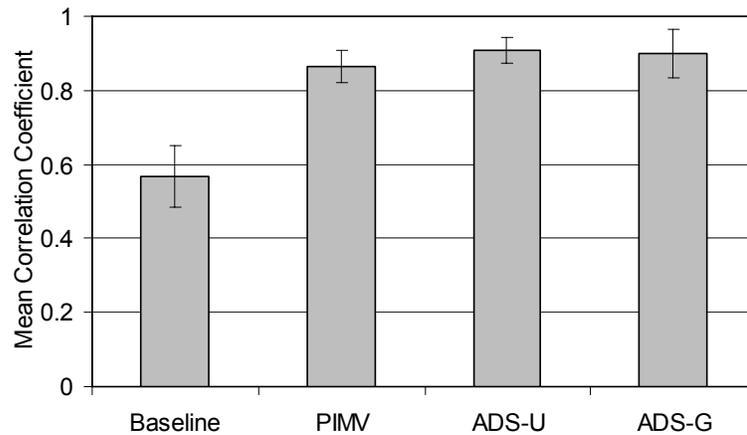


Figure VII-4. The mean correlation coefficient (and σ) between plans optimized on 5 different point sampling instances using 4 different techniques.

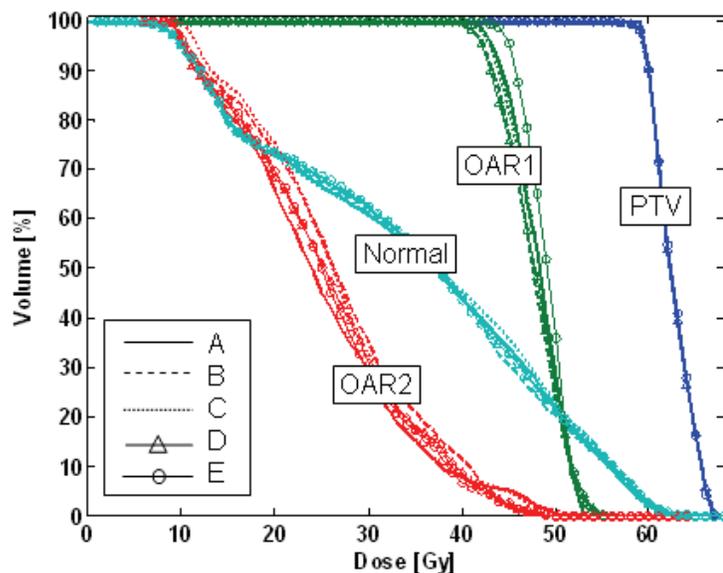


Figure VII-5. DVHs for the PIMV plans from all point samplings evaluated on point sampling A.

correlation coefficient between all beams in each of the smoothed cases relative to the beam in point sampling A. The smoothed beams correlate to a much higher degree than the baseline beam, making these plans more robust with respect to the specific random point sampling of the cases. The fact that the ADS-Gradient plans are more dependent on the points due to their use of the cost function gradient may have increased the standard deviation in that case, but it did not significantly alter the results compared to the other smoothing costlets.

Clinically speaking, each of the intensity patterns is still of high quality when evaluated on a different sampling of points. For example, Figure VII-5 shows the DVHs for the PIMV plans from all point samplings evaluated on point sampling A. There is no noticeable plan degradation. However, when we compare the actual cost function value in Figure VII-6, we see that there is an improvement in the cost function value of the smoothed plans relative to the baseline plan when evaluated on point sampling A. This suggests that there is a dosimetric robustness observed to point sampling when using the smoothing costlets.

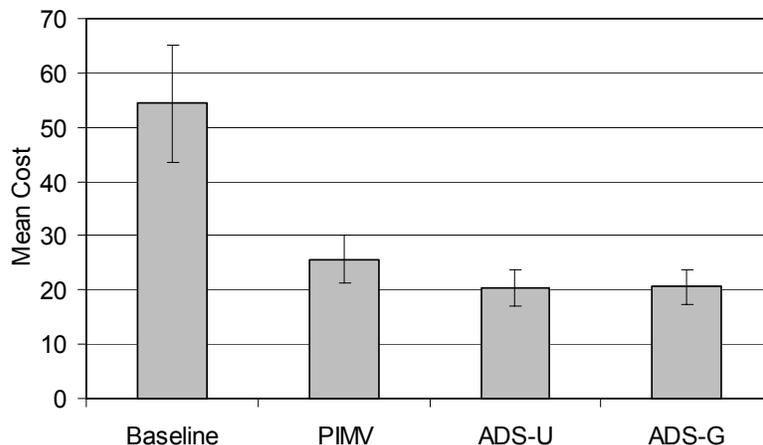


Figure VII-6. The mean cost (and σ) between plans optimized using 4 different techniques on random point sampling instances B-E and then evaluated on the point sampling of A. This would represent the cost of B-E if A were the true geometry.

VII.A.1.2. Point Subsampling

A possible limitation of accuracy in IMRT planning is the potential undersampling of regions of interest. The number of discrete points that can be used in the optimization process is constrained by memory limits and the number of beamlets in a plan. A larger number of points can also significantly slow the optimization process. Both of these issues can lead to the undersampling of the points in regions of interest. We demonstrated in Chapter II that this undersampling can result in higher plan modulation and degrade the dosimetric plan quality. This section will analyze whether plans optimized with the PIMV and ADS penalties are more robust to undersampling than unsmoothed IMRT plans in terms of both beam modulation and dosimetric quality. To test this, the point sampling in the phantom in Figure VII-1 was successively reduced by factors of 2, 3, 4 and 5 and optimized using the same four methods given in Section VII.A.1.1. To determine the degradation in plan quality as a result of undersampling, the “optimal” beams from each subsampled plan are analyzed using the full sample of points. The changes in the DVHs and total cost value are compared between plans with and without smoothing. The similarity of the optimized beams at each point sampling using the same modulation penalty are compared using the correlation coefficient relative to the fully sampled plan.

To ensure that the smoothing costlets weights chosen do not significantly affect the plan quality and bias the robustness results, the DVHs for the fully sampled plan and

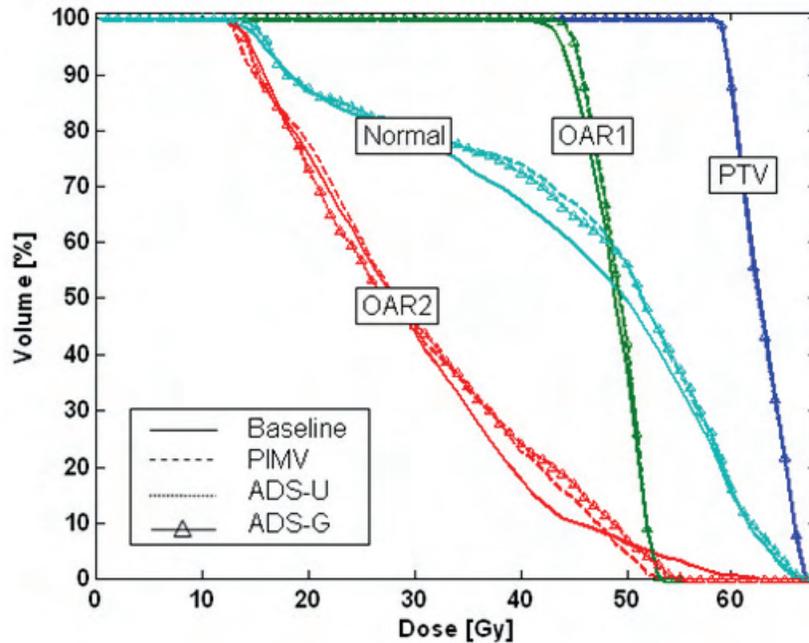


Figure VII-7. The DVHs for the phantom case when optimized with the full point sampling for all smoothing techniques compared to the baseline plan. This is shown to verify that the choice of the smoothing costlet weights for this comparison was conservative and resulted in equivalent smoothing for each of the cases.

each of the smoothed plans are shown in Figure VII-7. These DVHs are very similar, showing that a conservative weight was chosen for the smoothing costlets inside the cost function. This weight was held constant throughout each of the subsamplings, and similar DVH comparisons were observed in all cases.

The results of this study show that the use of the smoothing costlets makes the IMRT plan more robust to undersampling, both in terms of beam modulation and dosimetric quality. Figure VII-8 shows the average correlation coefficient between the optimal beams from each of the subsampled plans relative to fully sampled plan for each of the 4 techniques. The baseline plan with no smoothing shows a more significant decrease in beam correlation as the subsampling increases, while the correlation between the smoothed beams is higher and decreases much more slowly as the subsampling increases. There is no distinct advantage in any one of the smoothing costlets.

The fact that the beams are more stable as a function of the subsampling also results in less plan quality degradation when the optimization is performed on a reduced set of points. To demonstrate this, we evaluated the subsampled beamlet intensities using the fully sample of points to determine what the true cost function value would be

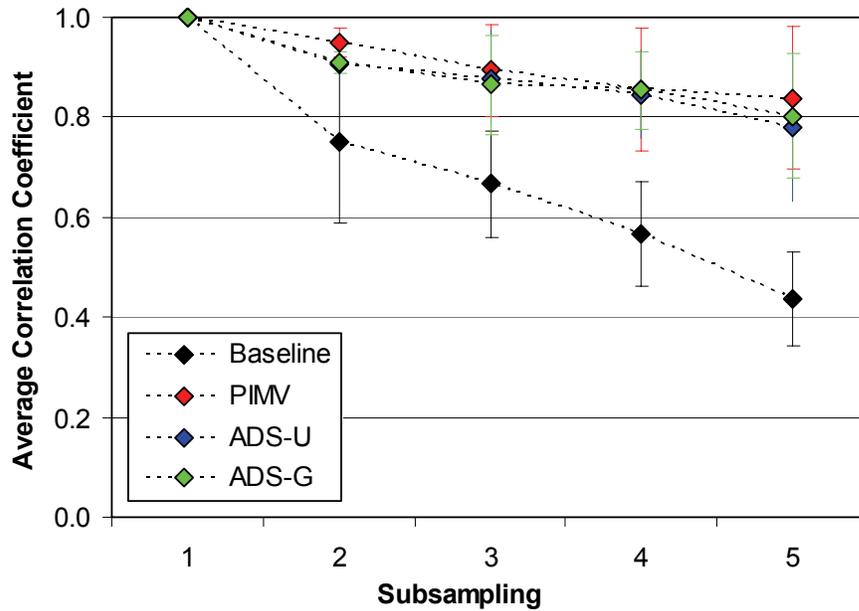


Figure VII-8. The average correlation coefficient between the optimal beams from each of the subsampled plans relative to fully sampled plan for each of the 4 techniques. The standard deviation of the correlation coefficient for each of the 5 beams is also shown.

versus the perceived optimal value obtained with the reduced number of optimization points. Figure VII-9 shows the difference between the true and perceived optimal cost function values as a function of subsampling for each of the techniques. The cost difference is plotted on a log scale, and an exponential curve was fit to each of the datasets. The fits were all excellent, with R^2 values of 0.99. The baseline cost degradation is not only higher, but it also increases at a higher pace. This can be seen by the exponential coefficient b , which is displayed in the lower right of the figure. The behavior of the degradation in plan cost as a function of subsampling was very similar for each of the smoothing costlets, with a slight advantage when using the ADS costlets. This is likely due to the fact that the PIMV costlet results in more uniform beams, which smooth out the intensity gradients in the field, resulting in the possibility of additional dose going to the normal structures.

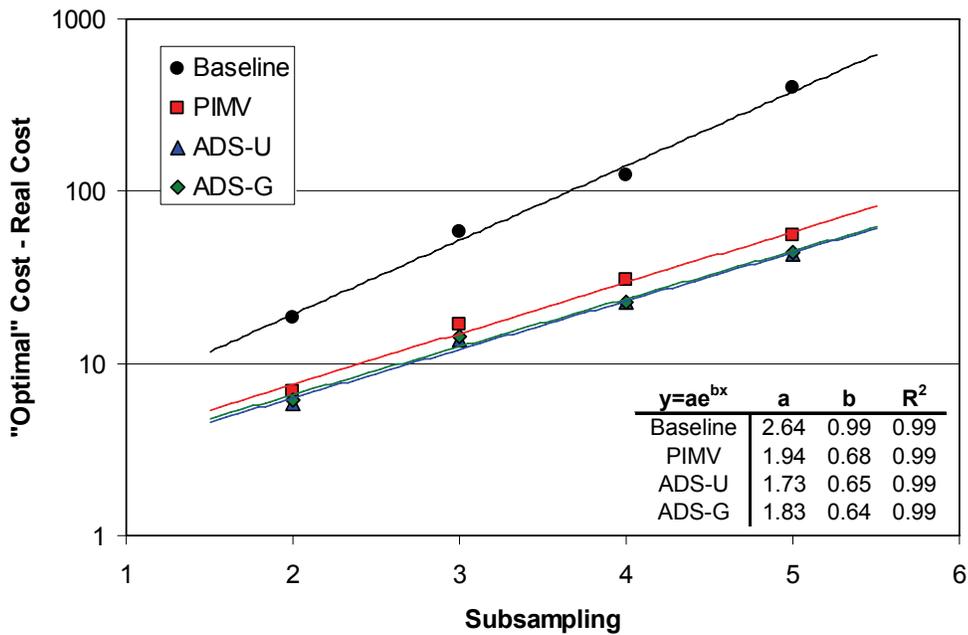


Figure VII-9. The difference in the true and perceived optimal cost function values as a function of subsampling for each of the techniques. Exponential fits to the data are shown in the lower right.

VII.A.2. Optimization Efficiency

It has been observed by several authors that the use of a smoothing filter during optimization can reduce the number of iterations required for optimization^{1,2}. Alber *et al.* suggests that this may be a by-product of the solution space becoming steeper near convergence with the addition of a modulation penalty in the objective function¹. We have made similar observations. The purpose of this section is to compare the optimization efficiency of standard IMRT with IMRT including several of the modulation penalties that we have developed.

VII.A.2.1 Methods

To determine the improvement in optimization efficiency that results from the use of a modulation penalty in the cost function, three clinical examples were optimized with and without modulation penalties, and the number of iterations required for convergence of the cost function were recorded. Both the quadratic plan intensity map variation (PIMV_q) penalty and adaptive diffusion smoothing (ADS) penalty with the cost function gradient

diffusion coefficients were employed. Since the convergence time may also be a factor of the modulation penalty weight, the optimizations were run with increasing weights of the two penalties.

Data from three patients previously treated for brain, prostate, and head/neck cancer were used for this study. Point sampling density was as high as possible within the memory constraints of the software. The cost functions used closely approximate the current clinical treatment planning directives at the University of Michigan for each treatment site. The brain case consisted of 5 non-axial beams (originally placed by a dosimetrist) with $1 \times 1 \text{ cm}^2$ beamlets. The prostate and head/neck cases consisted of 7 axial beams each. The prostate case was planned with $0.5 \times 0.5 \text{ cm}^2$ beamlets and the head/neck case was planned with $1 \times 1 \text{ cm}^2$ beamlets, due to the size of the target volumes.

The optimization search algorithm employed was the quasi-Newton method and the convergence parameters were set to zero to prevent the risk of premature termination of the optimization (see Chapter II, Section II.B.3.1).

VII.A.2.2. Results and Discussion

The use of the smoothing costlets during optimization resulted in quicker optimization time, with fewer iterations of the quasi-Newton algorithm in all but one instance. In the brain, when applying the ADS-Gradient penalty at a low weighting factor, the time was slightly increased from the baseline optimization. This was likely due to the increased time required to calculate the diffusion-smoothed beam and gradients. However, we note that the code was not optimized for speed, and therefore this disparity could be lessened or removed. Also, traffic on the network as well as individual computer differences could play a factor in the time calculations.

Figure VII-10 shows the number of iterations to full convergence, as well as total optimization time results for each of the patients. The savings in both time and iterations increases with increasing plan complexity, and the PIMV penalty showed a clear advantage in both categories in all cases. In fact, the PIMV penalty resulted in an 80 % reduction in the number of iterations required to optimize all patients when using the higher weighting factor. The ADS penalty performed almost equally well in the prostate although the PIMV penalty still had an advantage in the time factor. However, as we mentioned, the ADS code is not currently optimized for speed. Because of the fact that the PIMV plans are more uniform, we would expect them to converge in fewer iterations

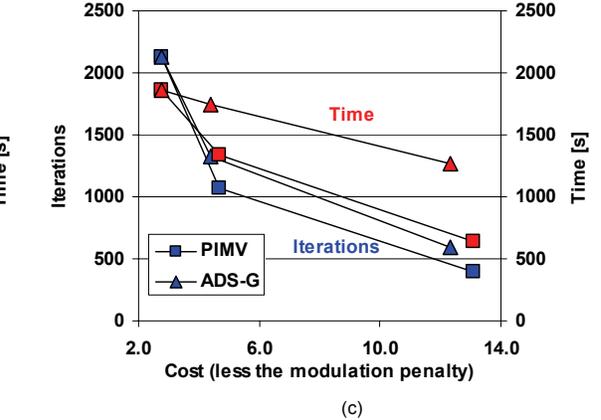
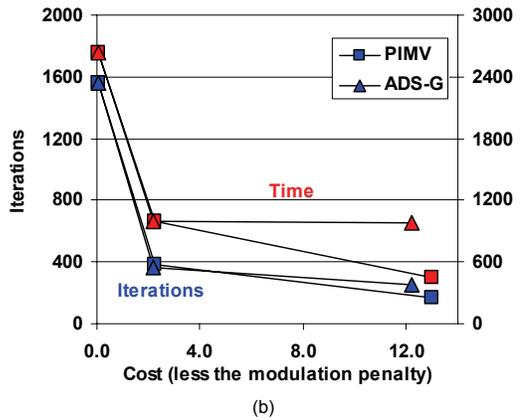
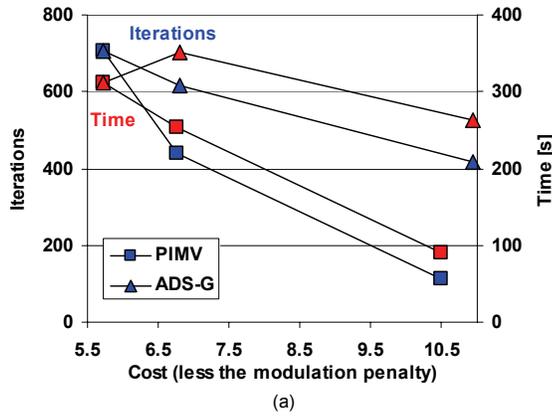


Figure VII-10. The time and number of iterations required for full convergence in both optimization techniques as a function of the total cost function value (not including the smoothing costlet penalties) for the (a) brain, (b) prostate, and (c) head/neck studies. The leftmost datapoint in all cases represents the baseline plan, or a smoothing penalty weight of zero. Note the two scales.

than the ADS plans. However, the disproportionate increase in time per iteration in the ADS plans may be reduced with code improvements.

To demonstrate the conservative and clinically acceptable weighting factors that were applied, each of the optimized prostate case DVHs are shown in Figure VII-11. Each structure is nearly identical. However, the ADS-Gradient penalty has a slight advantage in providing an MU reduction of 47 % compared to 39 % for the PIMV case. The MU reductions in the brain and head/neck are similar for both smoothing costlets. There is a general trend for the ADS plans to be smoother overall with higher (but very smooth) intensity gradients across the field. The PIMV beams are generally more uniform, but with more “noise” in the modulation patterns. These features and more delivery efficiency comparisons are demonstrated in Chapters V and VI.

The main point and conclusion of this section is that the addition of a smoothing penalty does not make the optimization problem more difficult or costly to solve. Instead, this drives the optimization problem to the cost function minima with fewer iterations and less time. Our results agree with the work of other authors, suggesting

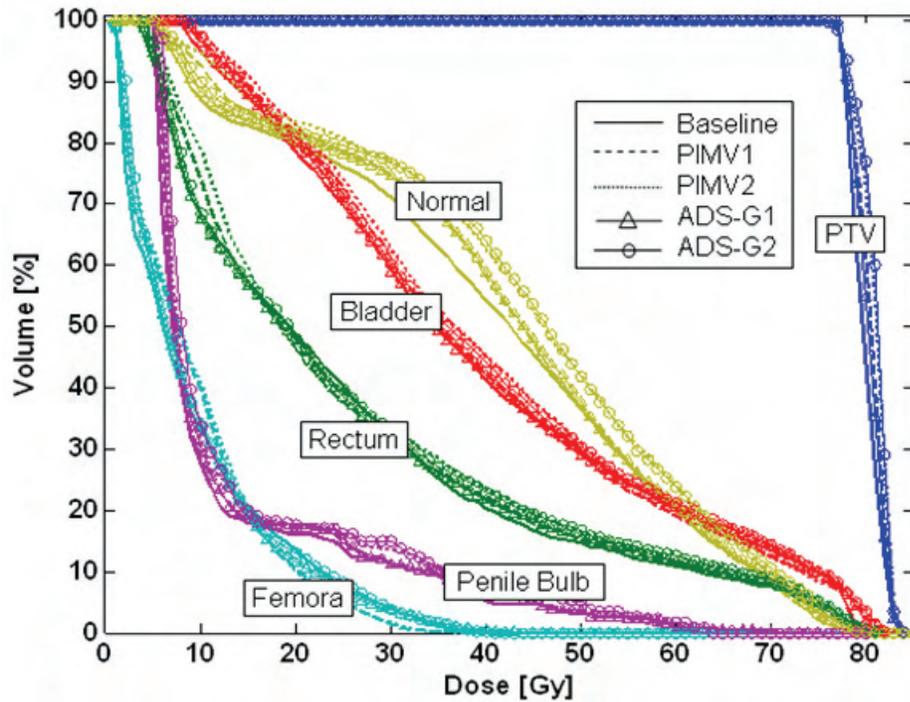


Figure VII-11. DVHs (for all of the techniques in the prostate cases) demonstrate the clinical acceptability of the weighting factors chosen. The different trials (PIMV1 and 2, ADS-G1 and 2) are relative indicators of the weighting factors.

that the addition of modulation penalties can increase the steepness of the cost function near the minimum and result in more efficient optimization^{1,2}. The time savings in the prostate and head/neck was significant and could likely be improved further with optimization of the smoothing costlet code implementation.

VII.B. Delivery Accuracy and Efficiency

VII.B.1. Motivation

Chapter II reviewed many of the consequences of highly complex IMRT fields. One of these concerns was the fact that complexity often leads to lengthy and inefficient treatment deliveries that are susceptible to delivery errors and artifacts, which are

exacerbated when using the high dose rates that are usually necessary for acceptable treatment times for complex cases³. These issues complicate quality assurance procedures and can create significant discrepancies between the planned and delivered dose distributions. Investigation of these issues (and others) motivated the work in the dissertation, which aims to control beam complexity while preserving the advancements seen with IMRT. We have alluded to the fact that smoother beams will make delivery sequences more accurate and more efficient, and the purpose of this section is to experimentally verify the impact of one of our modulation penalties, the ADS penalty, on the delivery aspects of IMRT. Study of this issue includes performing and analyzing ion chamber and film measurements for several cases that have been planned with standard beamlet optimization and then comparing the results with those for plans optimized with an ADS penalty included in the cost function.

VII.B.2. Methods

The adaptive diffusion smoothing penalty is described in Chapter V. Each of the ADS plans discussed in this section uses the cost function gradient-based diffusion coefficients, which are also discussed in Chapter V.

To experimentally verify the delivery aspects of the ADS beams, several phantom and clinical examples (brain, prostate and head/neck) were optimized with and without the ADS penalty. All plans included a baseline cost function, which was applied for standard beamlet optimization. Each plan was then reoptimized twice, using the ADS penalty. In the first ADS plan, called “ADS-Equivalent”, the achieved objectives from the standard IMRT were converted to constraints for the plan reoptimization. This method ensured that the ADS penalty was not allowed to make tradeoffs with any of the baseline plan objectives. These plans represent the amount of the smoothing that can be done without any sacrifice to the plan objectives. In the second ADS plan, called “ADS-Relaxed”, the plan objectives were relaxed, based on their importance in the plan. In most cases, the relaxation included a 2 Gy relaxation in the mean normal tissue doses (as long as a mean dose objective was not violated) and a 2 Gy relaxation in the target maximum dose. These concessions were chosen to be clinically realistic (and even conservative) sacrifices for decreases in plan MU and overall beam smoothing.

All plans in this study were optimized using UMOpt, and beamlet dose calculations were performed using a convolution/superposition algorithm based on work by Mackie *et al.*⁴. Optimized plans were then sequenced for step and shoot delivery

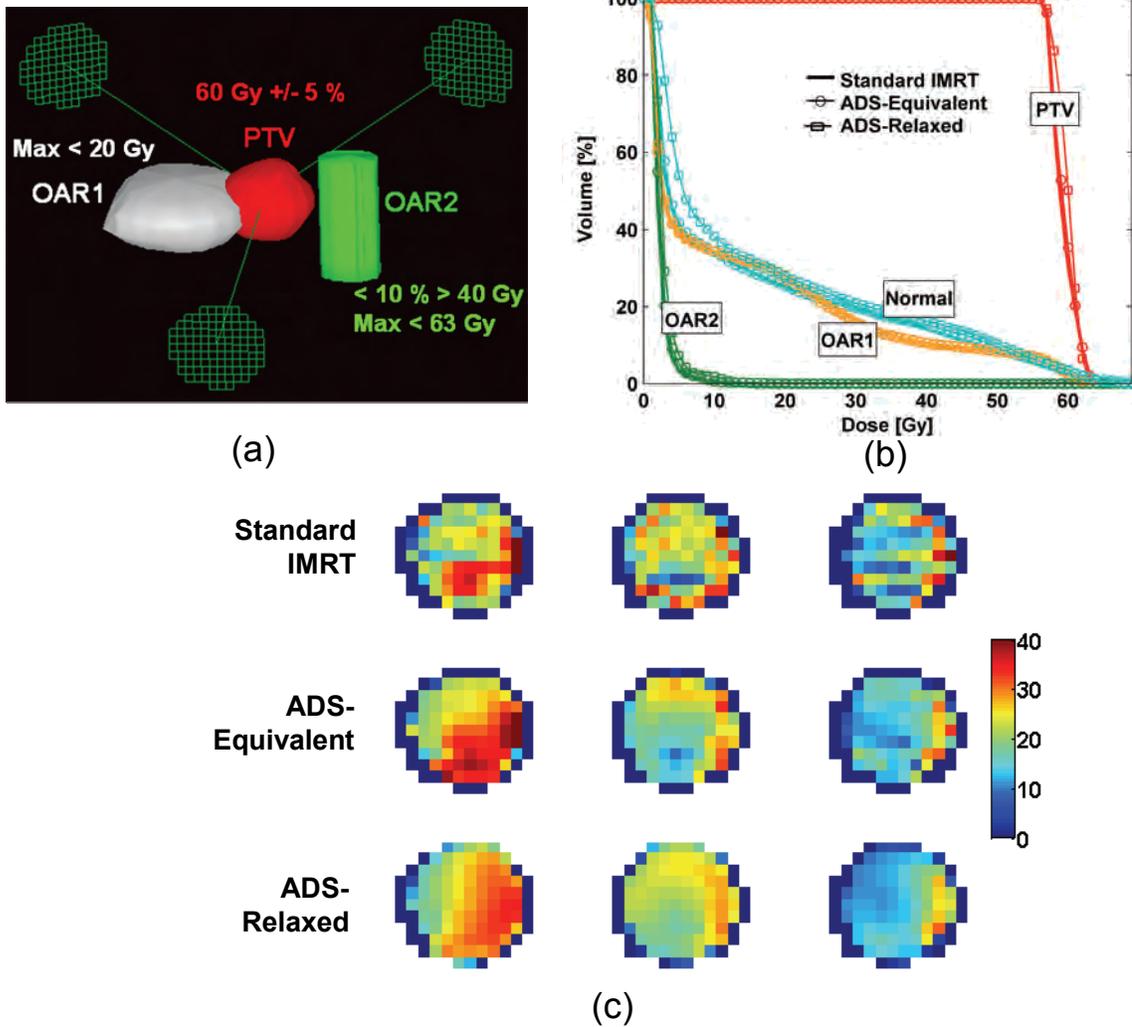


Figure VII-12. (a) Geometry, (b) DVHs, and (c) intensity distributions for Standard vs. ADS-Equivalent and ADS-Relaxed IMRT plans in the 3-field phantom case.

Table VII-2. Beam area outside +/- 5 cGy agreement between calculations and film measurements

Site	Technique	Percentage
Phantom 1	Standard IMRT	17.0%
	ADS-Equivalent	13.2%
	ADS-Relaxed	12.3%
Phantom 2	Standard IMRT	16.9%
	ADS-Equivalent	10.0%
	ADS-Relaxed	11.2%
Prostate	Standard IMRT	10.8%
	ADS-Equivalent	6.7%
	ADS-Relaxed	6.8%
Head/Neck	Standard IMRT	3.7%
	ADS-Equivalent	1.8%

using a leaf sequencer based on the method published by Bortfeld *et al.*⁵ It should be noted that the sequencing algorithm is designed to agree with the intended intensity pattern to a 1 % threshold and does not aim to reduce delivery segments or delivery time by sacrificing this agreement.

To remove any effects due to inhomogeneities, each of the plans was transferred to a 30 x 30 x 30 cm³ solid water geometry and recalculated for both ion chamber and film measurements. The high resolution convolution/superposition calculations were based on a 1 mm x 1 mm grid and then interpolated to a 2 mm x 2 mm grid for output and analysis. A 0.6 cc ion chamber was used to measure the composite plan dose and verify the absolute dose delivered to the ion chamber volume. To verify the individual IMRT fields, film measurements were performed with either Kodak XV or Kodak EDR film (depending on the maximum dose observed in each plan) in the solid water phantom. Film measurements were performed at 5 cm depth and 95 cm SSD with the gantry fixed perpendicular to the phantom for all fields. All plans used 6 MV photons with a dose rate of 600 MU/min (our clinical standard). Dose calibration curve films were exposed at the time of measurements. All films were processed and then digitized using a laser film digitizer. Film analysis was performed in IGOR Pro software with automatic registration capabilities. Differences between calculations and measurements were analyzed by dose difference displays, isodose contours, and the beam area falling outside a +/- 5 cGy agreement criteria. This was chosen to be consistent throughout all plans, since the maximum doses varied between techniques for the same case. Because of the known unreliability of film measurements in low-dose regions (due to the over-response of film to low energy photons), and the considerable low-dose noise present in EDR film, data points below 10 cGy and 5 cGy were excluded from the analysis for EDR and XV film, respectively.

In addition to the measurements, we also saved dynamic MLC log (Dynalog) files from the treatment deliveries for future examination. A Dynalog file is a record of the actual MLC delivery details recorded at every 0.05 s by the MLC controller of the treatment machine. The files contain information on the intended and actual leaf positions and whether the beam is on or off.

VII.B.3. Results

Five cases (two phantoms and brain, prostate and head/neck clinical examples) were optimized to generate standard IMRT, ADS-Equivalent, and ADS-Relaxed plans (ADS-

Relaxed plans are not presented for head/neck since many limits were already reached by the standard plan). The geometry and plan objectives and dose-volume histograms for the 3-field phantom case are shown in Figure VII-12(a) and (b). The DVHs demonstrate that the Standard and ADS-Equivalent plans are of similar quality, while the ADS-Relaxed plan made only slight sacrifices in the overall normal tissue dose. Figure VII-12(c) shows each of the optimized beams and illustrates the reduction in plan complexity that occurs with the use of the ADS penalty. There is an overall reduction in modulation while preserving the important gradients at the field edges and in the regions that overlap with normal structures.

Film measurements were compared with high resolution convolution calculations in each of the cases. Dose difference displays consistently demonstrated that the ADS plans agreed better to calculations than the standard IMRT plans. The percentage of beam area falling outside the ± 5 cGy criteria for all cases is shown in Table VII-2. The

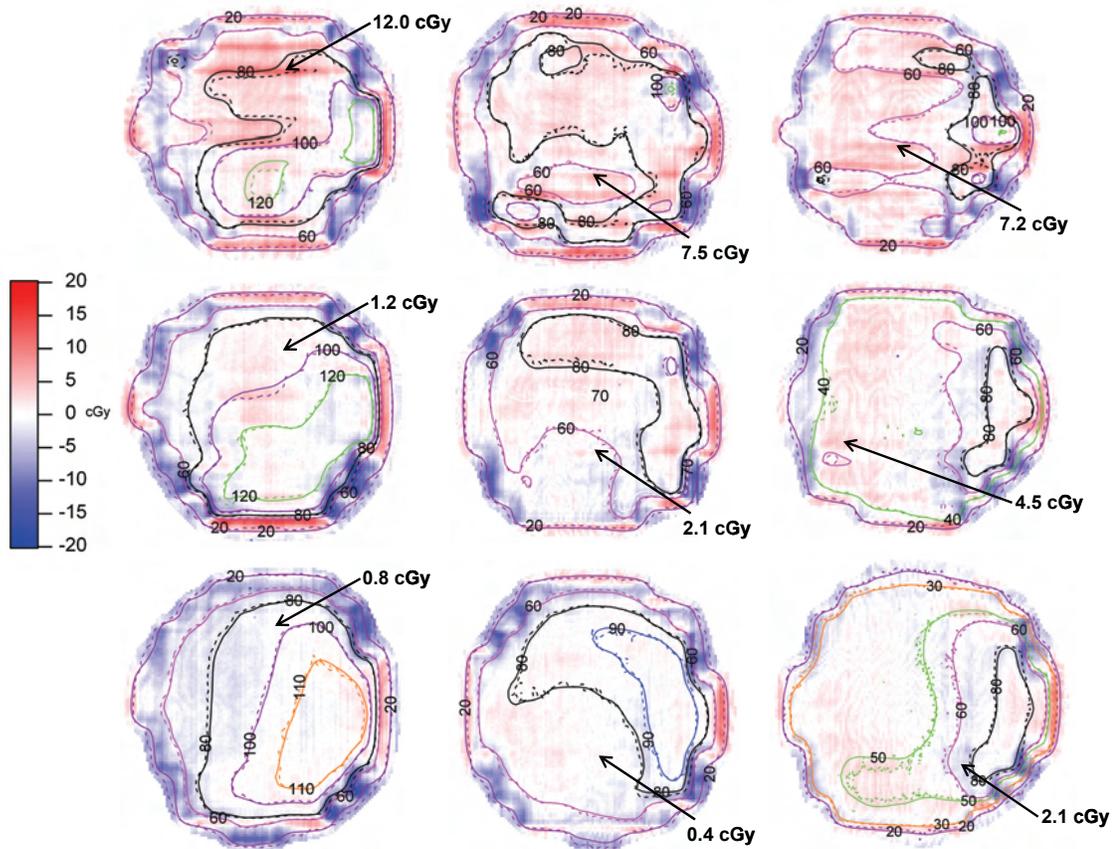


Figure VII-13. Dose difference displays (calculations – film) for the 3 field phantom case for standard IMRT (top row), ADS-Equivalent (middle row), and ADS-Relaxed (bottom row) plans. Absolute dose difference examples are given by the arrows.

brain case was excluded because of substantial noise in the EDR film.

Figure 13 shows the dose difference display for the 3-fld phantom as calculation – film. The noise in the lower dose regions of the display is attributed to the noise present in the EDR film measurements. The reduction in tongue-and-groove effect as well as an overall improvement in the agreement is noted as we go from the standard IMRT plan in the top row to the ADS-Relaxed plan in the bottom row.

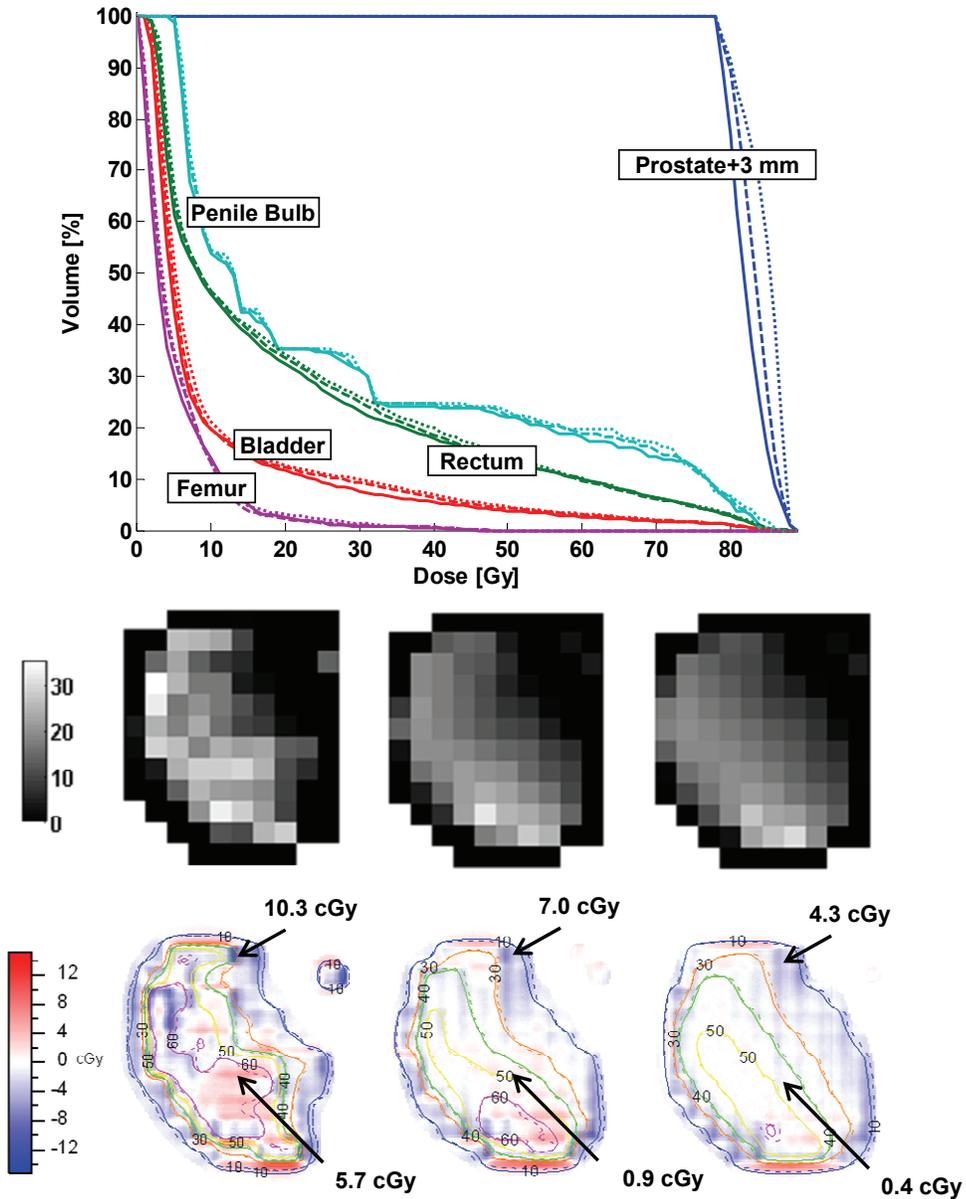


Figure VII-14. Prostate example: Top: DVHs (Standard IMRT solid, ADS-Equivalent dashed, ADS-Relaxed dotted) and Bottom: IMRT intensity maps and dose difference displays for Standard IMRT (left), ADS-Equivalent (middle), and ADS-Relaxed (right). Differences are calculations – film and isodose contours for both are shown as overlays in cGy along with example absolute dose differences (arrows).

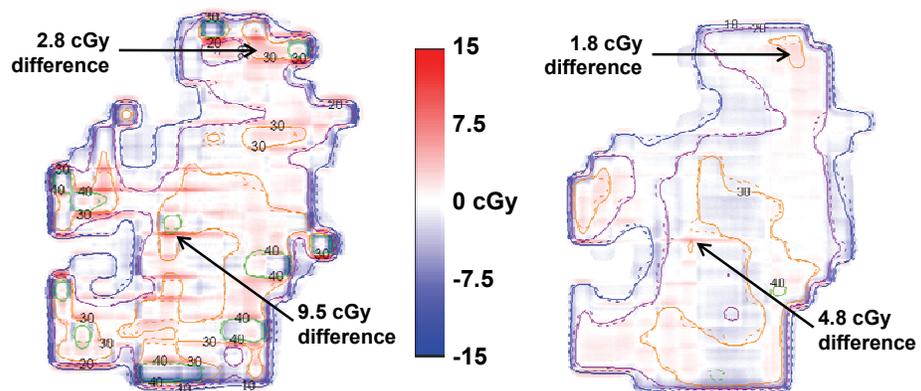
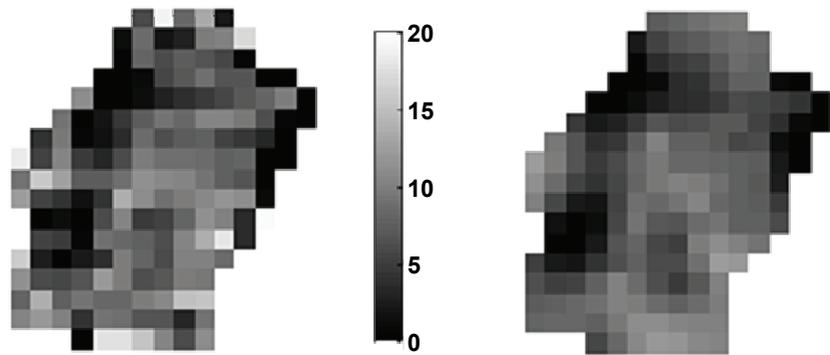
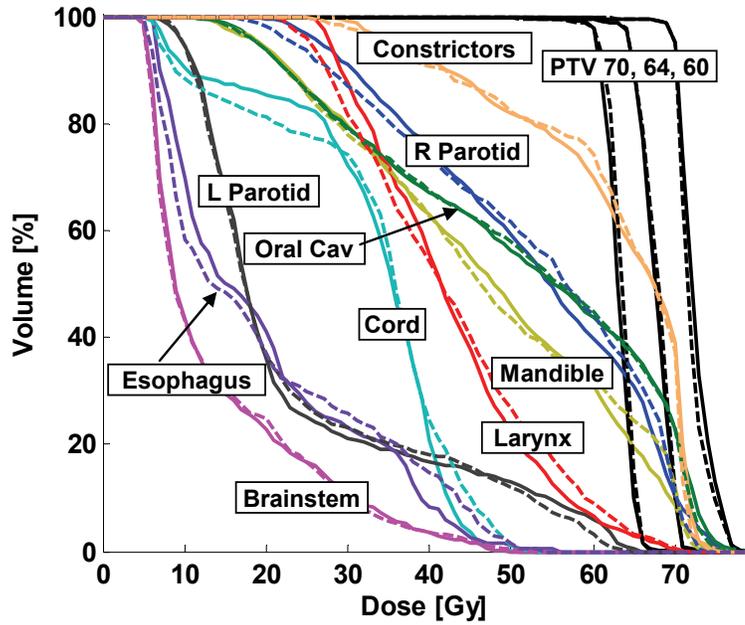


Figure VII-15. Head/neck example: Top: DVHs (Standard IMRT solid, ADS-Equivalent dashed) and Bottom: Planned IMRT intensity maps and dose difference displays (calculations – film) for Standard IMRT (left) and ADS-Equivalent (right). Examples of absolute dose differences are shown by the arrows.

Figures VII-14 and 15 show the DVHs, a typical IMRT beam intensity distribution, and dose difference displays for the clinical prostate and head/neck examples, respectively. The cost function employed in these cases and the brain case was based on in-house IMRT protocols. These dose difference displays demonstrate the improved qualitative agreement between calculations and film in the clinical body sites. It is apparent that there is a noticeable decrease in the amount of tongue-and-groove underdosage that occurred in the ADS plans compared to the standard IMRT plans. This may make it unnecessary to recalculate and evaluate the beams before delivery, since the ADS plans agree well with the convolution calculations used inside the optimization process.

Ion chamber measurements taken of the composite plans verified that the absolute dose of each plan was in an acceptable range of ± 2 cGy. Dose differences observed from ion chamber measurements between standard and ADS IMRT plans were thus all less than 2 cGy and not considered clinically significant. In addition, the composite dose calculations are not expected to be perfectly accurate because the positioning of the sensitive volume of the ion chamber in exactly the same position in the solid water phantom that was chosen in the calculation is very difficult. Therefore, these measurements were used as a verification tool instead of a comparison tool.

The relative time and MU required for delivery of all ADS cases is shown in Figure VII-16. We note that the leaf sequencer used in this work was not optimized to reduce delivery time.

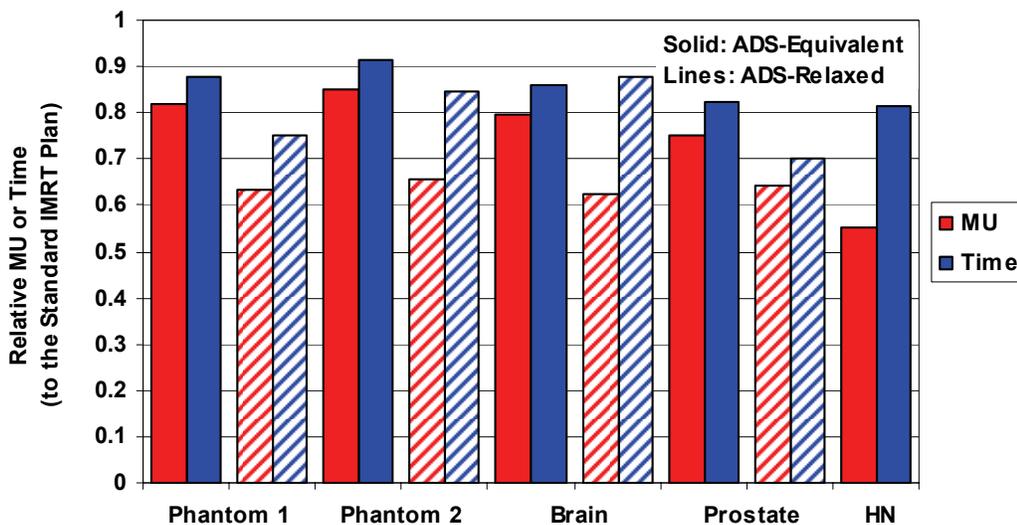


Figure VII-16. Relative MU and time for all cases for both ADS-Equivalent and ADS-Relaxed plans relative to standard IMRT.

VII.B.4. Conclusions

The ADS penalty can be used inside the IMRT optimization cost function to promote smoothing that does not affect the dosimetric quality of the plan. The resulting IMRT beams are smoother and can be delivered with fewer MU and in less time than standard IMRT plans. Dose differences between convolution calculations and film demonstrate that the smoother ADS plans are less susceptible to tongue-and-groove underdosage as well as to small areas of disagreement in the planned and delivered fields caused by large gradients in neighboring beamlets. This improved agreement with ADS plans may make it unnecessary to recalculate IMRT plans after optimization to evaluate the dose degradation that occurs during delivery. Future projects in this area could include analysis of the Dynalog files saved from delivery and comparisons with Monte Carlo simulations that accurately account for the transport through the MLC and patient geometry.

VII.C. Sensitivity to Geometric Uncertainty

We suggested in Chapter II that smoother IMRT fields may be less sensitive to geometric setup errors than highly complex fields. Schwarz et al. have shown that plans including high gradient fluence areas are more sensitive to geometric uncertainties than those with fairly uniform intensities⁶, and Duan *et al.* have published that the dosimetric error observed in IMRT increases in beams with high intensity gradients versus those dominated by more low intensity gradients across the field. In addition, Tomé and Fowler have shown that underdosing even small volumes of aggressive tumors can significantly degrade the probability of tumor control⁷. The purpose of this section is to test the PIMV and ADS penalties to determine whether they improve the geometric robustness of the IMRT plans. We examine both tumor and normal tissue dose-base metrics as well as the radiobiologic metric, equivalent uniform dose (EUD).

VII.C.1. Methods

Intensity modulated radiation therapy is a very conformal treatment modality that benefits greatly from precise daily localization of the patient in the correct treatment position. This localization is performed daily at our institution for IMRT patients through portal imaging and matching with the reference geometry. Because of this daily setup correction method, the main component of setup error that must be evaluated in our IMRT patients is the residual error remaining after daily imaging and correction. To

analyze the effect of setup errors of this magnitude, we begin with the modified IMRT benchmark phantom in Figure VII-1. This case was optimized using the cost function shown in Table VII-1, and then 10 residual setup errors were simulated. We have chosen to model residual prostate motion in this study since prostate is a common IMRT treatment site, and the residual setup error for prostate treatments has been studied at the University of Michigan. Litzenberg *et al.* have quoted that the average residual setup error (σ) over all fractions for four patients treated in the supine position (with daily imaging and correction) was 1.5 mm in the left-right direction, 1.6 mm in the anterior-posterior direction, and 1.2 mm in the inferior-superior direction⁸. To simulate errors of this magnitude for the phantom, we took 10 random samples of the (LR, AP, IS) residual setup error from normal distributions with mean = 0 mm and the given σ for each direction. The resulting errors for the 10 samples are shown in Figure VII-17. Next, the IMRT plan isocenter was shifted according to the simulated setup error for each of the 10 fractions and re-calculated so that the dosimetric effect of any changes in distance to the patient or patient heterogeneities could be taken into account by the dose calculation.

The unshifted geometry was used to plan the IMRT treatment using the baseline cost function in Table VII-1. The dosimetric results achieved in the baseline optimization

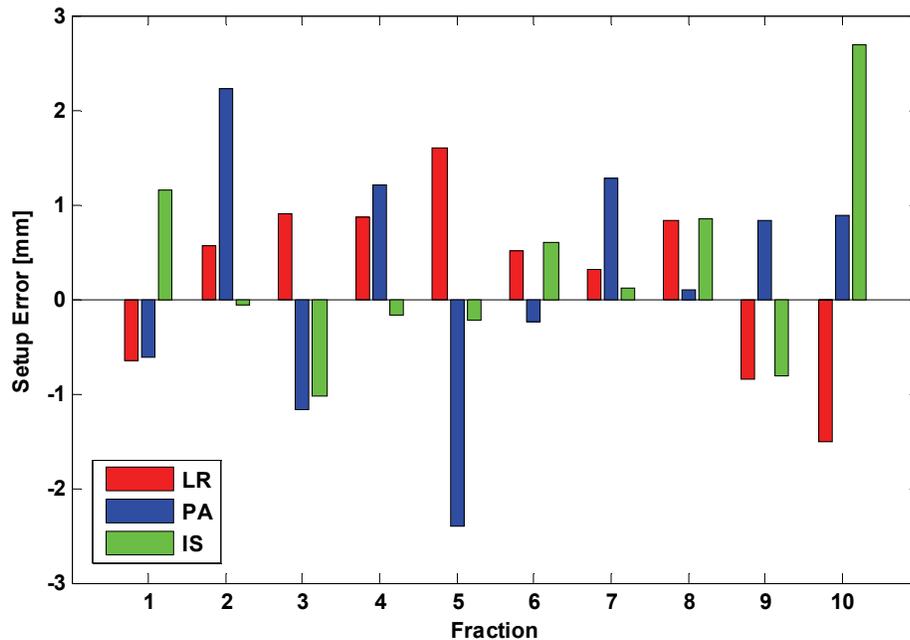


Figure VII-17. The sampled residual setup errors (σ) in each direction for 10 fractions to be simulated with the phantom case.

were then constrained, and the PIMV and ADS (both Uniform and Gradient) penalties were minimized. In addition to these dosimetrically equivalent plans, the ADS-Gradient penalty was applied during the initial optimization stage at two different weights to determine the effect of the increased amount of smoothing on plan robustness. The weights were conservative, so as to not negatively affect the objectives.

Finally, to determine the geometric sensitivity of each of the plans, the optimized beamlet patterns were evaluated on each of the 10 individual fraction geometries. The interfraction changes in the plan quality as well as the cumulative changes (taking into account all 10 fractions) in plan quality in the baseline versus smoothed plans were recorded and analyzed to determine whether the smoothed plans showed any advantage in robustness to the residual setup errors.

VII.C.2. Results and Discussion

The dose-volume histograms and beams for each of the plans that were used in this analysis of geometric sensitivity are shown in Figure VII-18. As we showed before, the DVHs are very similar, with slight changes in the OAR1 shape, and the beams are considerably noisier when using the baseline cost function only versus including a modulation penalty.

The effect of the simulated setup errors on the dosimetric quality of the plans in Figure VII-18 was evaluated in terms of dose differences resulting from individual fractions as well as the cumulative dose differences resulting from all of the fractions. As one would expect, the size and number of interfraction differences was larger than the cumulative dose differences. The percentage of points with greater than 2 Gy dose differences in the individual fractions and the percentage of points with 2 Gy or more cumulative dose differences are shown in Table VII-3. The percentage of cumulative errors greater than 2 Gy was extremely small. The percentages in the OAR1 and OAR2 represent under 20 total points ($< 1\%$ of the points) in each structure. The interfraction differences were more numerous, with a trend toward a decreasing number of differences greater than 2 Gy when the smoothing costlets were used.

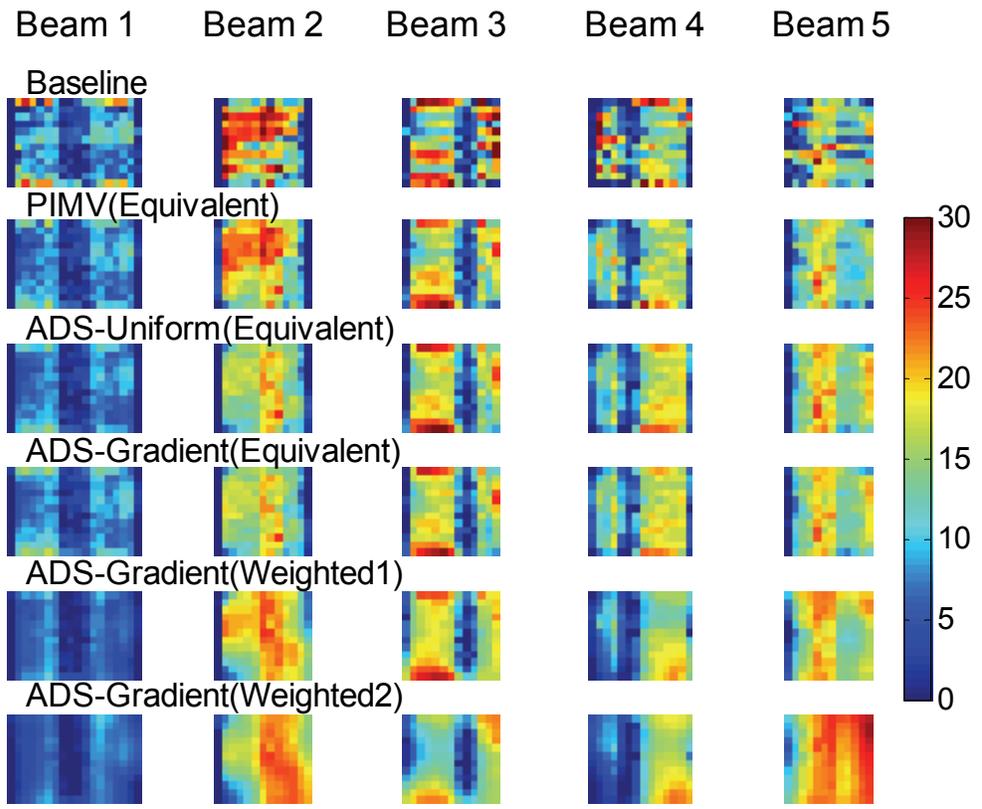
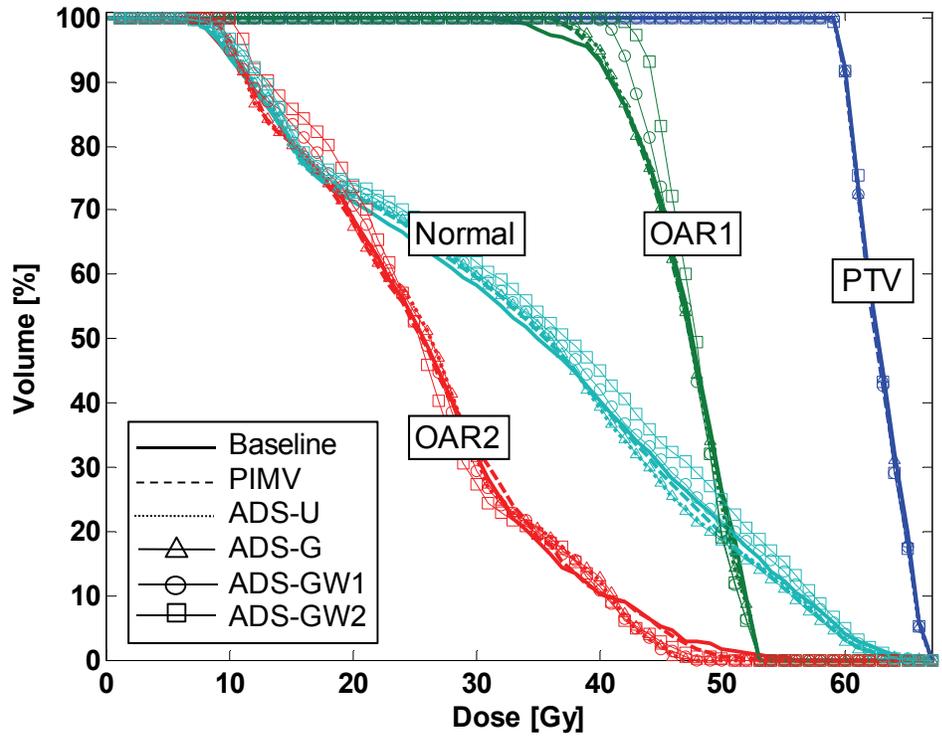


Figure VII-18. The DVHs and beams for each of the techniques used to study the impact of smoothing costlets on the robustness of the plan to geometric residual setup errors.

Table VII-3. Percentage of points with > 2 Gy dose differences

	Interfraction			Cumulative		
	PTV	OAR1	OAR2	PTV	OAR1	OAR2
Baseline	6.4%	14.1%	21.0%	0.67%	1.4%	4.8%
PIMV (Equivalent)	5.6%	13.2%	21.9%	0.41%	0.34%	3.0%
ADS-U (Equivalent)	5.3%	11.9%	22.0%	0.23%	0.34%	1.8%
ADS_G (Equivalent)	5.3%	12.2%	21.5%	0.23%	0.34%	2.1%
ADS_G (Weighted1)	4.5%	6.4%	20.1%	0.10%	0.00%	2.1%
ADS_G (Weighted2)	4.0%	5.6%	19.3%	0.00%	0.00%	1.1%

Another observation is that the actual dose differences were of greater magnitude in the baseline plan for both interfraction and cumulative errors. As an example, the PTV point dose differences are all plotted in Figure VII-19 for the Baseline plan (blue stars) and the ADS-Gradient plans (red circles). The ADS-Equivalent plan is shown in (a), the ADS-Gradient Weighted1 plan is shown in (b) and the ADS-gradient Weighted2 plan is shown in (c) for fraction 4. The distribution of the dose errors in the ADS-Gradient plans becomes narrow as the plans are more smoothed. This can be seen as the red markers become more focused around zero from a to c. There are also fewer outliers with high point dose differences in for the ADS plans as the smoothing increases from (a) to (c). This was typical behavior in all structures and fractions. Fraction 4, shown in Figure VII-19 was average in terms of the magnitude of errors. In another example, the histograms of cumulative errors for the PTV, OAR1, and OAR2 are shown in Figure VII-20 for baseline plan (blue) and the ADS-Gradient(Weighted2) plan (red). The errors shown here represent the best case scenario if the residual setup error was the only source of error in the plan. However, delivery errors and breathing motion will likely increase the dose differences, and both of these have been shown to be larger for more complex fields.

It is worth noting that the main source of error in this study was the actual shift in the dose distribution due to the residual error. The use of smoothing costlets does not remove this error, but it does reduce the magnitude of some of the point dose errors that result. To show this overall effect, we have plotted the equivalent uniform dose (EUD) for all of the cases in Figure VII-21. The trend in EUD shows that the small residual errors do not have a large effect on the PTV, which was modeled to be a fairly resistant tumor at $\alpha = -15$. However, the change in EUD in the most out of field structure, OAR2, was significant in several of the fractions. The use of smoothing did not have a large

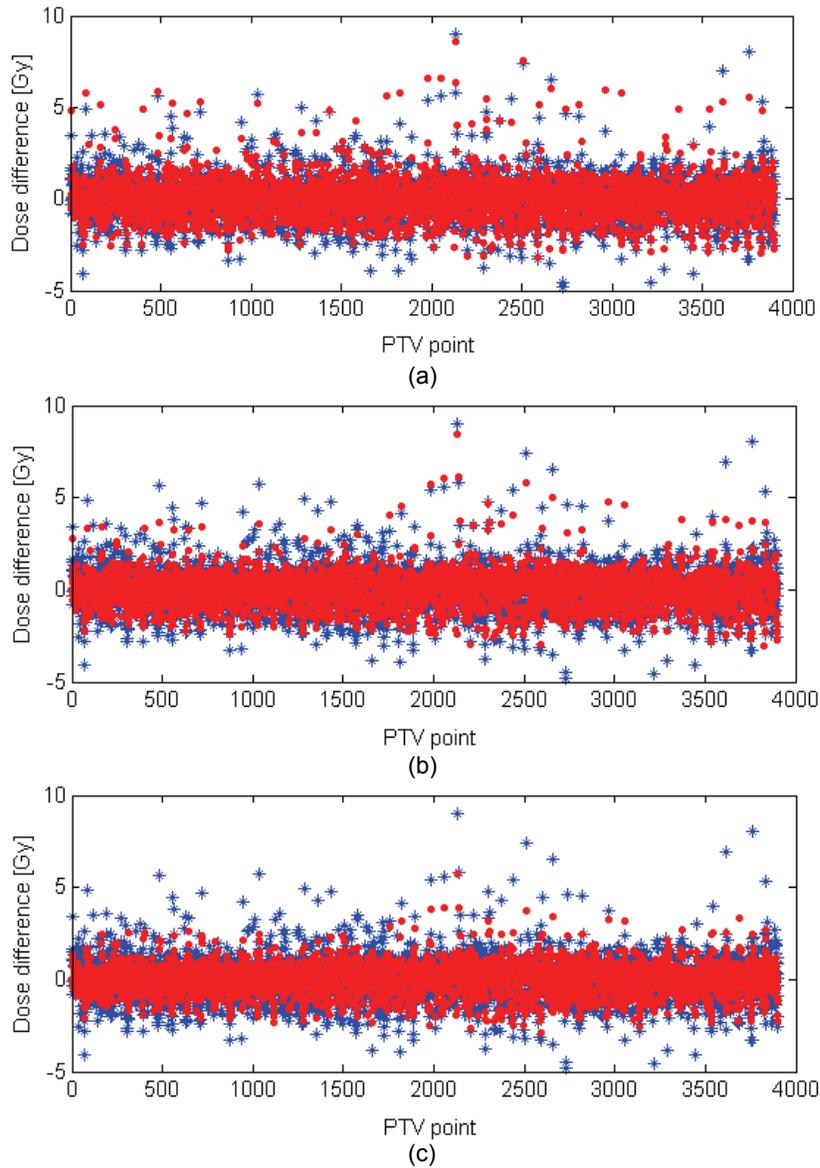


Figure VII-19. PTV point dose differences in the Baseline plan (blue) and the ADS-Gradient (red) (a) Equivalent plan (b) Weighted1 plan and (c) Weighted2 plan.

impact on EUD compared to the baseline plan in this example, although we have noted much larger impact on EUD with more targeted smoothing of radioresistant tumor sub-volumes (discussed in Chapter VI).

Several preliminary conclusions can be drawn from this study on the potential impact of smoothing on geometric sensitivity. First, the biggest source of dosimetric error seen here was due to the shift in isocenter. From there, we observed subtle differences in the dose discrepancies resulting from setup error between the baseline

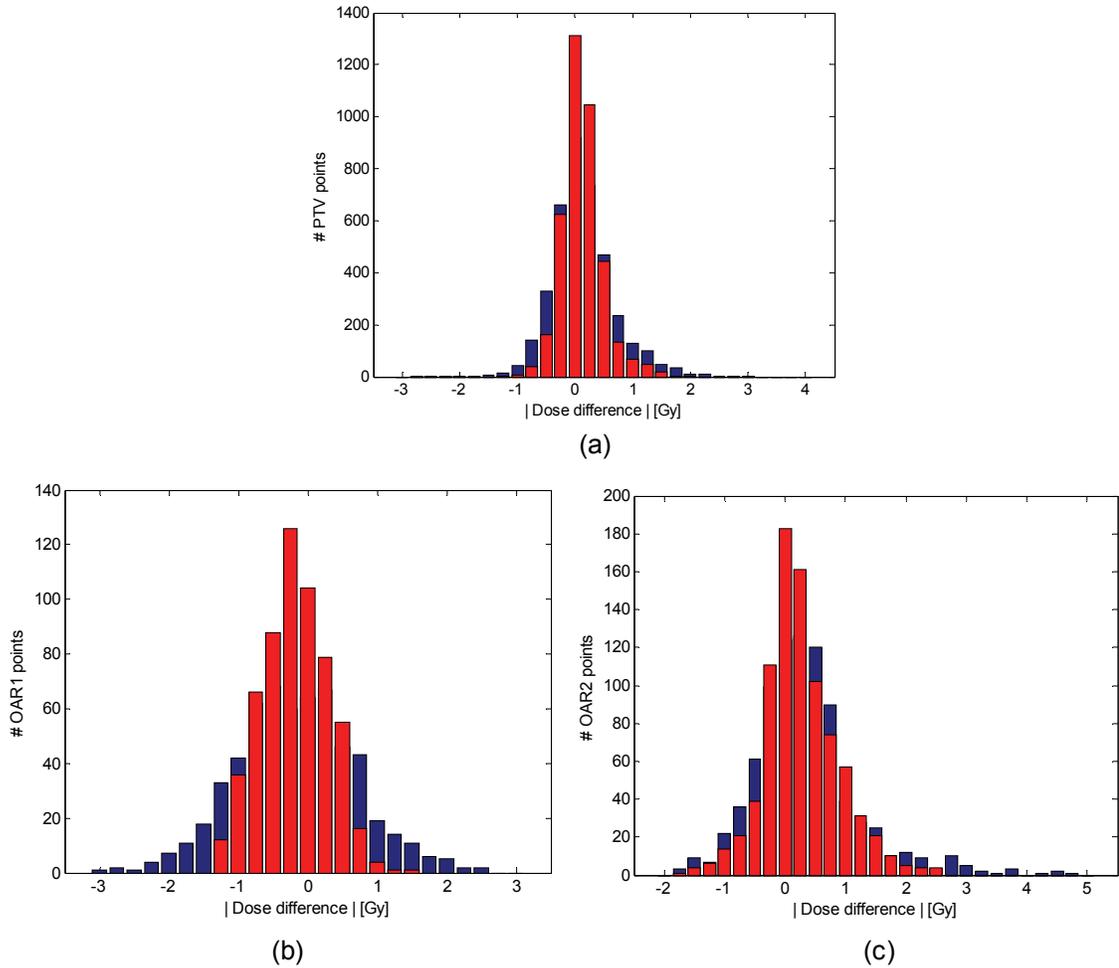


Figure VII-20. Histograms of cumulative point dose differences for the (a) PTV, (b) OAR1, and (c) OAR2 are shown in for baseline plan (blue) and the ADS-Gradient, Weighted2 plan (red).

IMRT plan and plans that included smoothing costlets. All plans had similar dosimetric quality, although the final two ADS-Gradient plans had less modulation. These fields exhibited smaller percentages of fractional and cumulative point dose errors that were greater than 2 Gy. The errors were also of smaller magnitude as the smoothing increased. There did not appear to be an increase in dose to the central cord-like organ at risk as a result of the smoothing.

The results shown here relate only to residual setup errors. A good IMRT protocol should apply daily positioning to correct for errors, and therefore it would be an exaggeration of the effect of setup error to include simulation of errors greater than the residual setup error after correction. However, the results here are also a best-case scenario of the dose differences that occur. Intrafraction organ motion, breathing

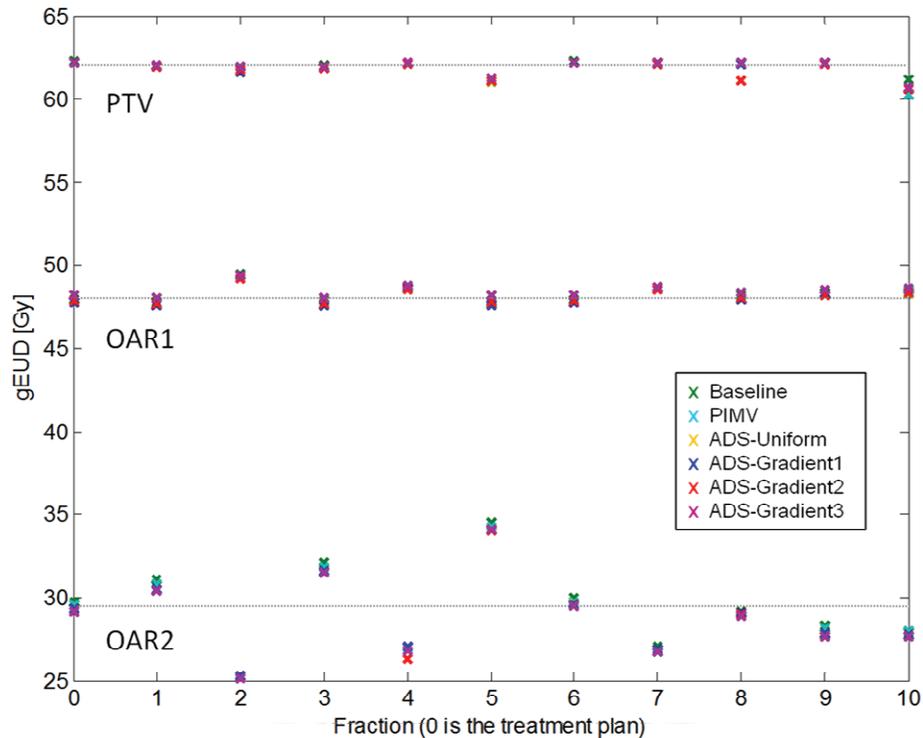


Figure VII-21. Generalized equivalent uniform dose for all structures and techniques as evaluated in each simulated fraction. The horizontal lines approximate the gEUD values in the original plan.

motion, and dose delivery errors will also affect the dose errors. Both of these issues have been shown to result in larger dose differences in complex fields. These reductions in individual and cumulative point dose differences could have a positive impact on the planning of adaptive radiation therapy. The cumulative improvements in the dose agreement between planned and delivered fields due to the impact of smoothing should make it easier to calculate the dose-to-date with confidence, so that appropriate and accurate adaptations can be made to the plan.

VII.D. Potential Impact on Adaptive Radiotherapy

Adaptive radiation therapy is a broad term that can apply to any radiotherapy technique that aims to adapt the original treatment plan or setup to changing information about the patient or even the patient population. Some common areas of adaptive radiotherapy research and clinical implementation include (i) protocols to adapt the patient positioning according to daily imaging (ii) adapting the clinical target volume margins to account for change setup errors, (iii) adapting the patient's treatment plan to account for dose errors

or prescription changes resulting from new information obtained during the treatment course. The previous section demonstrated that standard and smoothed IMRT beams are both fairly robust to small-magnitude residual geometric setup errors, but that fewer and smaller dose discrepancies occurred in the smoother plans. Section VII.B. also showed that smoother plans are more accurate from a delivery standpoint, making it less necessary to account for delivery errors through re-planning. Both of these advantages of smoothed fields should make the dose-to-date calculation for adaptive therapy more accurate due to less uncertainty between the planned, sequenced, and delivered dose distributions. The focus of this section is to evaluate the potential impact that modulation penalties may have on the re-planning of IMRT treatments to account an adaptive dose escalation protocol in a head and neck example.

VII.D.1. Methods

We simulated a three stage adaptive treatment course in a simple head and neck patient for this preliminary study. The three stages are highlighted in Figure VII-22, showing the dose prescription and simulated geometry changes in each stage. The scenario preserves the volume of the patient so weight loss and other possible deformations are not modeled. In the first stage, we plan a case to the PTV, and all objectives are met except for the oral cavity mean dose objective. Halfway through treatment, the PTV volume has shrunk so it is decided to escalate dose to the new volume. We allow a 5 % volume underdosing in the PTV1-PTV2 volume to try to bring down the mean dose to the oral cavity and allow dose escalation. The preservation of the cord and parotid dose and the dose escalation are higher priorities than the oral cavity. Finally, with 25 % of the treatment remaining, the PTV has shrunk again. We continue dose escalation to PTV2 and PTV3, but allow a 20 % volume underdosing in PTV1-PTV2 in hopes of higher dose escalation without degrading the achieved objectives. To simulate the effect of the actual treatment, each stage was recalculated using a small systematic setup error and the actual MLC delivery files were used to simulate delivery artifacts, such as tongue-and-groove effect, that are not modeled in the beamlet dose calculations. Therefore, at the subsequent planning stage, any errors that would degrade the objectives would have to be corrected before dose escalation. The simulated error magnitudes in the left-right, anterior-posterior, and inferior-superior directions are shown in the parenthesis in Figure VII-22. To see if there is any advantage to using the ADS-Gradient penalty during each optimization stage, we planned a baseline treatment

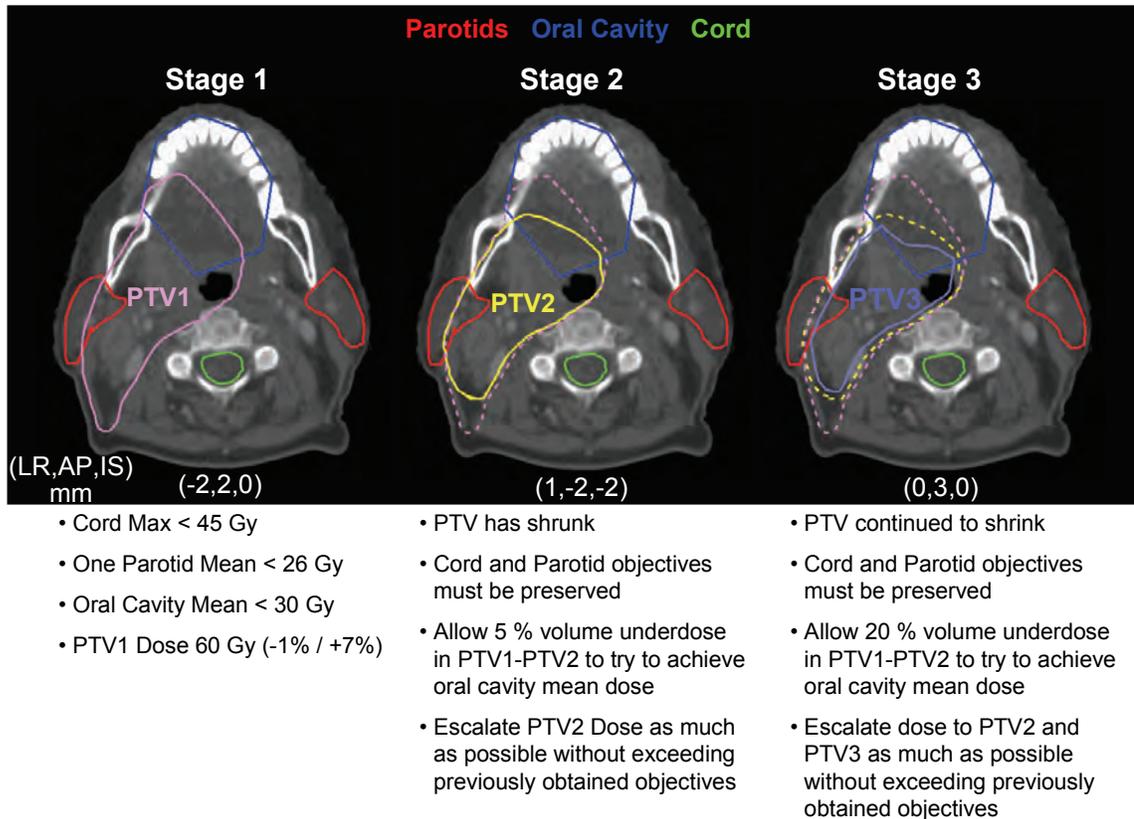


Figure VII-22. The simulated three stage dose escalation adaptive plan for a head/neck case.

course and an ADS treatment course. Both cases had the same simulated errors in each stage.

VII.D.2. Results

The use of ADS during the optimization of the simulated adaptive dose escalation case did not appear to have a significant effect on the ability to safely escalate dose to the tumor, although the dose homogeneity in the escalated volume was better. Figure VII-23 shows the final simulated dose volume histogram after the simulation of each of the 3 stages of treatment for both the baseline and ADS treatment courses. There was a slight increase in mean dose to the spinal cord, although this could have been prevented by constraining the cord dose in the stage 1 optimization. The remaining structures all have very similar DVHs, showing that ADS did not interfere with meeting any of the plan objectives. Similar results were seen in Stage 1 and Stage 2. The biggest advantage of including the ADS penalty during the optimization of an adaptive protocol is the

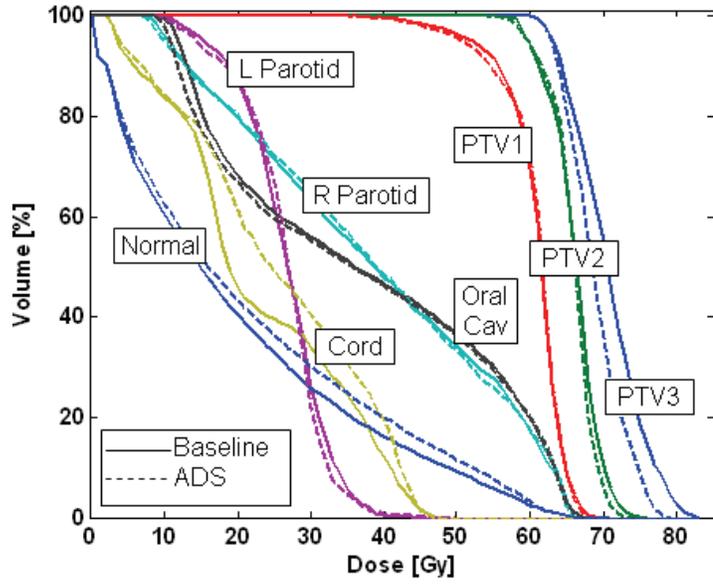


Figure VII-23. Final composite dose-volume histograms for the Baseline and ADS dose escalation treatment courses.

smoothing that occurs in the later stages of planning. In each stage, we observed an increase in the overall modulation and MU of the plan.

Figure VII-24 shows the MU required per fraction for each stage of the optimization for both treatment courses, and Figure VII-25 shows the optimized beams for each stage with and without ADS. The large increase in modulation at the later stages is likely a result of the plans having to compensate for the simulated setup and

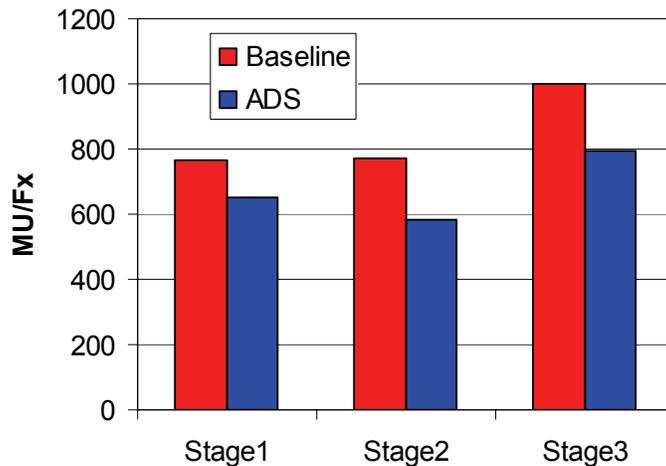


Figure VII-24. The MU required per fraction for each stage of the adaptive dose escalation simulation for the Baseline and ADS treatment courses.

delivery errors to preserve plan quality. These large fluctuations that occur in the later stages of the optimization process could be very detrimental to a real adaptive protocol. We have shown earlier in this chapter that highly complex fields have a higher chance of failing quality assurance which could interrupt treatment and cause significant problems in the clinic if the patient had to be switch to a non-IMRT plan during the middle of therapy.

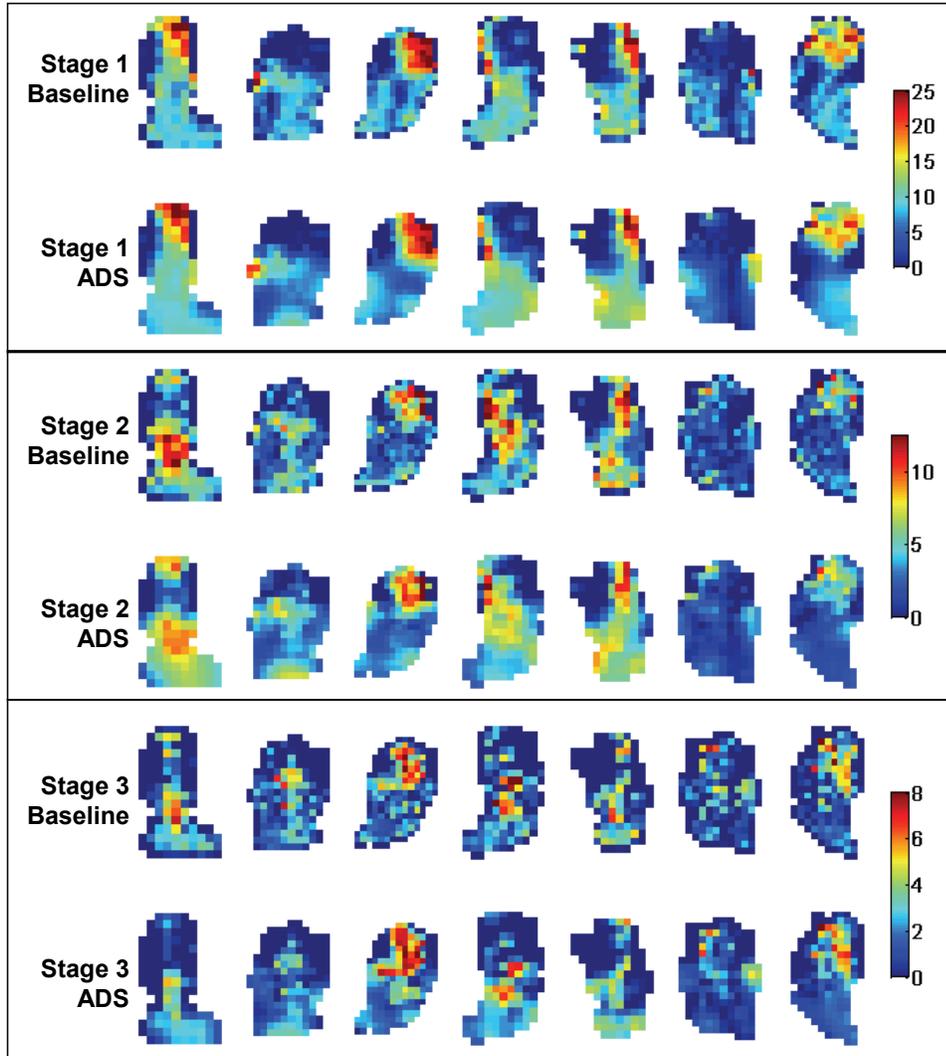


Figure VII-25. Optimal IMRT beams as each stage in the adaptive dose escalation example for the Baseline and ADS treatment courses.

This adaptive therapy simulation has shown that IMRT beams may become highly complex during the later stages of an adaptive protocol due to the combination of the dose prescription changes and the compensation for dosimetric errors due to setup

error and delivery. Many of the positive aspects of smoothing that have been shown in the chapter may potentially improve the efficacy of adaptive therapy protocols. In this example, we showed that ADS could be used during each optimization stage to promote smoothness without significantly affecting the dosimetric quality of the adaptive plan. The beamlet patterns in each treatment stage were smoother and required fewer MU than the baseline beamlet patterns. We note that the baseline treatment course at later stages became highly modulated, which could have negative affect on treatment delivery and quality assurance.

VII.D.3. Discussion

This section has demonstrated that modulation penalties could have a positive impact on the treatment planning of adaptive radiation therapy. In the multi-stage dose escalation study that was investigated, the use of ADS during each stage results in significant decreases in modulation and MU. In later stages, this smoothing may result in a higher probability that the plan will pass IMRT QA. This is due to the considerable modulation observed in the final IMRT planning stages, due to the need to preserve treatment planning objectives under setup errors and delivery artifacts. The use of ADS did not impact the ability to meet the normal tissue objectives or the dose escalation goals. Future studies on the impact of modulation penalties in different types of adaptive therapy scenarios would be very useful to determine whether there are situations in which smoothing has a larger or smaller impact on the final result.

VII.E. Conclusions

This chapter has investigated the potential clinical impact of using modulation penalties inside the IMRT planning objective function. Previous chapters have demonstrated that modulation penalties can significantly smooth IMRT beams with negligible loss in plan quality. This smoothing has resulted in a decrease in the number of monitor units required to a deliver IMRT plans that have been optimized with modulation penalties versus those optimized without modulation penalties. In addition to this increase in delivery efficiency, we have also demonstrated that modulation penalties can have a positive impact on the treatment planning stage in terms of optimization time and robustness to point sampling effects. Plans optimized with smoothing penalties required significantly fewer iterations and less time to optimize than standard IMRT plans. This effect was most noticeable in the PIMV costlet, with a nearly 80 % reduction

in optimization time and iterations. Both PIMV and ADS plans were also shown to be more robust to different random samplings of the regions of interest and to undersampling of the regions of interest. The latter is an especially promising observation since many optimization plans are limited in the number of points that can be defined.

Measurements of standard versus ADS IMRT plans demonstrated that smoother plans are more accurately delivered by the treatment machine. Specifically, film measurements of ADS-optimized plans agreed better to dose calculations than the more modulated standard IMRT plans at the same levels of the dosimetric quality. The ADS plans also exhibited fewer regions of tongue-and-groove underdosing.

Plans optimized with modulation penalties were also shown to be slightly more robust to residual setup errors than standard IMRT plans, although the clinical significance of these differences would have to be verified by more in-depth studies of geometric sensitivity for different treatment sites. Still, our preliminary investigation has shown that smoother plans exhibit smaller and fewer point dose differences in the face of small residual setup errors.

Finally, an adaptive dose escalation example in head-and-neck cancer revealed that ADS smoothing may have a positive impact on the planning and delivery of adaptive radiation therapy. Standard and ADS IMRT plans were able to produce similar dosimetric results in the multi-stage dose escalation example, but the ADS plans required much less modulation in each stage. In the final stages, when the standard IMRT plans become exceedingly modulated, the use of the ADS may be very important to promote smoothing and ensure that plans can be delivered accurately and efficiently. Additional studies on different adaptive therapy scenarios and protocols may be useful to determine whether ADS could have a significant dosimetric impact on adaptive planning.

References

- ¹ M. Alber and F. Nusslin, "Intensity modulated photon beams subject to a minimal surface smoothing constraint," *Phys. Med. Biol.* **45**, N49-N52 (2000).
- ² S. Spirou, N. Fournier-Bidoz, J. Yang, C. Chui, and C. Ling, "Smoothing intensity-modulated beam profiles to improve the efficiency of delivery," *Med. Phys.* **28**, 2105-2112 (2001).
- ³ A. M. Stell, J. G. Li, O. A. Zeidan, and J. F. Dempsey, "An extensive log-file analysis of step-and-shoot intensity modulated radiation therapy segment delivery errors," *Med Phys* **31**, 1593-1602 (2004).
- ⁴ T. R. Mackie, J. W. Scrimger, and J. J. Battista, "A convolution method of calculating dose for 15-MV x rays," *Med. Phys.* **12**, 188-196 (1985).
- ⁵ T. R. Bortfeld, D. L. Kahler, T. J. Waldron, and A. L. Boyer, "X-ray field compensation with multileaf collimators," *Int J Radiat Oncol Biol Phys* **28**, 723-730 (1994).
- ⁶ M. Schwarz, J. Van der Geer, M. Van Herk, J. V. Lebesque, B. J. Mijnheer, and E. M. Damen, "Impact of geometrical uncertainties on 3D CRT and IMRT dose distributions for lung cancer treatment," *Int J Radiat Oncol Biol Phys* **65**, 1260-1269 (2006).
- ⁷ W. A. Tomé and J. F. Fowler, "On cold spots in tumor subvolumes," *Med Phys* **29**, 1590-1598 (2002).
- ⁸ D. Litzenberg, L. A. Dawson, H. Sandler, M. G. Sanda, D. L. McShan, R. K. Ten Haken, K. L. Lam, K. K. Brock, and J. M. Balter, "Daily prostate targeting using implanted radiopaque markers," *Int J Radiat Oncol Biol Phys* **52**, 699-703 (2002).

CHAPTER VIII.

SUMMARY

Cancer is a potentially deadly disease that affects millions of lives across the world. Radiation therapy is an integral part of treatment for many of those suffering from cancer because of its ability to precisely target tumor volumes while sparing as much normal tissue as possible. Radiation therapy technology has grown considerably over the past several decades. Precise targeting and delivery of radiation to the patient is possible with advanced volumetric imaging and radiation delivery techniques. Inverse planned intensity modulated radiation therapy is a technique that allows an optimization algorithm to choose the radiation intensity profile that should be delivered from a variety of different beam directions to best meet the dose prescription goals of the treatment. These goals include the prescription dose for the tumor and dose limits for organs and normal tissue. Inverse planned IMRT has the potential to create treatment plans that likely would not be possible with conventional planning techniques. IMRT is expected to provide a benefit in many areas of cancer treatment—most notably in the treatment of prostate and head and neck cancer—although the advantages of this relatively new technology are still being investigated.

An aspect of IMRT that causes concern, however, is the high level of complexity observed in the beam intensity distributions. Composed of small sub-beams, or beamlets, from each beam direction, the distribution of intensities can be highly non-intuitive, with large fluctuations and apparently random noise. This is a large departure from the uniform intensity beams used in conventional radiation therapy. Because the increase in beam complexity in IMRT is so great compared to conformal radiation therapy, many investigators wonder whether the beamlet patterns created by inverse planning are truly necessary.

In order to investigate the overall value of these remarkably complex treatments in creating high quality radiation therapy plans versus the potentially negative impact of

treating patients with highly complex intensity patterns, this study of IMRT beam complexity and methods for its control was undertaken.

Chapter I contains a brief history of radiation therapy and an overview of the inverse planning process. IMRT planning consists of (i) acquiring 3D imaging of the patient and outlining the important regions of interest (i.e. the tumor and organs) on the 2D slices of imaging data (ii) creating a 3D representation of the regions of interest for the optimization system by placing discrete dose evaluation points throughout each of the regions, (iii) arranging beams around the patient and splitting each beam into a theoretical grid of beamlets, (iv) calculating the dose to the points from each of the unit intensity beamlets, (v) designing a cost function that consists of the treatment planning goals and assigns a penalty for not reaching the goals based on their importance, (vi) using an optimization algorithm to find the intensities of the beamlets that best minimize the total penalty or cost of the cost function, (vii) translating the intensity patterns into a set of instructions to the treatment machine so that it can deliver the intensities using the shaping capabilities of a multi-leaf collimator, and (viii) performing quality assurance and delivering the plan to the patient. The end of Chapter I introduced the problem with IMRT beam complexity and discussed some of the previous work which has addressed this problem.

In order to appreciate the potentially negative consequences of IMRT beam complexity, Chapter II discusses in more detail the reasons why it is undesirable to treat patients with highly complex fields. First of all, complex fields are very difficult to deliver accurately and efficiently by the treatment machine. This delivery inefficiency requires the machine to be on for longer periods of time and results in extra radiation delivered to the patient from leakage. It also prolongs the treatment time, which may reduce the probability of controlling some tumors. The extra time and beam complexity combined can make IMRT treatments sensitive to patient setup error and organ motion. These issues, along with the potential for delivery errors, can not only increase the discrepancy between the planned and delivered dose, but can also increase the uncertainty in predicting or calculating the delivered dose. These potential disadvantages have led to efforts to try to identify some of the sources of IMRT beam complexity.

Several causes of beam complexity are reviewed in the second part of Chapter II. Previous research has shown that IMRT problems can be highly degenerate, which can lead to a great deal of noise in the IMRT intensity patterns. Other sources of additional beam complexity are the optimization algorithm and cost function definition.

Examples of these sources of beam complexity from published literature and the University of Michigan were given. For example, it was shown that the use of a stochastic optimization algorithm can result in an addition of noise into the optimized beamlet patterns. It was also demonstrated that IMRT beamlet patterns are highly dependent on the goals of the treatment plan and the discrete point sampling of the regions of interest. Undersampling the regions of interest can increase the field complexity as well as give a false representation of the difficulty of the plan. In addition, the beamlet patterns are very sensitive to the particular point placement in regions of interest, even when they are properly sampled. The main conclusions of Chapter II were that (i) there are a number of things that can be done to reduce beam complexity, such as using a deterministic optimization algorithm, starting the algorithm from zero beamlet intensities, and properly sampling the regions of interest, and (ii) much of the complexity observed in IMRT comes from the sensitivity of the cost function to the beamlet intensities and the fact that IMRT problems are degenerate. This means that the cost function can be minimized by a variety of different beamlet patterns, and that some intervention should be taken to steer the optimization algorithm to the smoothest possible solution that does not affect the plan quality.

Approaches to minimize and control IMRT beam complexity were presented in Chapters III through V. Chapter III introduced the use of the mathematical surfaces, as opposed to individual beamlets, to represent the IMRT beam. This method was implemented as a way to reduce the degrees of freedom in the optimization problem while also promoting smooth and continuous beamlet patterns. Details regarding the implementation of the infrastructure for basis function optimization into the University of Michigan's optimization system were given. Polynomial surface and composite Gaussian radial basis function sets were studied to represent the IMRT beams. These studies were performed by optimizing the function coefficients or parameters instead of individual beamlet intensities. The use of basis function optimization results in a considerable decrease in the number of optimization variables and was also shown to produce an increase in delivery efficiency. This method was very promising for simple geometries that did not require the full degrees of freedom of high resolution beamlet optimization. However, the method did cause an increase in the normal tissue dose and was not able to meet all of the planning objectives in complicated geometries. These limitations may be overcome by the use of more numerous and more customized basis functions, and there is a great deal of future work that could be done in this area.

However, the limitations in complex cases suggested that a larger number of degrees of freedom were necessary to plan complicated IMRT cases.

Chapter IV described the investigation of two different procedures that could be applied during optimization to reduce beam complexity. The first, which is simple and practical, is to not allow unconstrained intensity values in the optimization. To do this, restrictions were placed on the maximum beamlet intensities allowed in each plan in an effort to remove the high intensity peaks that are often found in IMRT beams. The second method was the use of a modulation penalty inside the cost function to promote smoothness. The use of a modulation penalty inside the cost function, as opposed to smoothing after optimization, allows the effect of the smoothing to be taken into account according to the dosimetric objectives. Chapter IV demonstrated that maximum intensity limits in IMRT are very useful for improving delivery efficiency and removing the high intensity peaks from the IMRT fields. In fact, this is a method that is now routinely used in clinical practice at the University of Michigan. However, restricting the intensity too much can overly constrain the optimization system and prevent it from meeting the objectives. In addition, maximum intensity limits usually resulted in an overall flattening of the IMRT fields, which was found to raise dose to the normal tissues. Therefore, the use of two different types of modulation penalties in the cost function was also investigated. The first penalty, a Savitzky-Golay filtering method, penalized the difference between the original IMRT beam and the beam after least squares filtering with either a 1D or 2D polynomial window. When smoothing was performed in the direction of the MLC motion only, delivery efficiency was improved, but the beam was step-like and non-intuitive in the opposite direction. Smoothing in 2D was superior in terms of overall smoothing and delivery efficiency. However, both methods exhibited poor behavior as their importance was increased in the objective function. This instability occurred because the filtered version of the beam was not always a smoother version of the beam, since the filtering process could introduce unwanted modulation into the field. The second modulation penalty that was developed and tested penalized a measure of the overall field modulation called the plan intensity map variation (PIMV). Use of this penalty was superior to the other methods in terms of delivery efficiency and dosimetric metric quality. Using high penalty weights resulted in an increase of dose to normal tissue but moderate weights resulted in an improvement in delivery efficiency while preserving plan quality.

One of the reasons that high penalty weights could not be used in the modulation penalties in Chapter IV was the fact that these methods penalized all modulation in the field. This meant that even the important intensity gradients, such as those between targets and normal tissues, were penalized. This problem appeared to be the main limitation in smoothing IMRT fields. Therefore, in Chapter V, a novel smoothing procedure that was developed from diffusion principles, called adaptive diffusion smoothing, was introduced. This method uses spatially variant diffusion coefficients that are automatically defined for each beamlet to preferentially smooth the undesirable and unnecessary modulation in the field while preserving important and desirable areas of modulation. By using the sensitivity of the cost function with respect to the beamlets to define the coefficients, IMRT fields could be optimally smoothed using a modulation penalty that penalized the difference between the original beam and diffusion smoothed beam. This penalty differs from the Savitzky-Golay filter penalty because the diffusion smoothed beam is always a smoother version of the original beam and the important modulation is not penalized. The ADS penalty was shown to be superior or equivalent to the PIMV penalty in all cases tested.

The potential of ADS is considerable, due to the ability to customize the diffusion coefficients for different purposes. Chapter VI presented several applications of ADS, including the use of ADS to make controlled tradeoffs between the plan objectives and smoothing with the aid of a multi-criteria optimization method called lexicographic ordering (LO). This method was also used to show the clinical impact that ADS smoothing could have on the treatment of breast cancer. Reduction of leakage radiation involved in these IMRT treatments through ADS smoothing may reduce the potential for inducing secondary cancers in this young patient population.

Chapter VI also demonstrated that ADS could be used to promote smoothing over very critical tumor regions that could be susceptible to high risks from underdosing even small sub-volumes. The use of specialized diffusion coefficients to penalize modulation over these critical areas allows for a clinically significant increase in tumor coverage in face of geometric uncertainties.

Finally, Chapter VII covered the potential clinical impact that smoothing could have on the IMRT planning and treatment process. The use of the modulation penalties in the cost function was shown to make plans more robust to the point sampling effects that were discussed in Chapter II, as well as make the optimization process itself more efficient. It was experimentally verified that IMRT cases that were planned with ADS

could be delivered more accurately than standard IMRT plans and showed better agreement with the dose calculations. Fewer delivery artifacts were observed, and ADS plans showed fewer dose discrepancies that could cause failures in the IMRT plan quality assurance procedure. Chapter VII also showed that plans optimized with modulation penalties show slightly fewer and smaller dose discrepancies after simulated residual setup errors. All of these features will aid in the planning of adaptive therapy, which requires accurate calculation of the actual dose that has been delivered. Chapter VII simulated a simple adaptive dose escalation protocol that demonstrated that beams can become highly complex without the use a smoothing penalty during adaptation. This new complexity occurs because the adapted plan must overcome the dose errors caused in earlier treatments as well as meet the goals for the updated dose prescription. Smoothing may have a very important place in adaptive therapy planning to ensure that plans are not so complex that they may fail IMRT quality assurance or cause significant delivery errors.

This dissertation has focused on the investigation and control of beamlet complexity in intensity modulated radiation therapy. Previous research in this area has focused on leaf sequencing algorithms and non-optimal smoothing procedures inside and outside of the optimization loop. In most cases, earlier smoothing methods have not been able to distinguish between desirable and undesirable modulation and the plan quality ultimately suffered as a result. In this work several new methods that are able to reduce IMRT beam complexity without sacrificing plan quality were introduced. The most advanced method, adaptive diffusion smoothing, can distinguish between desirable and undesirable modulation in the field by using the cost function sensitivity to each of the beamlets in the plan. This method is unique and adaptable to variety of situations in radiation therapy planning.

Future work identified by this dissertation includes the development of more intelligent point sampling schemes in IMRT to reduce the effect that point sampling has on beam complexity. The use of basis function optimization to investigate the use of larger, and more customized basis function sets that will be robust to a variety of different geometries is also a worthwhile area of research to reduce the number of optimization variables in IMRT. Fewer variables result in much quicker optimization times with stochastic optimization methods, and therefore, in plans where significant local minima may be observed in the cost function, basis function optimization may have

an advantage over beamlet optimization with smoothing due to the reduction in the number of variables.

The further development of adaptive diffusion smoothing method for radiobiological targeting of both tumor and normal tissues may be very useful as more and more biological and functional imaging data are being used for input into IMRT dose prescriptions. It may also be reasonable to use adaptive diffusion smoothing to make plans more robust to internal organ motion if the diffusion coefficients can be customized to predictable motion data, such as a breathing trace. Adaptive diffusion smoothing itself could be applied to variety of different problems in IMRT that have not yet been studied due to the unique ability to customize the diffusion coefficients to achieve different clinical goals.

Intensity modulated radiation therapy is an important treatment modality in radiation oncology. Its importance and usefulness will only increase as more high resolution imaging data is used to (i) prescribe heterogeneous dose distributions based on the local properties of the tumor and surrounding tissue and (ii) dynamically change the dose prescription to take into account feedback about radiation damage and response in tumors and normal tissues. The use of the methods proposed in this dissertation can have a positive impact on the efficacy of treating patients with IMRT by reducing the unnecessary complexity in IMRT beams while ensuring that the dosimetric benefit of intensity modulation can be fully achieved.