IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **55** (2010) N521–N531 doi:10.1088/0031-9155/55/22/N01

NOTE

A unified framework for 3D radiation therapy and IMRT planning: plan optimization in the beamlet domain by constraining or regularizing the fluence map variations

B Meng[1,5], L Zhu[4], B Widrow[2], S Boyd[2] and L Xing[1,3]

[1] Department of Radiation Oncology, Stanford University, Stanford, CA 94305, USA
[2] Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
[3] Molecular Imaging Program (MIPS), Stanford University, Stanford, CA 94305, USA
[4] Nuclear and Radiological Engineering and Medical Physics Programs, The George W Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

E-mail: lei@stanford.edu

Received 29 April 2010, in final form 24 September 2010
Published 28 October 2010
Online at stacks.iop.org/PMB/55/N521

Abstract

The purpose of this work is to demonstrate that physical constraints on fluence gradients in 3D radiation therapy (RT) planning can be incorporated into beamlet optimization explicitly by direct constraint on the spatial variation of the fluence maps or implicitly by using total-variation regularization (TVR). The former method *forces* the fluence to vary in accordance with the known form of a wedged field and latter *encourages* the fluence to take the known form of the wedged field by requiring the derivatives of the fluence maps to be piece-wise constant. The performances of the proposed methods are evaluated by using a brain cancer case and a head and neck case. It is found that both approaches are capable of providing clinically sensible 3D RT solutions with monotonically varying fluence maps. For currently available 3D RT delivery schemes based on the use of customized physical or dynamic wedges, constrained optimization seems to be more useful because the optimized fields are directly deliverable. Working in the beamlet domain provides a natural way to model the spatial variation of the beam fluence. The proposed methods take advantage of the fact that 3D RT is a special form of intensity-modulated radiation therapy (IMRT) and finds the optimal plan by searching for fields with a certain type of spatial variation. The approach provides a unified framework for 3D CRT and IMRT plan optimization.

[5] Author to whom correspondence should be addressed.

1. Introduction

Despite the widespread acceptance of intensity-modulated radiation therapy (IMRT) and volumetric-modulated arc therapy (VMAT) (Crooks *et al* 2003, Ma *et al* 2009, Otto 2008, Yu 1995, Zhang *et al* 2010), 3D RT based on open or wedged fields remains a modality of choice for the treatment of some cancer patients because of its relative simplicity in planning and quality assurance (QA). In current practice, 3D planning is still a manual process involving multiple trial-and-error. Considerable effort may be required to compute a clinically acceptable plan and the final results may strongly depend on the planner's experience and understanding of the planning system. In the past decade, a number of automated 3D-planning techniques have been proposed but their clinical acceptance has been less than satisfactory (Beaulieua *et al* 2004, Bedford and Webb 2003, Chen *et al* 2008, Langer and Leong 1987, Lee *et al* 1997, Rosen *et al* 1991, Starkschall 1984, Xing *et al* 1997, 1998). Most, if not all, of these existing methods attempt to find the optimal solution by decomposing an incident field into a sum of open and nominal wedged fields (Dai *et al* 2001, Li *et al* 1999, Milliken *et al* 1996, Petti and Siddon 1985, Xing *et al* 1997), which is, for historical reason, a departure from commonly used inverse plan framework. In this work, we propose wedged field optimization in the beamlet domain to unify the two seemingly different methods of modalities (Xing *et al* 2009). Computationally, this is realized by forcing or encouraging a 3D radiation field to vary monotonically in intensity in a direction perpendicular to the beam axis. Given that much of dose optimization work has already been done in the beamlet domain for IMRT, a wedge optimization proposed here is easily implementable in the existing IMRT inverse-planning platform, which is one of the advantages of the proposed approach over previously developed methods based on the principle of universal wedge or omni-wedge.

2. Method

2.1. Dose optimization in the dose beamlet domain

An incident beam is divided into a collection of 1×1 cm beamlets. The delivered dose distribution, d, depends linearly on the intensity of the beamlets, x:

$$d = Ax \tag{1}$$

where x is a 1D vector consisting of beamlet intensities of the contributing fields. Each column of matrix A is a beamlet kernel, corresponding to the dose distribution achieved by a beamlet with unit intensity. The beamlet kernels were pre-computed based on the patient model derived from the patient's simulation CT images, the treatment machine settings, and the beam geometry by using a voxel-based Monte Carlo algorithm (VMC++) (Bogner *et al* 2009, Kawrakow 1997, Rickhey *et al* 2008). Although the size of matrix A is typically very large, it is sparse and reduction of computation cost and memory usage is possible. The sum of the square errors of the delivered dose relative to the prescribed dose is used as the optimization objective function. The treatment-planning problem is cast as

minimize

$$(Ax - d)^T (Ax - d) \tag{2}$$

subject to

$x \geqslant 0$

x is achievable using the wedge constraints.

The solution to the above problem depends on the machine delivery constraints. Two different but related methods are implemented to find the optimal 3D RT plan in the described

beamlet domain: (i) a direct constraint that *forces* the fluence to follow a wedge field; and (ii) a TVR which *encourages* the derivative of the beam fluence to be piece-wise constant (or sparse second derivatives). The fluence behaves either like a wedged or segmentally wedged field when the derivative of the fluence is piece-wise constant.

2.2. Plan optimization using a direct wedge field constraint

For a given field, we can restrict the spatial variation of the beamlet intensity by applying a mathematical constraint during optimization. While the constraint can take an arbitrary form, here we enforce the fluence to vary in the form of a wedged field with a linear change in intensity along a direction perpendicular to the central axis of the beam. The vector normal to the wedged field is defined by $(1, a_f, b_f)$, and the relative weight of the field is defined by c_f, where the subscript f is the field index. The corresponding optimization problem becomes

minimize

$$(Ax - d)^T (Ax - d) \tag{3}$$

subject to

$$x \geqslant 0$$

$$x_{u,v,f} + a_f \times u_f + b_f \times v_f + c_f = 0$$

$$u_f = \{1, 2, \ldots, N_u\}, \quad v_f = \{1, 2, \ldots, N_v\}, \quad f = \{1, 2, \ldots, N_f\}$$

where the beamlet intensity x is parameterized by the variables u_f, v_f and f, and the variable u_f (v_f) is the row (column) index of the beam intensity for the field indexed by f. N_u is the total number of possible MLC leaf positions for each leaf, N_v is the total number of MLC leaf pairs of the field and N_f is the number of fields. We assume that each treatment field defining the dose matrix has a rectangular shape when it is fully open. Note that the rectangular fields define only the calculation matrix, and the actual beamlets that we 'turned on' are determined by the BEV projections of the PTV. The system variables include the beamlet intensity x and (a_f, b_f, c_f) for each field. It is important to note that equation (3) can be formulated into a quadratic problem, and solved using the standard quadratic programming (Boyd and Vandenberghe 2004). The algorithm was implemented in Matlab, using the MOSEK optimization software package (http://www.mosek.com) (Zhu *et al* 2008, Zhu and Xing 2009).

2.3. Treatment planning using L-1 TV regularization

The fluence map of a wedged field has a constant first derivative or is sparse in the domain of second derivatives. This fact can be utilized to construct an L-1 norm of the second derivatives of the intensity map to help us to find a solution with sparse second derivatives of the fluence. Along this line, we reformulate the optimization problem to

minimize

$$(Ax - d)^T (Ax - d) + \beta \sum_{f=1}^{N_f} \sum_{u_f=3}^{N_u} \sum_{v_f=3}^{N_v} \left(\left| x_{u_f,v_f,f} + x_{u_f-2,v_f,f} - 2x_{u-1,v_f,f} \right| + \left| x_{u_f,v_f,f} + x_{u_f,v_f-2,f} - 2x_{u_f,v_f-1,f} \right| \right) \tag{4}$$

subject to

$$x \geqslant 0$$

where the parameter β is an empirical constant controlling the strength of the L-1 regularization. Because a wedged field has constant derivatives on the plane defined by the MLC leaves, the absolute values of the second derivatives of each beamlet in both u and v directions are included in equation (4). The second term of an L-1 norm does not appear to be linear or quadratic. However, the above optimization can be reformulated as a quadratic programming problem (Boyd and Vandenberghe 2004). The problem is then efficiently solved using a standard quadratic programming, similar to the study described in 2.2.

Although not exactly equivalent, the L-1 norm regularization in equation (4) encourages constant second derivatives in both u and v directions, which plays a similar role as the wedge constraint on the intensity maps, especially when the value of regularization constant β is large. In general, however, optimization of x in (4) does not guarantee a strict wedged field and a segmented wedge-shaped fluence map (i.e. a fluence map with a few pieces having different slopes may be resulted). A fitting can be applied to approximate the fluence map to a single wedge field, at the cost of slightly degraded dose distribution as compared to the plan before the fitting. In practice, a fluence map with piece-wise constant slopes can be easily delivered by using a dynamic jaw or MLC even without being transformed into a single-slope wedge field, which will be discussed in section 4.

2.4. Evaluation

Two clinical cases—a brain case and a head and neck case—were used to test the proposed methods. In the brain patient, five fields with gantry angles at 0°, 55°, 90°, 235° and 270° were used. The treatment isocenter was placed at the center of PTV. Each field contained 14 $\times$ 10 beamlets, with a beamlet size of 1 $\times$ 1 cm at the isocenter plane. To save computational time in the MC dose kernel calculation, the CT images were downsampled to a voxel size of 2.8125 $\times$ 2.8125 $\times$ 3 mm. The resultant dose distributions, DVHs and other dose statistics including maximum and mean doses were used to assess the treatment plans obtained using the two different methods.

For the head and neck patient, seven fields (20°, 120°, 145°, 180°, 215°, 240° and 340°) were used and each field contained 8 $\times$ 8 beamlets. Other parameters of the dose calculation and plan evaluation procedure were the same as those of the brain case study.

3. Results

Working in the beamlet domain allows us to model the spatial variation of the beam fluence naturally and efficiently. Both constrained wedge field optimization and TV regularization are capable of providing clinically sensible 3D RT solutions with monotonically varying fluence maps. The results for the two cases are summarized in the following.

3.1. Brain case

After the dose distribution of each beamlet was calculated using MC simulation, the proposed direct wedge constraint algorithm and the L-1 norm algorithm took about 21 and 10 s, respectively, on a 2 GHz PC for the brain case. Figure 1 shows the optimized beamlet intensity of the first incident field with the direct wedge constrained L-1 norm regularization, respectively. The L-1 norm drives the beam intensity map to converge to piece-wise constant slopes (figure 1(b)), which is consistent with that obtained from the direct wedge constraint method (figure 1(a)). In obtaining the results, the regularization constant was set to be

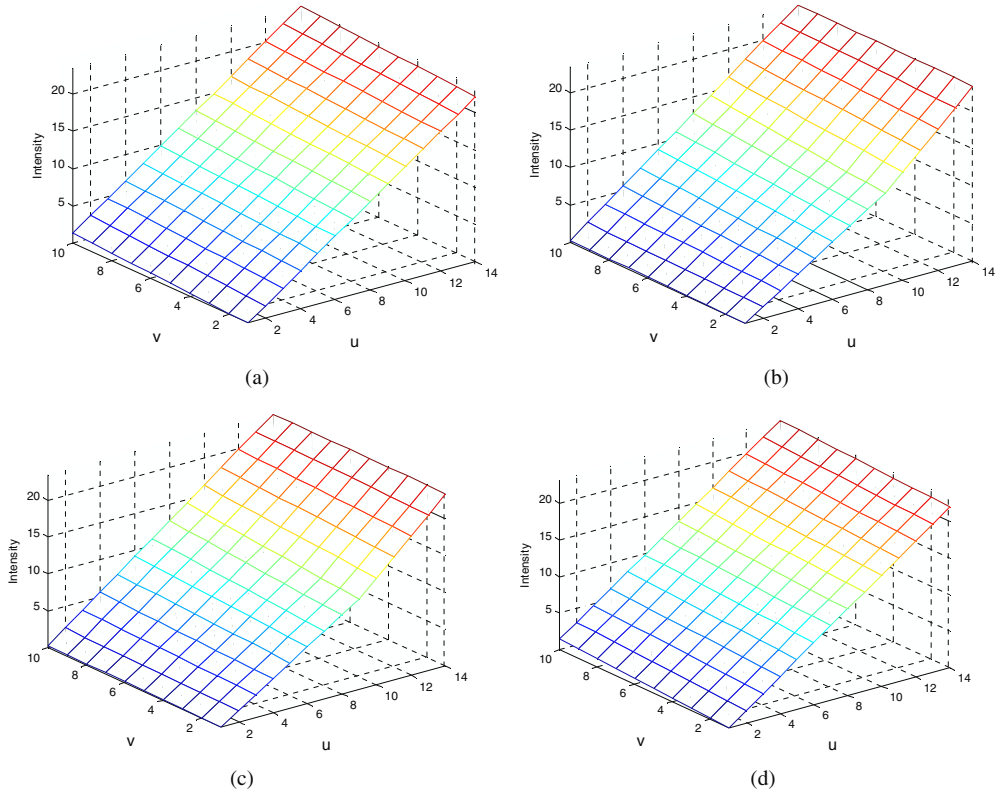

Figure 1. Optimized beam intensity map of the first field for a brain study obtained with direct wedge constraint (a) and total-variation regularization (b)–(d) with $\beta = 0.01$,0.1,0.5 respectively.

$\beta = 0.01$ in the objective function (3). When we increase β, stronger regularization is applied to the influence map. The results are shown below.

To illustrate the role of L-1 TV regularization, figure 2 shows the second derivatives in longitudinal and lateral directions without and with L-1 norm regularization. Note that the scales of the two histograms are largely different. The second derivatives of intensity map using no regularization can be up to 100, and it reduces to within 0.04 with the use of L-1 norm regularization.

Figure 3 shows the DVHs of the PTV obtained using the two different methods. Axial dose distributions obtained by using the two approaches are plotted in figure 4. For delivery based on a single-slope wedge, each of the fluence maps resulted from TVR is fitted to a wedged field and the final DVH after fitting is slightly compromised, as indicated in the dotted dash curve in figure 3. Overall, the final DVHs and dose distribution are very close to the ones without fluence fitting, indicating that any of the three methods can yield similar results. It is seen that the L-1 method is nearly as powerful as the direct wedge constraint norm method. The wedge fitting slightly degrades the DVH of the original L-1 regularization-based dose optimization. Although the direct wedge constraint method and the L-1 norm regularization method provide similar DVHs, the maximum dose intensity from the L-1 norm method is slightly lower than the other, because the L-1 norm method takes non-uniform second derivatives into consideration.

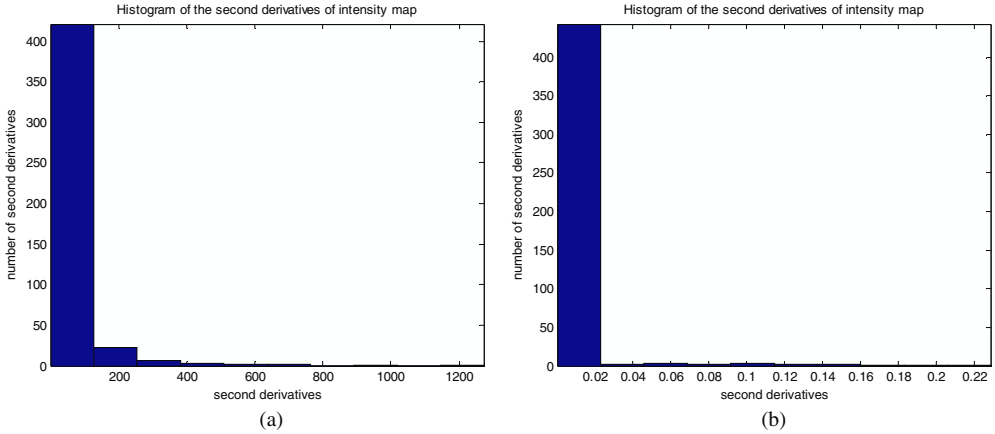

Figure 2. Histograms of the second derivatives of the beam intensity in both lateral and longitudinal directions without regularization (a) and with L-1 TV regularization (b).

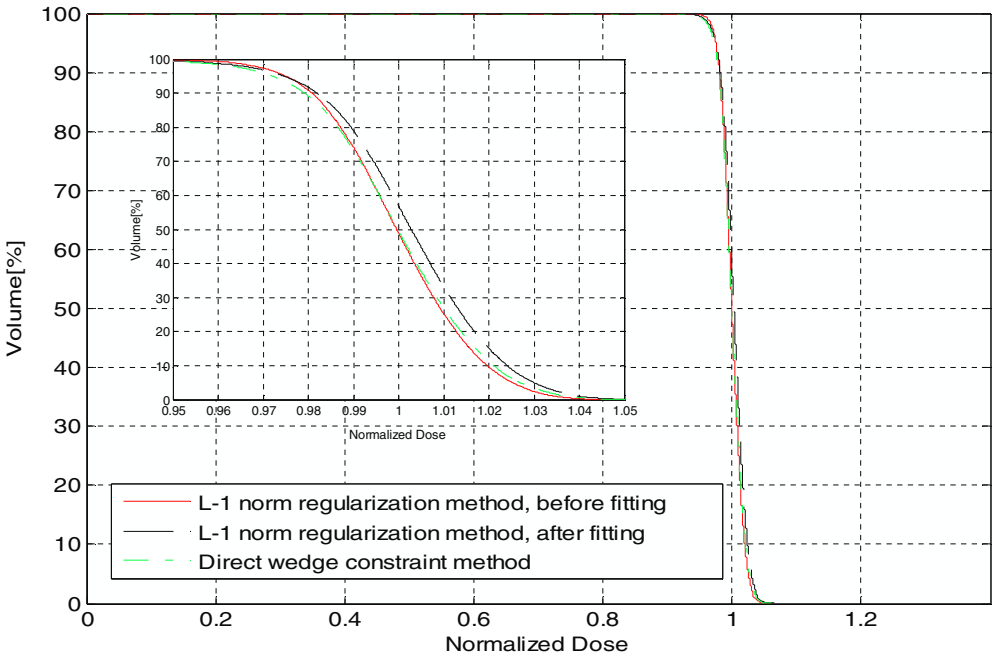

Figure 3. PTV DVHs of a five-field 3D brain RT plan obtained by using the direct wedge constraint method and the L-1 TVR method.

3.2. Head and neck case study

Figure 5 shows the first fluence map obtained by using the direct wedge constraint method and the TVR method for the head and neck patient. While the intensity maps obtained using the two approaches are very similar, some subtle differences between them are observed. The direct wedge constraint method confines the fluence map to a single slope, whereas the TVR method provides fluence maps with slowly changing second derivatives.

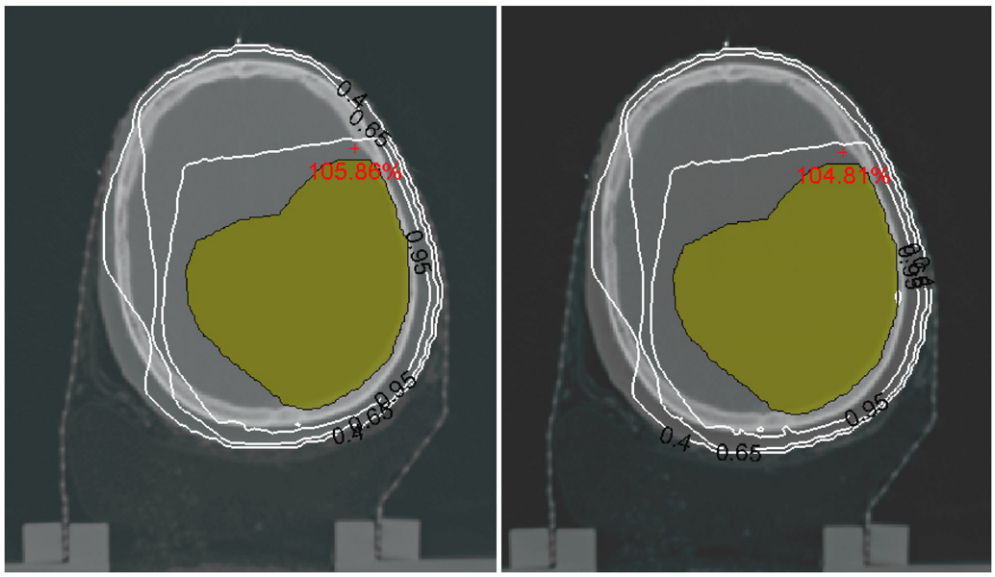

Figure 4. Dose distributions obtained using the direct wedge constraint method (left) and the TVR method (right). The iso-dose lines correspond to 95%, 65% and 40% of the prescribed dose.

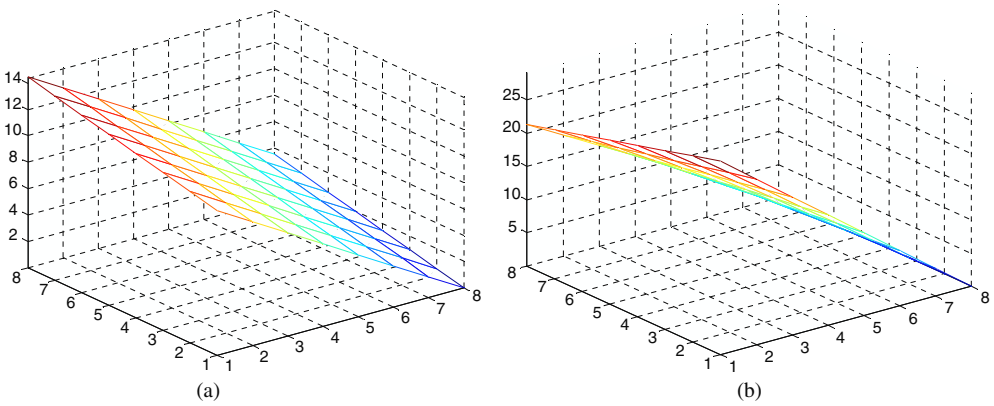

Figure 5. Optimized beam intensity map of the first field for the head and neck study obtained with direct wedge constraint (a) and total-variation regularization (b), respectively.

The PTV DVHs obtained by using the proposed methods are plotted in figure 6. Figure 7 shows the iso-dose distributions obtained using the TVR optimization and the direct wedge constraint optimization algorithms. Once again, it is seen that the direct wedge constraint method and the TVR method yield very similar results, suggesting that both methods can be employed for 3D RT plan optimization. Not surprising, approximating fluence map with a piece-wise constant slopes by a single-slope wedged field degrades the PTV doses slightly. But they all yield highly conformal and essentially identical dose distributions (the L-1 norm regularization method beats the direct wedge constraint method by about 0.4% of the maximum dose intensity).

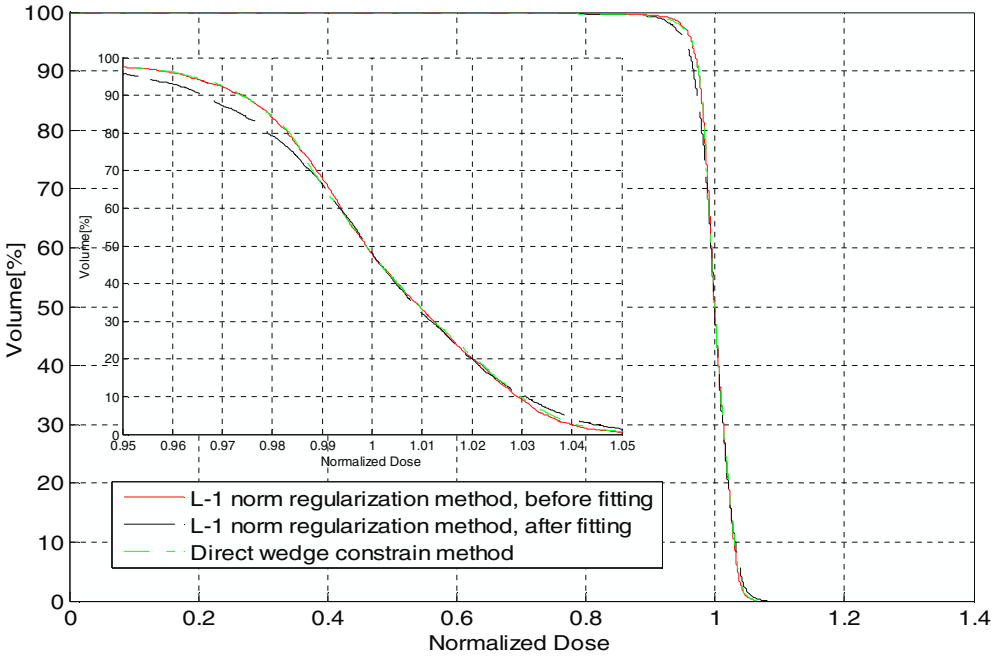

Figure 6. PTV DVHs resulted from the direct wedge constraint method and the TVR method for the seven-field head and neck case.

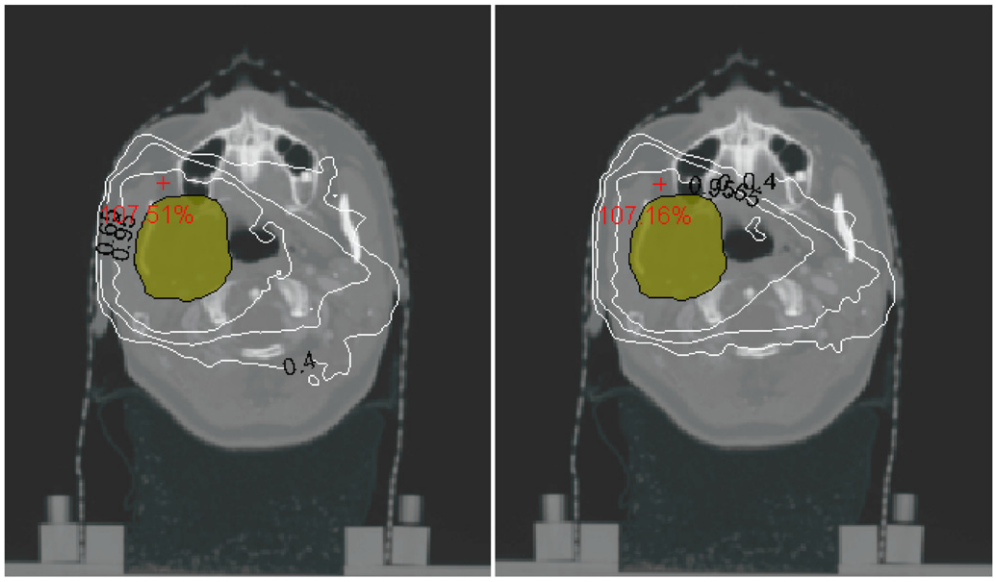

Figure 7. Dose distributions of the head and neck plans obtained using the direct wedge constraint method (left) and the TVR method (right). The iso-dose lines correspond to 95%, 65% and 40% of the prescribed dose.

4. Discussion

3D RT is a special form of IMRT and it is clinically sensible to have a unified platform for the planning and delivery of the two approaches, which have been separated for historical reason. Working in the beamlet domain allows us to combine the two types of plan optimizations naturally, which is more advantageous over methods based on the principles of universal wedge (Petti and Siddon 1985) or omni-wedge (Li *et al* 1999, Xing *et al* 1997). Beamlet domain optimization unites IMRT inverse planning and 3D RT and makes it straightforward to implement 3D RT planning on an existing IMRT plan platform.

Two methods have been developed for automated 3D RT planning in the beamlet domain: direct wedge constraint method and TVR-based optimization. Both approaches are capable of providing clinically sensible 3D RT solutions with monotonically varying fluence maps. For dose delivery based on the use of customized physical or dynamic wedges, the constrained wedge field optimization seems to be more useful because the optimized fields are directly deliverable using wedge filters. The TV-based optimization, on the other hand, may yield fields possessing piece-wise constant derivatives, which translates into two or more segmented wedged fields and cannot generally be realized without approximating the field by the closest wedge filter. Overall, we found that the resultant dose distributions of the two approaches are very similar. Computationally, the TV-based approach is slightly more efficient.

In general, the use of a wedge-shaped fluence map compromises the resultant dose distribution because of the reduced degree of freedom as compared to an arbitrarily shaped fluence map. Practically, the need for a wedge-shaped fluence map, instead of an arbitrarily shaped fluence map, arises in two situations. Constraining the fluence map to wedge shapes is necessary in 3D CRT with customized wedges. In this case, the optimization algorithm proposed here is useful because it has the potential to automate the planning process. Surprisingly, after more than a decade of IMRT, even in an advanced radiation oncology clinic with all linacs equipped with MLCs and IMRT capabilities, there are still a huge number of patients receiving 3D CRT treatments. Being able to automate the planning for this patient population is thus of practical significance. Furthermore, the proposed algorithm can produce plans with non-customized wedge angles/orientations, which can provide additional insight to the planner as to the level of improvement if non-customized wedge fields were used for treatment of the planned patient. For machines (including cobalt machines) without IMRT capability, wedged fields are obviously the only choice. We note that the methodology of constraining fluence maps to a known form is quite general and provides an effective way for incorporating prior knowledge of the incident beams into the treatment plan optimization process.

In current clinical practice, only customized wedge filters are employed. The situation is not changed even with the introduction of dynamic wedges, which produce a wedge-shaped field based on the dynamic movement of the jaw according to a pre-commissioned STT (segmented treatment table). A relevant question raised here is whether it is necessary to have pre-commissioned wedge filters in future 3D RT. Given the widespread acceptance of IMRT and dynamic delivery, it seems more natural to replace the use of customized wedge filter by individually optimized fluence maps that are easily deliverable with dynamic jaw or MLC for patient treatment. The benefits of this type of treatment include, but are not limited to, a more optimal treatment plan, reduced efforts in commissioning and routine QA of wedge filters. Moreover, abandoning the use of physical wedges will greatly simplify the design of the LINAC system and improve the clearance between the LINAC head and the patient. At this point, this type of quasi-IMRT treatment may not satisfy the needs of all clinical situations. For example, the size of the dynamic wedge field may not be as large as that of a physical

wedge. The relative orientation of the dynamic wedge field with respect to the MLC leaves is another common concern. A natural solution is to use MLC for both wedged field production and beam shaping. In a broad sense, the system we are dealing with becomes a field-in-field IMRT. This type of intensity-modulated field is readily deliverable using currently available dynamic MLC, but requires us to abandon the conventional 'wisdom' of wedged fields. With the recent advancement in dynamic MLC-based delivery, such as the auto-movement of MLC carriage, simultaneous MLC leaf movement and the jaw movement for reduction of inter-leave leakage, we foresee that the need for customized physical or dynamic wedges is diminishing in radiation oncology practice. The need for wedged fields in radiation therapy will be eventually diminished as IMRT planning, delivery and QA becomes more and more efficient and robust.

This work presents a novel application of the general TVR and compressed sensing techniques (Boyd and Vandenberghe 2004, Candes and Wakin 2008, Zhu and Xing 2009). The classical Shannon–Nyquist sampling theorem specifies that to avoid losing information when capturing a signal, one must sample at least two times faster than the signal bandwidth. In many applications, the Nyquist rate is so high that too many samples result. In RT treatment planning, increasing the sampling rate (i.e. the loosing fluence constraints) is either impractical or clinically less favorable. Compressed sensing provides a practically valuable approach for finding optimal solutions with under-sampled data. The 3D RT plan optimization, in which only open or wedged fields are employed, is an example of using reduced sampling (system variables) to produce a clinically sensible dose distribution. A key issue here is to find a domain where the content is sparse. We pointed out, for the first time, that the sparsity of fluence maps is in the domain of the second derivatives. The calculation is similar to that of compressed sensing-based IMRT inverse planning (Zhu and Xing 2009), where the sparsity is reflected in the domain of first derivative of the fluence maps with a goal of finding piece-wise constant fluence maps.

Finally, we note that a linear fluence with an arbitrary slope and orientation is used to represent a wedge filter. In reality, depending on the specific design of the wedge used, the fluence slope may deviate slightly from the ideal linear variation. Both methods can be extended to adapt the specifics of the wedge filters by treating the measured fluence variation as prior knowledge, which is then used in a way analogous to the linear variation assumption used in this work. In the 'futuristic' 3D RT with individualized wedges as discussed above, this issue vanishes because the form of spatial fluence variation of a 'wedge filter' can be specified by the user.

5. Conclusion

An inverse-planning framework that unites 3D RT and IMRT planning has been established. Two independent methods, L-1 TVR-based optimization and direct wedge constraint optimization, are investigated and a similar level of success has been achieved. The proposed methods take advantage of the fact that 3D RT is a special form of IMRT and finds an optimal plan by searching for fields with a certain type of spatial variation in the beamlet domain. Beamlet domain 3D RT plan optimization can be easily implemented in an existing IMRT-planning platform and may have practical implications of eliminating the need for customized wedge filters, which can potentially reduce efforts in the commissioning and routine QA of wedge filters.

Acknowledgment

This project was supported in part by National Cancer Institute (1R01 CA104205 and 5R01 CA98523).

References

Beaulieua F *et al* 2004 Simultaneous optimization of beam orientations, wedge filters and field weights for inverse planning with anatomy-based MLC fields *Med. Phys.* **31** 1546–57

Bedford J L and Webb S 2003 Elimination of importance factors for clinically accurate selection of beam orientations, beam weights and wedge angles in conformal radiation therapy *Med. Phys.* **30** 1788–804

Bogner L *et al* 2009 Fast direct Monte Carlo optimization using the inverse kernel approach *Phys. Med. Biol.* **54** 4051–67

Boyd S and Vandenberghe L 2004 *Convex Optimization* (New York: Cambridge University Press)

Candes E and Wakin M 2008 An introduction to compressive sampling *IEEE Signal Process. Mag.* 21–30

Chen G P, Ahunbay E and Li X A 2008 Automated computer optimization for 3D treatment planning of breast irradiation *Med. Phys.* **35** 2253–8

Crooks S M, Wu X, Takita C, Watzich M and Xing L 2003 Aperture modulated arc therapy *Phys. Med. Biol.* **48** 1333–44

Dai J, Zhu Y and Wu X 2001 Verification of the super-omni wedge concept *Phys. Med. Biol.* **46** 2447–55

Kawrakow I 1997 Improved modeling of multiple scattering in the Voxel Monte Carlo model *Med. Phys.* **24** 505–17

Langer M and Leong J 1987 Optimization of beam weights under dose-volume restrictions *Int. J. Radiat. Oncol. Biol. Phys.* **13** 1255–60

Lee S *et al* 1997 Conformal radiotherapy computation by the method of alternating projections onto convex sets *Phys. Med. Biol.* **42** 1065–86

Li J, Boyer A and Xing L 1999 Clinical implementation of wedge filter optimization in 3D radiotherapy treatment planning *Radiother. Oncol.* **53** 257–64

Ma Y, Popple R, Suh T and Xing L 2009 Beam's-eye-view dosimetrics-guided inverse planning for aperture-modulated arc therapy *Int. J. Radiat. Oncol. Biol. Phys.* **75** 1587–95

Milliken B D, Hamilton R J and Rubin S J 1996 The omni wedge: a method to produce wedged fields at arbitrary orientations *Med. Phys.* **23** 337–42

Otto K 2008 Volumetric modulated arc therapy: IMRT in a single gantry arc *Med. Phys.* **35** 310–7

Petti P L and Siddon R L 1985 Effective wedge angles with a universal wedge *Phys. Med. Biol.* **30** 985–91

Rickhey M, Moravek Z and Bogner L 2008 Inverse treatment planning and integration of segmentation procedures *Z. Med. Phys.* **18** 163–9

Rosen II *et al* 1991 Treatment plan optimization using linear programming *Med. Phys.* **18** 141–52

Starkschall G 1984 A constrained least-squares optimization method for external beam radiation therapy treatment planning *Med. Phys.* **11** 659–65

Xing L *et al* 1997 Optimization of relative weights and wedge angles in treatment planning *Med. Phys.* **24** 215–21

Xing L *et al* 1998 A three-dimensional algorithm for optimizing beam weights and wedge filters *Med. Phys.* **25** 1858–65

Xing L, Lee L and Timmerman R D 2009 *Image Guided and Adaptive Radiation Therapy* (Baltimore, MD: Lippincott Williams & Wilkins)

Yu C X 1995 Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy *Phys Med. Biol.* **40** 1435–49

Zhang P *et al* 2010 Optimization of collimator trajectory in volumetric modulated arc therapy: development and evaluation for paraspinal SBRT *Int. J. Radiat. Oncol. Biol. Phys.* **77** 591–9

Zhu L, Lee L, Ma Y, Ye Y, Mazzeo Y and Xing L 2008 Using total-variation regularization for inverse planning with field specific numbers of segments *Phys. Med. Biol.* **53** 6653–72

Zhu L and Xing L 2009 Search for IMRT inverse plans with piece-wise constant fluence maps using compressed sensing techniques *Med. Phys.* **36** 1895–905