Dose-shaping using targeted sparse optimization

George A. Sayre and Dan Ruan[a)]
Department of Radiation Oncology, University of California – Los Angeles School of Medicine, 200 Medical Plaza, Los Angeles, California 90095

(Received 10 September 2012; revised 16 May 2013; accepted for publication 16 May 2013; published 12 June 2013)

Purpose: Dose volume histograms (DVHs) are common tools in radiation therapy treatment planning to characterize plan quality. As statistical metrics, DVHs provide a compact summary of the underlying plan at the cost of losing spatial information: the same or similar dose-volume histograms can arise from substantially different spatial dose maps. This is exactly the reason why physicians and physicists scrutinize dose maps even after they satisfy all DVH endpoints numerically. However, up to this point, little has been done to control spatial phenomena, such as the spatial distribution of hot spots, which has significant clinical implications. To this end, the authors propose a novel objective function that enables a more direct tradeoff between target coverage, organ-sparing, and planning target volume (PTV) homogeneity, and presents our findings from four prostate cases, a pancreas case, and a head-and-neck case to illustrate the advantages and general applicability of our method.

Methods: In designing the energy minimization objective ($E_{\text{tot}}^{\text{sparse}}$), the authors utilized the following robust cost functions: (1) an asymmetric linear well function to allow differential penalties for underdose, relaxation of prescription dose, and overdose in the PTV; (2) a two-piece linear function to heavily penalize high dose and mildly penalize low and intermediate dose in organs-at risk (OARs); and (3) a total variation energy, i.e., the L_1 norm applied to the first-order approximation of the dose gradient in the PTV. By minimizing a weighted sum of these robust costs, general conformity to dose prescription and dose-gradient prescription is achieved while encouraging prescription violations to follow a Laplace distribution. In contrast, conventional quadratic objectives are associated with a Gaussian distribution of violations, which is less forgiving to large violations of prescription than the Laplace distribution. As a result, the proposed objective $E_{\text{tot}}^{\text{sparse}}$ improves tradeoff between planning goals by "sacrificing" voxels that have already been violated to improve PTV coverage, PTV homogeneity, and/or OAR-sparing. In doing so, overall plan quality is increased since these large violations only arise if a net reduction in $E_{\text{tot}}^{\text{sparse}}$ occurs as a result. For example, large violations to dose prescription in the PTV in $E_{\text{tot}}^{\text{sparse}}$-optimized plans will naturally localize to voxels in and around PTV-OAR overlaps where OAR-sparing may be increased without compromising target coverage. The authors compared the results of our method and the corresponding clinical plans using analyses of DVH plots, dose maps, and two quantitative metrics that quantify PTV homogeneity and overdose. These metrics do not penalize underdose since $E_{\text{tot}}^{\text{sparse}}$-optimized plans were planned such that their target coverage was similar or better than that of the clinical plans. Finally, plan deliverability was assessed with the 2D modulation index.

Results: The proposed method was implemented using IBM's CPLEX optimization package (ILOG CPLEX, Sunnyvale, CA) and required 1–4 min to solve with a 12-core Intel i7 processor. In the testing procedure, the authors optimized for several points on the Pareto surface of four 7-field 6MV prostate cases that were optimized for different levels of PTV homogeneity and OAR-sparing. The generated results were compared against each other and the clinical plan by analyzing their DVH plots and dose maps. After developing intuition by planning the four prostate cases, which had relatively few tradeoffs, the authors applied our method to a 7-field 6 MV pancreas case and a 9-field 6MV head-and-neck case to test the potential impact of our method on more challenging cases. The authors found that our formulation: (1) provided excellent flexibility for balancing OAR-sparing with PTV homogeneity; and (2) permitted the dose planner more control over the evolution of the PTV's spatial dose distribution than conventional objective functions. In particular, $E_{\text{tot}}^{\text{sparse}}$-optimized plans for the pancreas case and head-and-neck case exhibited substantially improved sparing of the spinal cord and parotid glands, respectively, while maintaining or improving sparing for other OARs and markedly improving PTV homogeneity. Plan deliverability for $E_{\text{tot}}^{\text{sparse}}$-optimized plans was shown to be better than their associated clinical plans, according to the two-dimensional modulation index.

Conclusions: These results suggest that our formulation may be used to improve dose-shaping and OAR-sparing for complicated disease sites, such as the pancreas or head and neck. Furthermore, our objective function and constraints are linear and constitute a linear program, which converges to the global minimum quickly, and can be easily implemented in treatment planning software. Thus, the

authors expect fast translation of our method to the clinic where it may have a positive impact on plan quality for challenging disease sites. © *2013 American Association of Physicists in Medicine*. [http://dx.doi.org/10.1118/1.4808363]

I. INTRODUCTION

Dose-volume histogram (DVH) metrics provide a benchmark to characterize and assess the dose delivered to organs-at-risk (OARs). Widely used as a plan quality evaluation criterion, DVH properties are often "prescribed" by the radiation oncologists to indicate the desired tumor coverage or normal-organ sparing for the specific site under treatment, requiring that a specified volume-fraction of a contoured structure is below or above a specified dose.[1] For example, 95% of the planning target volume (PTV) in prostate cases must be at or above the prescription dose.[2] For organs-at-risk, DVH constraints require that the fractional volume receiving a specific dose is smaller than a given value to ensure proper normal tissue recovery. Such constraints are designed to reflect the parallel/serial property and radiobiological property of the OARs within the region of interest.[3] Parallel organs, such as the liver, are able to tolerate damage better to a subvolume than serial organs, such as the spinal cord.

Dose-volume constraints are intuitive to prescribe but difficult to implement and are analogous to the value-at-risk problem, which has been proved to be NP hard.[4] Optimization with respect to such criteria/constraints is computationally costly, unscalable, and nonconvex. Therefore, *dose-volume* constraints are often approximated with more practical *dose-based* functionals to make optimization tractable and convex.[5] In the clinic, it usually requires several iterations of inverse optimization by an experienced planner to satisfy DVH constraints.[6] A more fundamental limitation of the DVH, arguably, arises from the statistical nature of the DVH itself. DVHs are degenerate in the sense that plans with the same (or similar) dose-volume histograms may exhibit markedly different spatial dose distributions.[7] In fact, this insight is reflected in the oncologists' practice of interactively reoptimizing to adjust the isodose contours, rather than DVH curves, in the plan refinement process. This distinction is important, because dose-planning physicians use both DVH and spatial information to determine plan quality.[2,8,9]

Research into improving spatial dose distributions (dose-shaping) has been mostly focused on voxel-based importance weight assignment in the objective function.[10–13] These voxel-based methods enable the exploration of a larger Pareto surface than the conventional structure-based approaches. However, a systematic method to utilize these high degrees of freedom remains largely elusive—the voxel-based importance factors are often adjusted based on either subjective planner experience[10] or automated heuristic schemes.[11–13] In addition, these approaches aim at dose-profiling and do not directly target specific geometric properties of the dose distribution, such as the spatial distribution of dose gradients and hot/cold spots.

These limitations are often addressed in the clinic with several empirical strategies to control dose falloff and hot spots. Most of these methods impose dose constraints on virtual volumes, i.e., volumes with no physical basis that are created for the sole purpose of improving planning results. These techniques only differ in their approach to constructing the virtual volumes and whether *a priori* information is needed to improve plan quality. In one strategy,[14] virtual critical structures are constructed in sensitive regions outside the target where hot spots occurred in an initial optimization step, allowing dose planners to directly reduce the size and magnitude of hot spots. However, this method requires the dose planner to assign a maximum dose constraint to each virtual volume, a step that further complicates the iterative process of dose-planning and possibly reduces the sparing of other critical structures. Another strategy, target volume expansion,[15] applies dose constraints to serial expansions of the PTV in an effort to control dose falloff and has similar drawbacks.

In this study, we propose to address these limitations by formulating a novel objective function using robust cost functions that encourage a sparse set of violations that implicitly shape the spatial distribution of dose values and dose gradients, and provide a flexible method for their control. Furthermore, we will provide a proof of principle for our method using four IMRT prostate cases, one IMRT pancreas case, and one IMRT head-and-neck case. We will accomplish proof of principle by: (1) describing the mathematical motivation and logic behind our method; (2) providing qualitative and quantitative comparisons of dose-shaping performance with the clinical plans; and (3) discussing potential applications and future work.

II. MATERIALS AND METHODS

II.A. Ideal dose-shaping properties: Motivation for objective function development

In general, a good treatment plan delivers: (1) homogeneous dose to the PTV; and (2) sharp dose falloff away from the homogeneous region at the boundary of the PTV. When the PTV and OARs overlap or are relatively adjacent, full PTV coverage with the prescription dose cannot be achieved without sacrificing OAR protection, necessitating a proper allocation of dose gradients. Furthermore, the underlying physiological, physical, and mechanical characteristics of the human anatomy, x-ray beam energy deposition, and treatment machine capability may further limit the achievable spatial dose distribution.[9,16] The rigidity of these tradeoffs varies from site-to-site and patient-to-patient, and depends on factors such as: patient size, PTV-OAR overlap, and PTV-shape concavity.[8,17,18] For example, OARs with substantial PTV-overlap cannot be spared as well as OARs with small overlap, and horseshoe PTVs are harder to treat without compromising OARs than spherically or cylindrically shaped PTVs. However, despite the different challenges presented by different disease sites and individual plans, the fundamental goals and tradeoffs remain the same.

II.A.1. Fundamental optimization goals

The goals of treatment planning may be condensed into two qualitative objectives that are common to inverse optimization in radiotherapy: (1) conformity to the prescription dose in the vast majority of voxels, i.e., general conformity, for all regions; and (2) gradient control in the PTV(s) to help control the spatial dose distribution of the PTV(s), including hot spots.

General conformity to the prescription dose in each region is mainly dictated by overlap[8] and the geometric proximity of PTV and OAR surfaces. This geometric proximity, or relative adjacency, occurs when PTV and OAR surface voxels are separated by at most a few centimeters and are coupled by the dose deposition matrix (see the Appendix). In prostate cases, the PTV almost always overlaps with the bladder and rectum. In the pancreas case, several organs overlapped or adjoined the PTV, including the small bowel, kidneys, stomach, liver, and spinal cord. The head-and-neck case exhibited similar behavior, with the parotid glands, mandible, spinal cord, and brain stem overlapping or adjoining at least one of three PTVs. In these cases, it is much harder to deliver the prescription dose to the target(s) while sparing the OARs than in cases where the target(s) and OARs are separated by large distances.

Conventional radiotherapy optimization typically uses quadratic cost functions to penalize deviations from prescription in the PTV and OARs. As a result, maximum dose falloff is not achieved, except when PTVs and OARs are far apart, because quadratic penalties do not allow large deviations from prescription with respect to each of these structures. By contrast, robust penalty functions enable sharp dose drop-off to OARs by permitting a small number of relatively large deviations, i.e., *outliers*, to arise near the PTV's surface. The resulting solutions are said to have sparse violation with respect to prescription since most voxels conform to the prescription dose while a small fraction may deviate markedly from prescription. While there is a possibility that outliers may group together to form cold spots in PTV-OAR overlaps and/or OAR-adjacent regions on the PTV's surface, we do not anticipate any risk to achieving tumor control. The competing PTV and OAR costs will decrease the likelihood of such cold spots forming in the CTV by localizing them close to the PTV's periphery in overlap or relatively adjacent regions.

To increase PTV spatial homogeneity while largely maintaining OAR-sparing, we propose a novel application of the total variation energy that penalizes the spatial dose gradient in the PTV with a robust cost function. In doing so, we encourage the development of two classes of voxels in the PTV: (1) zero-gradient voxels that arise mainly in the PTV's interior and improve homogeneity; and (2) nonzero gradient voxels that arise mainly on the prescription-dose isocontour and encourage dose-falloff to the OARs.

II.A.2. Choice of robust cost function

In regression, a robust cost function is a cost function that is not unduly affected by a small set of extreme values, i.e., outliers, which deviate from normality.[19] For optimization, the robust cost function is chosen to balance the tradeoff between: (1) desirable outlier-sparing behavior and (2) algorithmic stability and efficiency. In our preliminary implementation, we chose to use the L_1 norm, because: (1) it is associated with a Laplace distribution of violations, which is more robust to outliers than the Gaussian distribution;[20] and (2) its convexity (though not strict) and linear form enables fast linear program (LP) solvers. Furthermore, L_1-based solutions would serve as an appropriate initial guess for iterative solutions of L_0-based objective functions[21–23] should they be deemed preferable in the future.

II.B. Development of objective function

In this subsection, we will prescribe appropriate robust cost functions to mathematically describe the behavior outlined in Sec. II.A. We begin by presenting a generic formalism for an objective function E_{tot}, as the sum of a data fidelity term and a regularization energy,

$$E_{\text{tot}} = E_f + E_r. \tag{1}$$

The data fidelity term E_f characterizes the discrepancy between the achieved distribution and prescription—decreasing this energy encourages general conformity to objective prescriptions. The regularization term E_r is used to impose *a priori* knowledge about solution behavior on the optimization process. In this study, the data fidelity term was used to penalize deviations from the prescription doses of the PTV(s) and OAR(s) to achieve desired target coverage and OAR-sparing, and the regularization term was used to penalize deviations of the spatial dose gradient in the PTV from zero to improve dose homogeneity in the PTV. Equations (2a)–(2d) reflect these objectives,

$$E_{\text{tot}}^{\text{sparse}} = \overbrace{\underbrace{\sum_{j\in\text{PTV}} E_{\text{PTV}_j}}_{E_{\text{PTV}}} + \underbrace{\sum_{m\in\text{OAR}} E_{\text{OAR}_m}}_{E_{\text{OAR}}}}^{E_f} + \overbrace{\underbrace{\sum_{j\in\text{PTV}} E_{\text{TVR}_j}}_{E_{\text{TVR}}}}^{E_r}, \tag{2a}$$

$$E_{\text{PTV}_j} = \frac{1}{N_{\text{PTV}_j}} \sum_{i=1}^{N_{\text{PTV}_j}} \max \begin{pmatrix} \alpha_l(D_L - d_i), \\ 0, \\ \alpha_h(d_i - D_H) \end{pmatrix}, \tag{2b}$$

$$E_{\text{OAR}_m} = \frac{1}{N_{\text{OAR}_m}} \sum_{i=1}^{N_{\text{OAR}_m}} \max \begin{pmatrix} \beta_l(d_i), \\ \beta_h(d_i - D_x) \end{pmatrix}, \tag{2c}$$

$$E_{\text{TVR}_j} = \frac{\gamma}{N_{\text{PTV}_j}} \left\| \nabla d_{\text{PTV}_j} \right\|_1. \tag{2d}$$

The first two terms on the right-hand side of Eq. (2a) are fidelity terms and encourage conformity with PTV and OAR dose prescriptions, respectively. The third term is the total variation energy E_{TVR}, a regularization term that encourages PTV homogeneity. Together, these terms produce a novel formulation that improves tradeoff between DVH fidelity and PTV homogeneity by permitting large penalty violations, i.e., outliers, to develop in regions where competing planning goals may be maximally benefitted. In doing so, overall plan quality is improved.

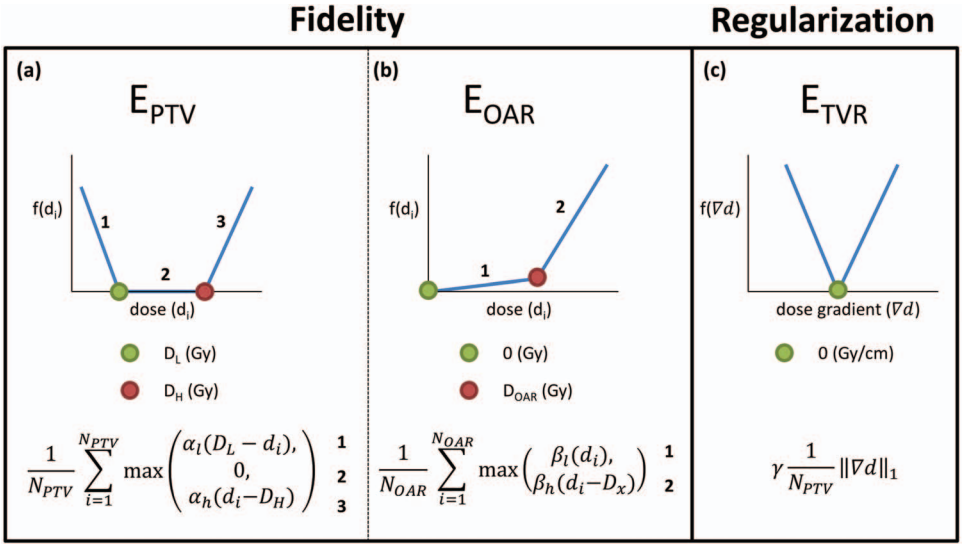

FIG. 1. Penalty descriptions: (a) PTV voxels are penalized by a linear well function that is made more severe by increasing weights α_l and α_h and more discriminating by decreasing D_H–D_L; (b) OAR voxels are penalized heavily above the elbow D_{OAR} with weight β_h, shown as a darkly colored circle, and lightly below it with weight β_l; and (c) spatial dose deviations in the PTV are penalized by the L_1 norm with weight γ. D_x is the offset needed for the high-dose leg and low-dose leg to intersect at the E_{OAR} elbow.

Comprehensive descriptions of E_{PTV}, E_{OAR}, and E_{TVR} for radiotherapy were gathered into Fig. 1 to visually illustrate Eq. (2a). The term E_{PTV} is a linear well function with two nonzero "legs" that bookend a central region that is not penalized, providing differential penalties for underdose, relaxation of prescription dose, and overdose, respectively. The mathematical form of E_{PTV} has several parameters that describe: (1) the slopes, α_l and α_h, of the low-dose and high-dose legs; and (2) the "underdose" prescription, D_L, and the "overdose" prescription, "D_H." Increasing the value of α_l and α_h increases the penalty on those voxels that deviate below D_L and above D_H, respectively, thereby reducing deviation from prescription. Decreasing the range of relaxation, i.e., $D_H - D_L$, makes E_{PTV} more discriminating, and drives the PTV dose distribution closer to uniform while leaving less flexibility for OAR-sparing. Increasing E_{TVR} weight γ also increases PTV homogeneity.

E_{OAR} is a linear piecewise function with two legs: (1) a low-dose leg that lightly penalizes doses less than D_{OAR}; and (2) a high-dose leg that heavily penalizes doses above D_{OAR}. The offset value D_x is defined such that the high-dose leg and low-dose leg of E_{OAR} intersect at D_{OAR}. Increasing the slope β_h increases the penalty on those voxels with dose above D_{OAR}, the point beyond which doses are considered toxic or otherwise undesirable. The slope β_l is made small to provide a mild driving force towards prescription for low-dose or unimportant areas. This piecewise behavior allows more beams to pass through less sensitive OAR regions to improve dose-shaping around PTV surfaces and reduce overall risk of organ complication. Here, the goal is not to approximate E_{OAR} or E_{PTV}, as an L_2 norm as in Ref. 24, but rather to improve tradeoff between the OARs and PTVs.

E_{PTV}, E_{OAR}, and E_{TVR} exhibit L_1-type behavior, or *generalized-L_1* behavior, because their functional forms may be recast as L_1 norms with linear constraints. Therefore, they retain the robust error-handling of L_1 norms and are naturally suited to balance PTV homogeneity with OAR-sparing using large deviations, or outliers, from dose and gradient prescriptions of the PTV and OARs.

II.B.1. Outlier allocation

Figure 2 depicts a simplified but illustrative scenario of tradeoff between a PTV, shown as a large circle, and an OAR, shown as a small ellipse, and demonstrates how minimizing the proposed cost function would allocates outliers, corresponding to cold and hot spot of potential concerns. If low-dose outliers, i.e., large deviations below D_L, were to arise outside of the overlap region, e.g., in the star in Fig. 2, then they would incur an increase in cost to E_{PTV} and E_{TVR}. However, if low-dose outliers were to arise inside the overlap region, e.g., in the triangle in Fig. 2(a), then they would incur the same increase to E_{PTV} and E_{TVR} but reduce E_{OAR}. Thus, it is more favorable in terms of $E_{\text{tot}}^{\text{sparse}}$ cost for low-dose outliers to arise in overlap regions where they will improve OAR-sparing, rather than in nonoverlapping regions where no benefit is achieved. Similarly, high-dose outliers localize away from PTV-OAR overlaps and regions of relative adjacency to help spare OARs of high dose, as illustrated in Fig. 2(c), where high-dose outliers are shown to incur an additional cost to E_{OAR} if they arise in the overlap region. The majority of gradient outliers will arise with low-dose outliers in PTV-OAR overlap regions and with high-dose outliers in insensitive regions to minimize $E_{\text{tot}}^{\text{sparse}}$ via increased OAR-sparing.

The interplay of energy terms in Fig. 2 generalizes to cases of multiple PTVs and OARs with two caveats: (1) low-dose outliers localize predominantly to PTV-OAR overlaps of more important OARs, i.e., those OARs given higher priority weights by the dose planner; and (2) high-dose outliers still primarily localize away from PTV-OAR overlaps, but will

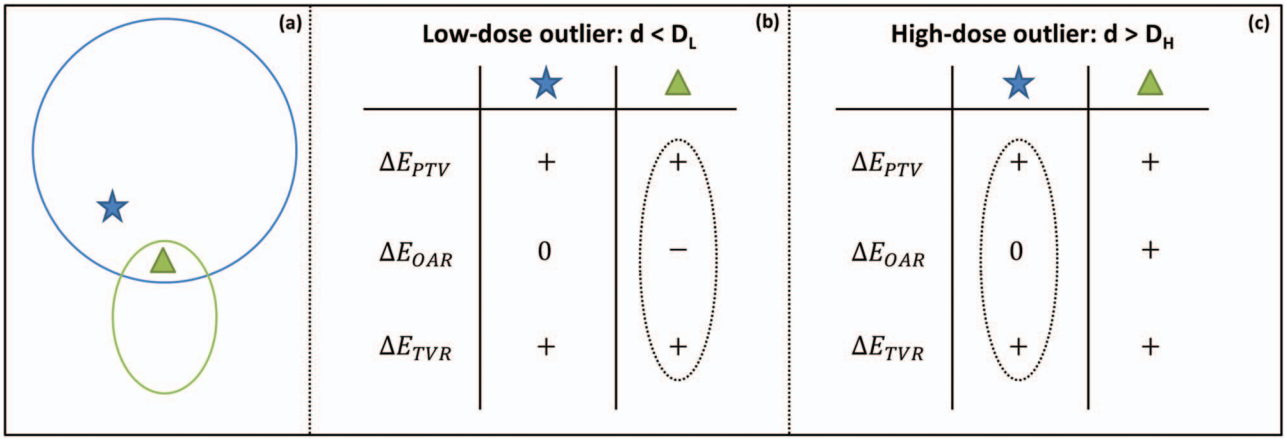

FIG. 2. Illustration of outlier localization: (a) geometry of simple, instructive scenario, (b) low-dose outlier localization, and (c) high-dose outlier localization. The charts in (b) and (c) depict how low- and high-dose outliers, respectively, will increase or decrease the cost functions of our formulation. A "+" indicates an increase in cost, a "0" indicates no change in cost, and a "−" indicates a decrease in cost. The outliers will arise in locations that minimize overall cost. Therefore, low-dose outliers arise in the overlap region and high-dose outliers arise outside of the overlap region.

arise in PTV-OAR overlaps of less important OARs should that net a reduction in $E_{\text{tot}}^{\text{sparse}}$. The majority of gradient outliers will still arise where low- and high-dose outliers do to minimize $E_{\text{tot}}^{\text{sparse}}$. For example, in prostate cases low-dose outliers will localize first to the rectum and second to the bladder and high-dose outliers will localize first to nonoverlapping regions in the PTV and second to the bladder, because the rectum is given higher priority than the bladder. Similarly, gradient outliers will localize with low-dose outliers first to the rectum and second to the bladder and with high-dose outliers first to nonoverlapping regions in the PTV and second to the bladder.

II.B.2. Balancing PTV homogeneity and OAR-sparing

The dose planner must decide how relatively important PTV homogeneity and OAR-sparing are for a particular plan. As discussed in Sec. II.B.1, OAR-sparing is dependent upon the low-dose outliers that arise in overlap and adjacent regions. In fact, maximal OAR-sparing is achieved by penalizing E_{TVR} lightly and penalizing the high-dose leg of E_{PTV} lightly or negligibly, so that large low-dose and gradient outliers may develop in overlap regions. Unfortunately, a plan with maximal OAR-sparing may also exhibit substantial PTV heterogeneity with large and prominent hot spots since large high-dose and gradient outliers are also allowed to develop when E_{TVR} is penalized lightly and the high-dose leg of E_{PTV} is largely ignored. As a result, such a plan would not be acceptable for treatment. Moderately increasing the magnitude of E_{TVR} would decrease the magnitude of hot and cold spots to some extent while largely maintaining the spatial distribution of the maximal-sparing plan. Moderately increasing the magnitude of the high-dose leg of E_{PTV} would alter the spatial distribution of hot and cold spots—and thus OAR-sparing—and have a more pronounced effect on reducing their magnitude than achieved by E_{TVR}. Heavily penalizing either cost function would yield a homogeneous PTV with reduced OAR-sparing. By tuning the relative importances of E_{TVR} and the high-dose leg of E_{PTV}, dose planners can achieve their desired balance between PTV homogeneity and OAR-sparing.

II.B.3. Comparison with previous LP techniques

Most previous LP methods were developed to decrease optimization time. In some cases, e.g., in the case of Romeijn *et al.*,[24,25] convex cost functions were approximated as piecewise linear functions with the intent of decreasing optimization time. In other cases, the aim was to make the problem convex, and thus solvable, for a specific application, such as adaptive planning[26] or Pareto navigation.[27] In contrast, our method was designed to address the underlying tradeoffs in radiotherapy planning, and this led to the development of our novel generalized L_1 formulation with total variation regularization. Each parameter of our objective function has physical implications, unlike the fitted values obtained by approximating convex functions with piecewise linear penalties, and may be adjusted in a straightforward manner to balance PTV homogeneity and OAR-sparing.

II.B.4. Optimization details

The sparse objective function $E_{\text{tot}}^{\text{sparse}}$ is a linear program and was solved using a primal-dual interior-point algorithm[28] with CPLEX, which was chosen for its robustness in solving large-scale, real-world problems.

II.C. Quantitative metrics

We introduce two quantitative metrics to be used in evaluating PTV dose distributions: (1) one-sided standard deviation (σ_+) and (2) one-sided root-mean-squared error (RMSE_+). These metrics describe homogeneity (σ_+) and overdose (RMSE_+), and are defined using the mean dose of voxels above the prescription dose, $\bar{d}_+$, and the prescription dose, D_p, respectively, as shown in Eqs. (3) and (4) below. Deviations below D_p are not penalized, because voxels below D_p

are likely needed to spare OARs. These metrics do not neglect target underdose since our plans were normalized to achieve: (1) the 95% clinical target coverage criterion for the prostate and pancreas cases; and (2) similar or better target coverage than the clinically optimized plans for each PTV in the head-and-neck case. Thus, our plans were on equal footing with the clinical plans in terms of target coverage so any gains in OAR-sparing by our plans could be clearly judged,

$$\sigma_+ = \sqrt{\frac{1}{|N_+|}\sum_{n\in N_+}(d_n - \bar{d}_+)^2}, \tag{3}$$

$$\mathrm{RMSE}_+ = \sqrt{\frac{1}{|N_+|}\sum_{n\in N_+}(d_n - D_p)^2}, \tag{4}$$

where N_+ is the set of voxels above D_p.

Quantification of OAR dose distributions for all cases was performed through DVH plots.

II.D. Clinical plan details

The prostate cases did not include seminal vesicles and were optimized with seven 6-MV beams located at 0°, 40°, 80°, 120°, 240°, 280°, and 320°. The pancreas case was optimized with seven 6-MV beams located at 0°, 50°, 100°, 150°, 210°, 260°, and 310°. The head-and-neck case included simultaneous integrated boost (SIB) and was optimized with nine equi-spaced 6-MV beams starting at 0°. The clinical planning objectives for each site are shown in Table I above and a fine-resolution fluence grid (2.5 × 2.5 mm) was used to optimize each case. The prescription doses were as follows: (1) 72 Gy for the prostate cases; (2) 45 Gy for the pancreas case; and (3) 69.3, 62.7, and 56.1 Gy for PTV_1, PTV_2, and PTV_3, respectively, for the head-and-neck case. PTVs were contoured by a dose-planning physician.

II.E. Clinical plan conversion to CERR framework

All clinical plans were optimized using Varian's Eclipse treatment planning system by an experienced dose planner, and clinical dose maps were calculated using the AAA algorithm.[29] We utilized CERR's QIB algorithm[30] for dose calculation, because it was not possible to obtain the pertinent dose deposition matrices from Eclipse. As such, algorithmic dissimilarity between Eclipse AAA and CERR QIB was a potential issue. Specifically, AAA and QIB, which are both based on the convolution superposition principle, differ in how they address heterogeneity: QIB assumes an infinite medium of water while AAA accounts for heterogeneity using the planning CT image. As such, for tumor sites with relatively low tissue heterogeneity, this algorithmic distinction does not necessarily translate to appreciable dose-calculation differences between AAA and QIB. However, it was still prudent to convert the Eclipse plans into the CERR framework to mitigate any potential dose-calculation bias.

To make this conversion, we used a three-step process. First, we imported the Eclipse beam angles and planning CT into CERR and calculated A_{CERR}, the global dose deposition matrix of CERR, using CERR QIB. Second, we imported the 3D clinical dose distribution d_{Eclipse} into CERR. Third, to match d_{clin} as closely as possible in CERR, we minimized the sum-of-squares error between: (1) d_{Eclipse} and (2) the matrix-vector product of A_{CERR} and $x_{\mathrm{Eclipse}\rightarrow\mathrm{CERR}}$, the optimal fluence map for d_{clin} in the CERR framework. Sufficient dose-matching was achieved in all cases, as evidenced by the body-wide root mean squared errors for the prostate cases (average 1.12 Gy), pancreas case (1.38 Gy), and head-and-neck case (1.41 Gy). Furthermore, good agreement between Eclipse and Eclipse→CERR fluence maps and DVHS was also observed, as characterized by the pancreas case in Fig. 3. The Eclipse plan was down-sampled for this comparison—and for the conversion from Eclipse→CERR—because its fluence resolution (2.5 × 2.5 mm) was much higher than what was used in CERR (1.0 × 1.0 cm).

TABLE I. Clinical planning objectives for the prostate, pancreas, and head-and-neck cases. Abbreviations: Fem. heads (femoral heads), N. liver (normal liver), S. bowel (small bowel), and D_p (prescription dose).

Prostate				Pancreas				Head and Neck			
Organ	Goal			Organ	Goal			Organ	Goal		
PTV	V_{Dp}	>=	95%	PTV	V_{Dp}	>=	95%	PTV	$D_{0.1\mathrm{cc}}$	<=	$1.1*D_p$
Rectum	V_{50}	<=	50%	N. liver	V_{15}	<=	900 cc	PTV_1	V_{Dp}	>=	95%
	V_{60}	<=	35%		D_{mean}	<=	15 Gy	PTV_2	V_{Dp}	>=	95%
	V_{65}	<=	25%	Kidneys	V_{10}	<=	10 cc	PTV_3	V_{Dp}	>=	95%
	V_{70}	<=	20%		D_{mean}	<=	10 Gy	Cochlea	D_{mean}	<=	45 Gy
	V_{75}	<=	15%	S. bowel	V_{20}	<=	10 cc	Cord	D_{max}	<=	50 Gy
Bladder	V_{65}	<=	50%		V_{30}	<=	9 cc	Chiasm	D_{max}	<=	55 Gy
	V_{70}	<=	35%	Stomach	V_{20}	<=	15 cc	Brain stem	D_{max}	<=	54 Gy
	V_{75}	<=	25%	Cord	D_{max}	<=	50 Gy	Mandible	D_{max}	<=	70 Gy
	V_{80}	<=	15%					Parotids	D_{mean}	<=	20 Gy
Fem. heads	V_{50}	<=	10%								
Penile bulb	D_{mean}	<=	50.0 Gy								

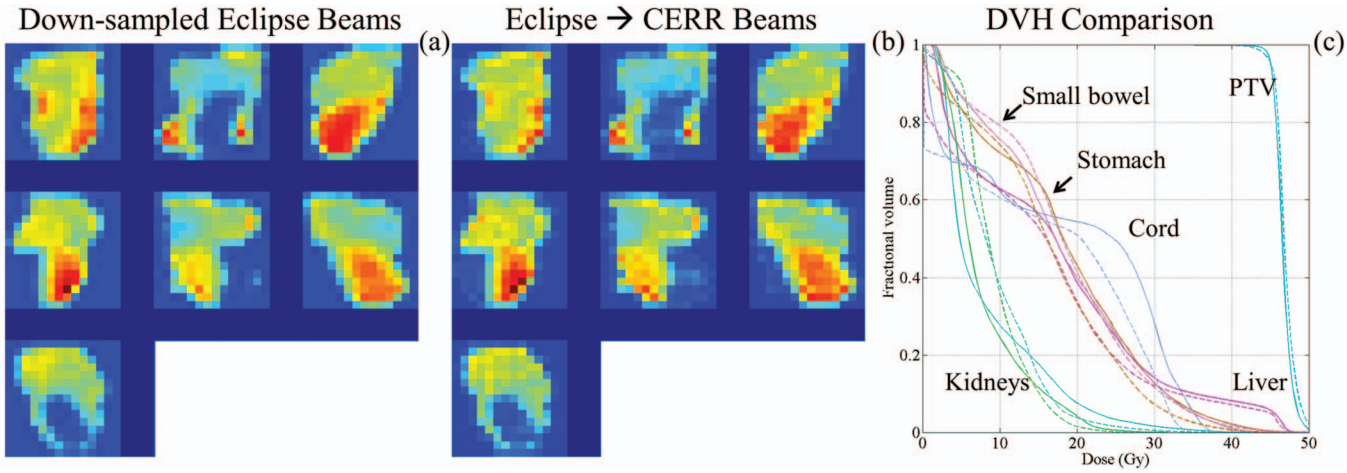

FIG. 3. Comparison between the Eclipse plan and the Eclipse → CERR plan for the pancreas case. (a) Down-sampled Eclipse beams. (b) Eclipse → CERR beams. (c) DVH comparison between Eclipse (solid lines) and Eclipse → CERR (dashed lines). The beams in (a) and (b) share a common color scheme.

II.F. Patient data

The proposed method was tested on four prostate cases, one pancreas case, and head-and-neck case. The retrospective use of de-identified patient data was approved by our institutional review board (IRB).

II.G. Plan deliverability assessment: 2D modulation index

The 1D modulation index (MI) was formulated by Webb as a proof-of-principle study to quantify beam modulation complexity, arguing that differences in influence between neighboring beamlets greater than the beam's standard deviation were more indicative of complexity/deliverability than the standard deviation itself.[31] Thus, MI is similar to the total variation in that it penalizes the total magnitude of intrabeamlet deviations, and is sensitive to both (1) high-frequency deviations, i.e., intrabeamlet deviations within relatively homogeneous regions, and (2) low-frequency deviations, i.e., coarse-scale deviations from piecewise-constant behavior. Later, the method was generalized to 2D (MI_{2D}) to quantify beam complexity of patient cases,[32] and was shown to correlate better with treatment deliverability than monitor units.[33] As such, ratios of MI_{2D} for E_{tot}^{sparse}-optimized plans to MI_{2D} for Eclipse→CERR plans and ratios of MI_{2D} for E_{tot}^{sparse}-optimized plans to MI_{2D} for clinical (Eclipse) plans were used in this work to compare the relative deliverability and efficiency of E_{tot}^{sparse}-optimized plans.

II.H. Planning process

The planning process of our method is relatively simple, as shown in the written steps below and Fig. 4.

1. Following contour exportation, DICOM-RT objects containing structure contours, CT scan, and the clinical dose map, are imported into CERR.
2. CERR's IMRTP function is then used to calculate the dose-deposition matrices for each structure with QIB.[30]
3. An initial set of weighting factors is chosen and an initial optimization is run.
4. Weighting factors are iteratively modulated by the planner until the plan sufficiently achieves planning goals.
5. The global dose map is calculated with the QIB algorithm.
6. The plan's MI_{2D} is calculated.

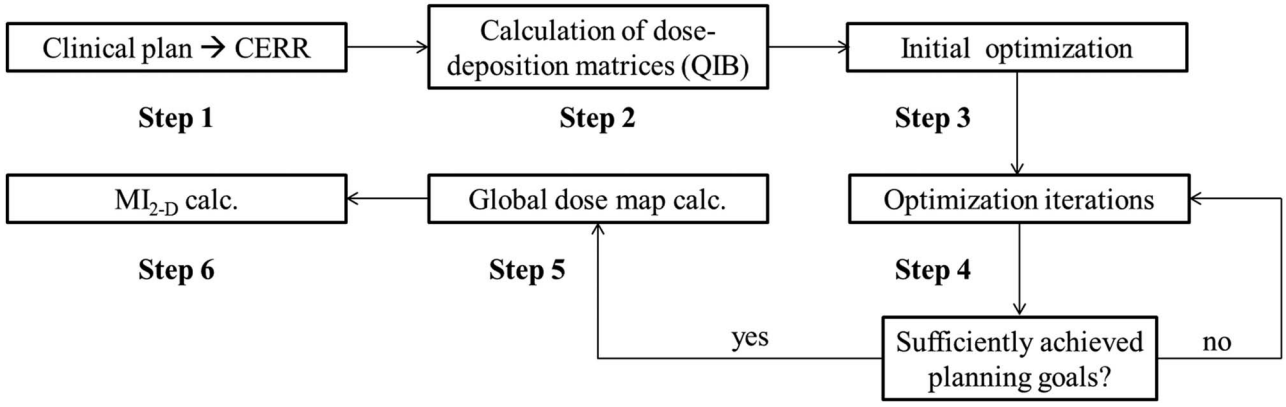

FIG. 4. Flowchart of our method's planning process.

II.I. Effect of fluence resolution on the optimization process

We defined our optimization problem in CERR using a coarse-grid fluence map (1.0 × 1.0 cm) for the purpose of improving planning efficiency, whereas a fine-grid fluence map (2.5 × 2.5 mm) was used in Eclipse. This difference in resolution represented a sixteen-fold decrease in the number of beamlets (degrees of freedom) available to our method for optimization, which physically limits dose-shaping ability.[34] Had the coarse fluence grid prevented our method from improving upon the clinical plans we would have refined its resolution as necessary. However, the improvements made by our method were substantial in their own right, and were made that much more pronounced with this limitation in mind.

II.J. Outline of investigations

We performed three main investigations for this work. First, several $E_{\text{tot}}^{\text{sparse}}$-optimized plans of each prostate case were compared against the Eclipse→CERR plan of the same case using DVH plots, characteristic image views, and the quantitative metrics introduced earlier in Subsection II.C. The purpose of these comparisons was to give us a better understanding of: (1) the practical differences between conventional and $E_{\text{tot}}^{\text{sparse}}$-based optimization; and (2) the spatial dose-shaping ability of $E_{\text{tot}}^{\text{sparse}}$. For the sake of brevity and presentation quality, the results of only one prostate case are presented in this manuscript, because the behavioral differences between $E_{\text{tot}}^{\text{sparse}}$-optimized plans and Eclipse→CERR plans changed negligibly from case-to-case.

Second, these comparisons were repeated for a pancreas case and a head-and-neck case to gauge the potential impact of our method on more challenging geometries. Third, two comparisons of plan deliverability were performed using MI_{2D} ratios. The first comparison was between fluence maps generated using our method and fluence maps generated from the Eclipse→CERR conversion to ensure a common optimization framework for deliverability comparison. The second comparison was between the fluence maps generated using our method and the clinical fluence maps generated with Eclipse to demonstrate the effect of fluence resolution on deliverability.

TABLE II. Summaries of $E_{\text{tot}}^{\text{sparse}}$-optimized plans with salient weighting factors for E_{PTV} and E_{TVR}.

Plan	E_{PTV} α_h	E_{TVR} γ
HC	∞ (constraint)	0
TV_1	0	3.5
TV_2	0	7
TV_{1c}	5	3.5
TV_{2c}	5	7

III. RESULTS

III.A. Prostate case

Each of the plans optimized with $E_{\text{tot}}^{\text{sparse}}$ for the prostate case was done according to the following, descending-order goals: (1) $E_{\text{tot}}^{\text{sparse}}$-optimized plans $V_{Dp} \geq$ Eclipse→CERR V_{Dp}, (2) rectum-sparing, and (3) PTV homogeneity and bladder-sparing. For the first plan, termed plan HC, we imposed a hard constraint on the PTV's dose distribution using the maximum dose found in the Eclipse→CERR plan as the upper limit and did not apply E_{TVR} regularization. This constraint is equivalent to letting α_h of the high-dose leg in E_{PTV} approach ∞, as shown in Table II. For the second plan, termed TV_1, the high-dose leg was removed, i.e., $\alpha_h = 0$, and an E_{TVR} with relatively low importance was assigned to the PTV. The third plan, TV_2, placed a higher importance on TV regularization than TV_1, the magnitude of which was controlled by weight γ, as shown in Table II. The third and fourth plans, TV_{1c} and TV_{2c}, were formed by adding high-dose legs with moderate importance, i.e., $\alpha_h = 5$, to TV_1 and TV_2, respectively. Therefore, TV_{1c} and TV_{2c} deterred high-dose regions, or hot spots, with two penalties—E_{TVR} and the high-dose leg of E_{PTV}—rather than just one, i.e., E_{TVR}, as TV_1 and TV_2 did. These formulations are summarized in Table II and the DVH results are shown in Fig. 5.

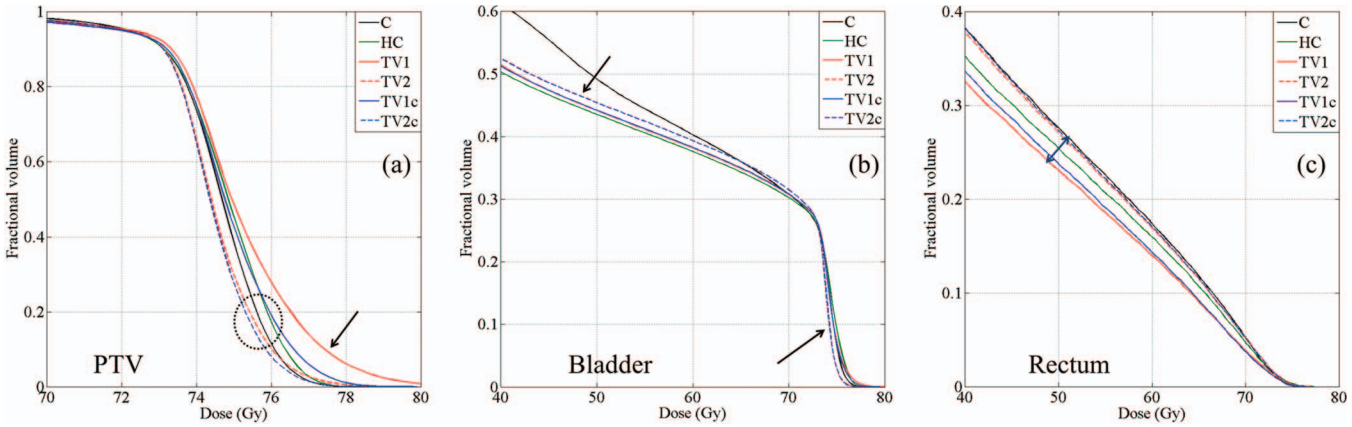

FIG. 5. (a)–(c) DVH plots for characteristic prostate case. Pictured are the clinical plan (C) and the five $E_{\text{tot}}^{\text{sparse}}$-optimized plans defined in Table II for the PTV, bladder, and rectum. Arrows and dotted circles indicate points of discussion in the text.

III.A.1. Controlling spatial dose distributions

The five plans described in Table II were designed to illustrate how nonspatial and spatial penalties can be used to influence $E_{\text{tot}}^{\text{sparse}}$-based optimization. To begin, we consider plan HC, which prevented all voxels from exceeding the maximum dose of the Eclipse→CERR plan. As a result, the high-dose regions that developed in plan HC were: (1) relatively low in magnitude; and (2) not very sensitive to plan geometry, as shown by black arrows in Figs. 6(b) and 7(d) where they developed in the PTV-bladder overlap. In contrast, plan TV_1, which lightly penalized spatial gradients but did not penalize overdose, encouraged a hot-spot distribution that was *more* sensitive to geometry but *less* sensitive to high-dose violations, as shown by lightly colored arrows in Figs. 6(c), 7(e), and 7(f). The hot spots that did develop were in nonsensitive locations but were also quite prominent, while dose falloff from the PTV's surface to the rectum was steeper than seen in all other plans, as shown by the lightly colored arrow in Fig. 5(c) and the dotted black arrows in Fig. 7, respectively. In this way, PTV homogeneity was sacrificed to maximize rectum-sparing in plan TV_1.

By contrast, plan TV_2 placed substantially more importance on PTV homogeneity by heavily penalizing E_{TVR}, which homogenized the PTV's spatial dose distribution through an anisotropic smoothing mechanism, as shown by comparing regions indicated by solid and dotted lightly colored arrows in Figs. 7(e) and 7(g) and Figs. 7(f) and 7(h), respectively. The smoothing effect also reduced OAR-sparing by penalizing dose deviations, and therefore dose falloff, from the PTV's surface. As a result, rectum-sparing was slightly worse for plan TV_2 relative to plan TV_1, as shown by the lightly colored arrow in Fig. 5(c), while PTV homogeneity was substantially increased, as shown in Fig. 5(a). Black arrows in Fig. 5(b) show that bladder-sparing for plan TV_2 was: (1) slightly improved for high doses due to improved PTV homogenization in the large PTV-bladder overlap; and (2) slightly decreased for lower doses due to the aforementioned smoothing mechanism's effect on dose falloff.

Plan TV_{1c} offered a more balanced approach where PTV inhomogeneity and high-dose regions were both *moderately* penalized. In fact, by comparing plan TV_{1c} with plan TV_1 and TV_2, intuition regarding the practical effects of E_{TVR} and the high-dose leg of E_{PTV} can be gained. Examination of the

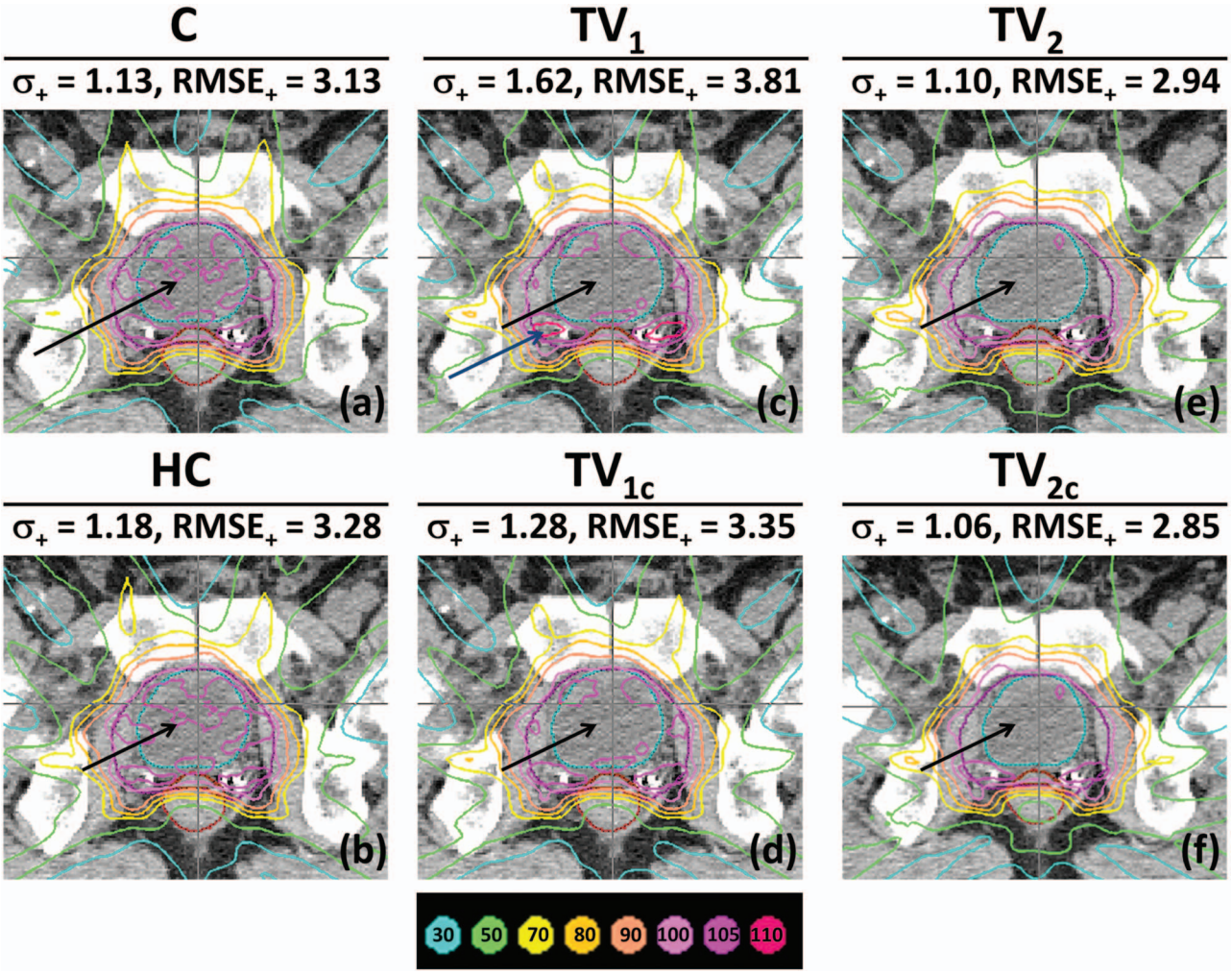

FIG. 6. Transaxial slice near the pubic symphasis. Pictured are the clinical plan (C) and the five sparse plans from Table II. Arrows refer to discussion points in the text. Isodose lines, expressed as a percentage of PTV prescription dose, are shown above in the legend.

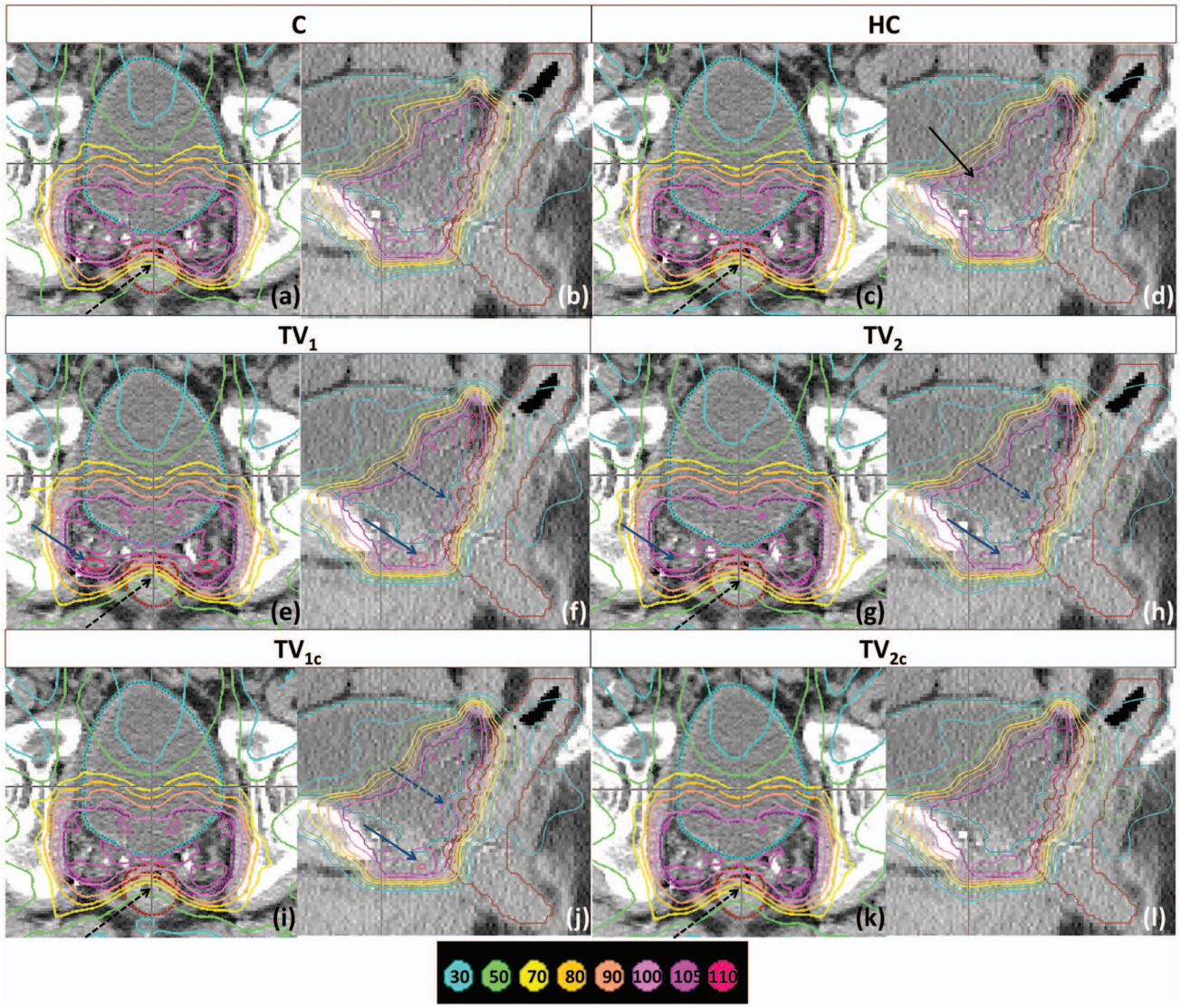

FIG. 7. Transaxial slice near seminal vesicles and selected sagittal slice near sagittal midline of PTV. Pictured are the clinical plan (C) and the five sparse plans from Table II. Arrows refer to discussions in the text. Isodose lines, expressed as a percentage of PTV prescription dose, are shown above in the legend.

110% isodose and 105% isodose regions indicated by solid and dotted lightly colored arrows, respectively, in Figs. 7(f) and 7(j), shows that E_{TVR} homogenized all spatial dose variation, regardless of dose value, because both of these regions were reduced in size and magnitude. In contrast, the high-dose leg of E_{PTV} only reduced dose delivered to voxels above a certain threshold, because the 110% isodose region was markedly reduced while the 105% region was not. Thus, the high-dose leg of E_{PTV} may be used to target prominent hot spots, like the 110% isodose regions seen in plan TV_1, while E_{TVR} may be used to improve overall PTV homogeneity and encourage steeper dose falloff using geometry-sensitive outlier allocation. The resulting net effects for this prostate case, as shown in Figs. 5(a), 5(c), and 7(e)–7(h), and 6(i) and 6(j), were that plan TV_{1c} was closer to plan TV_2 than plan TV_1 in terms of PTV homogeneity σ_+ and PTV overdose RMSE_+, as shown in Fig. 5, but closer to plan TV_1 in terms of rectum-sparing, as shown in Fig. 5(c). Plan TV_{2c} was nearly identical to plan TV_2 because the heavily penalized E_{TVR} was the main driver of dose falloff and PTV homogeneity in both plans.

III.B. Complicated geometries: The pancreas and head and neck

The prostate case of Sec. III.A was used as the basis for illustrating our method, because prostate plan geometries are relatively simple and can help gain a basic intuition about treatment planning behavior and the interaction of different energies, e.g., E_{PTV} and E_{TVR}, in a relatively straightforward process. Upon establishing insight from the prostate cases, we further tested the generality of the proposed method on complex pancreas and head-and-neck cases where more translational significance was expected.

III.B.1. Pancreas case

For the selected pancreas case, we primarily sought to improve PTV homogeneity and cord-sparing while

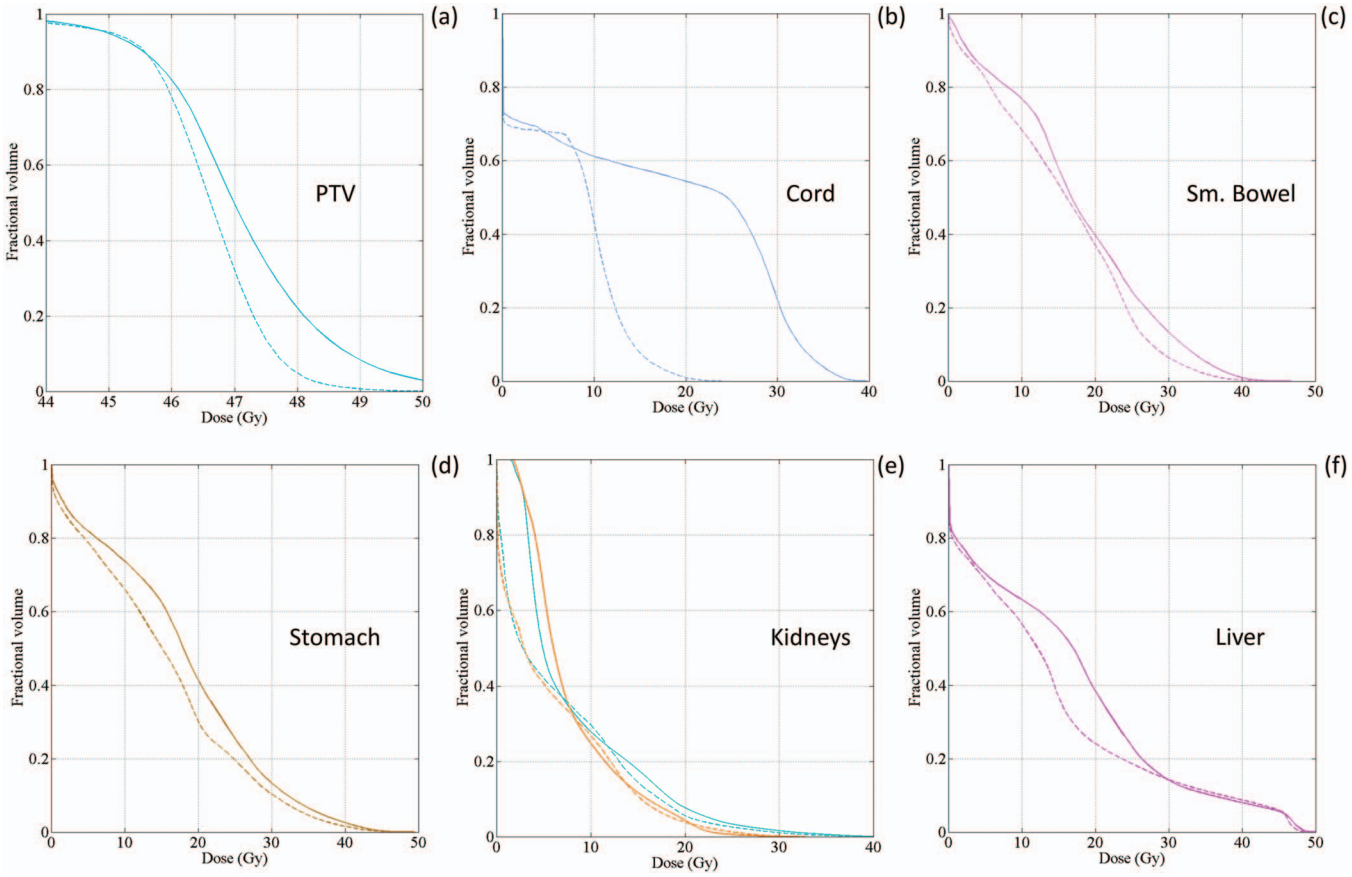

FIG. 8. DVH results for the pancreas case: (a) PTV, (b) spinal cord, (c) small bowel, (d) stomach, (e) kidneys, and (f) liver. Clinical plan plots have solid lines and our plan has dashed lines.

maintaining the DVHs of other organs. To this end, we applied a large weighting factor to E_{TVR}, $\gamma = 9.5$, and set the E_{cord} elbow to 10 Gy, significantly below the maximum seen in the Eclipse→CERR plan. We did not penalize high-dose deviations in the PTV, i.e., α_h of the high-dose leg was set to zero. DVH results from the plan optimized using our generalized-L_1 method and the Eclipse→CERR plan are shown in Fig. 8.

The DVH results indicated substantial improvements to both PTV homogeneity and cord-sparing while maintaining the dose-sparing of other organs. The maximum dose delivered to the PTV in our plan was approximately 2 Gy less than that delivered in the Eclipse→CERR plan, and gains in PTV homogeneity, $\sigma_+ = 0.79$ vs 1.18 Gy, and PTV overdose, $\mathrm{RMSE}_+ = 1.97$ vs 2.38 Gy, were clearly reflected in the DVH plot in Fig. 8(a). Furthermore, the maximum dose delivered to the cord in our plan was over 15 Gy less than that delivered to the Eclipse→CERR plan, as shown in Fig. 8(b).

The gains in overall plan quality made by our plan over the Eclipse→CERR plan become more pronounced when comparing transaxial, sagittal, and coronal views of each in Fig. 9. First, the spatial dose distribution of our plan is more uniform and devoid of the large hot spots seen in the Eclipse→CERR plan, as indicated by the black arrows in Figs. 9(b), 9(c), 9(e), and 9(f). Second, the large improvement in cord-sparing in our plan is also clearly seen, as is a larger low-dose region around the cord, which is labeled in Figs. 9(a) and 9(d). Third, our plan is more pleasing visually and generally has better conformity of high doses to the PTV surface where OARs are relatively adjacent, as shown by comparing regions indicated by the dotted ellipses in Figs. 9(b) and 9(e) and Figs. 9(c) and 9(f).

III.B.2. Head-and-neck case

Our goals for the head-and-neck case were to improve PTV homogeneity and sparing of the parotid glands, which have a major impact on quality of life. We applied large weighting factors to both E_{OAR}, for the parotid glands, and E_{TVR}, and set the E_{OAR} elbow of the parotid glands to 10 Gy. We used heterogeneous weights for E_{TVR} and the high-dose leg of E_{PTV}: (1) $\gamma = 15$ and $\alpha_h = 13.5$ for PTV_1, (2) $\gamma = 4$ and $\alpha_h = 25.5$ for PTV_2, and (3) $\gamma = 3$ and $\alpha_h = 20.2$ for PTV_3. DVH results from this plan and the Eclipse→CERR plan are shown in Fig. 10.

The DVH results in Fig. 10 clearly indicate that substantial improvements to both PTV homogeneity and parotid- and cord-sparing were made while maintaining or improving the dose-sparing of other organs. The maximum dose delivered to PTV_1, PTV_2, and PTV_3 in our plan was less than that delivered to the Eclipse → CERR plan by 2.35, 2.35, and 3.52 Gy, respectively. The quantitative PTV dose metrics followed a

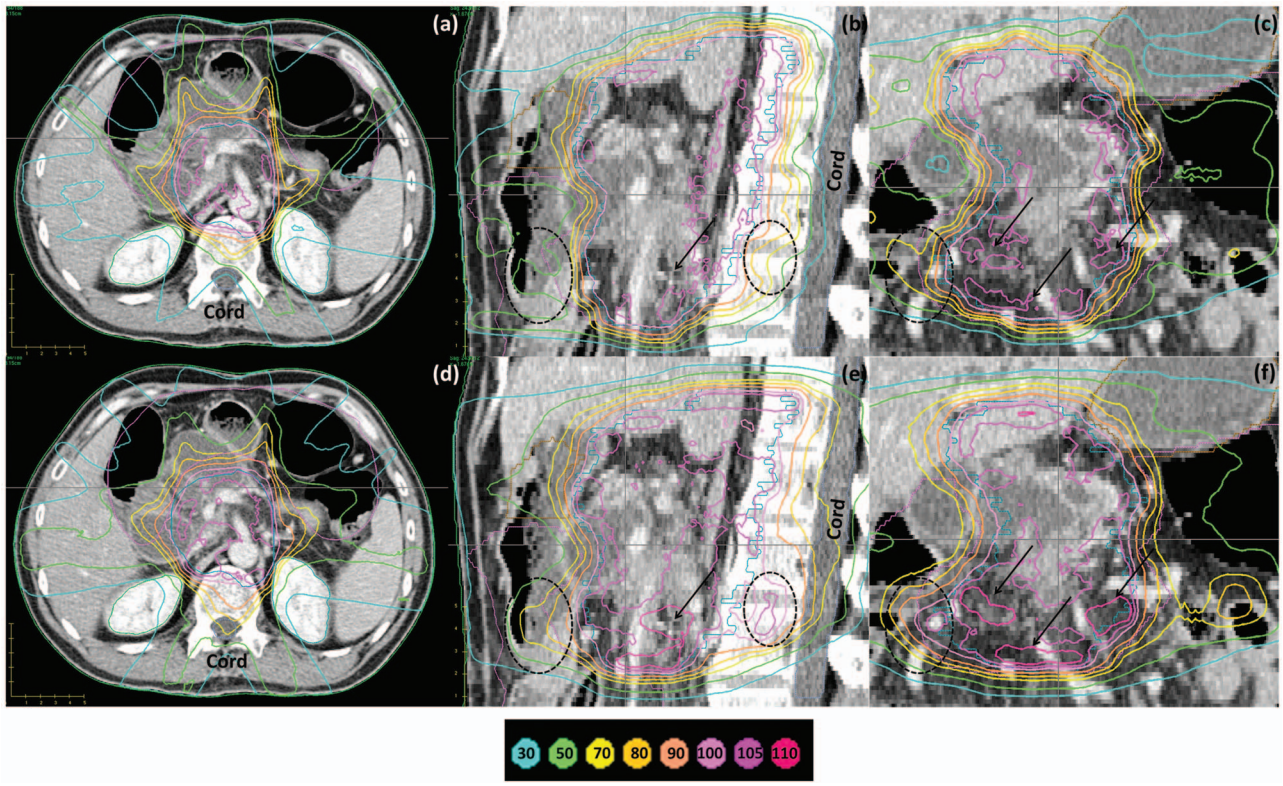

FIG. 9. Selected transaxial, sagittal, and coronal views of the pancreas case. Our plan is shown in the top row and the clinical plan is shown in the bottom row. Arrows and dotted ellipses indicate comparisons made in the text. Isodose lines, expressed as a percentage of PTV prescription dose, are shown above in the legend.

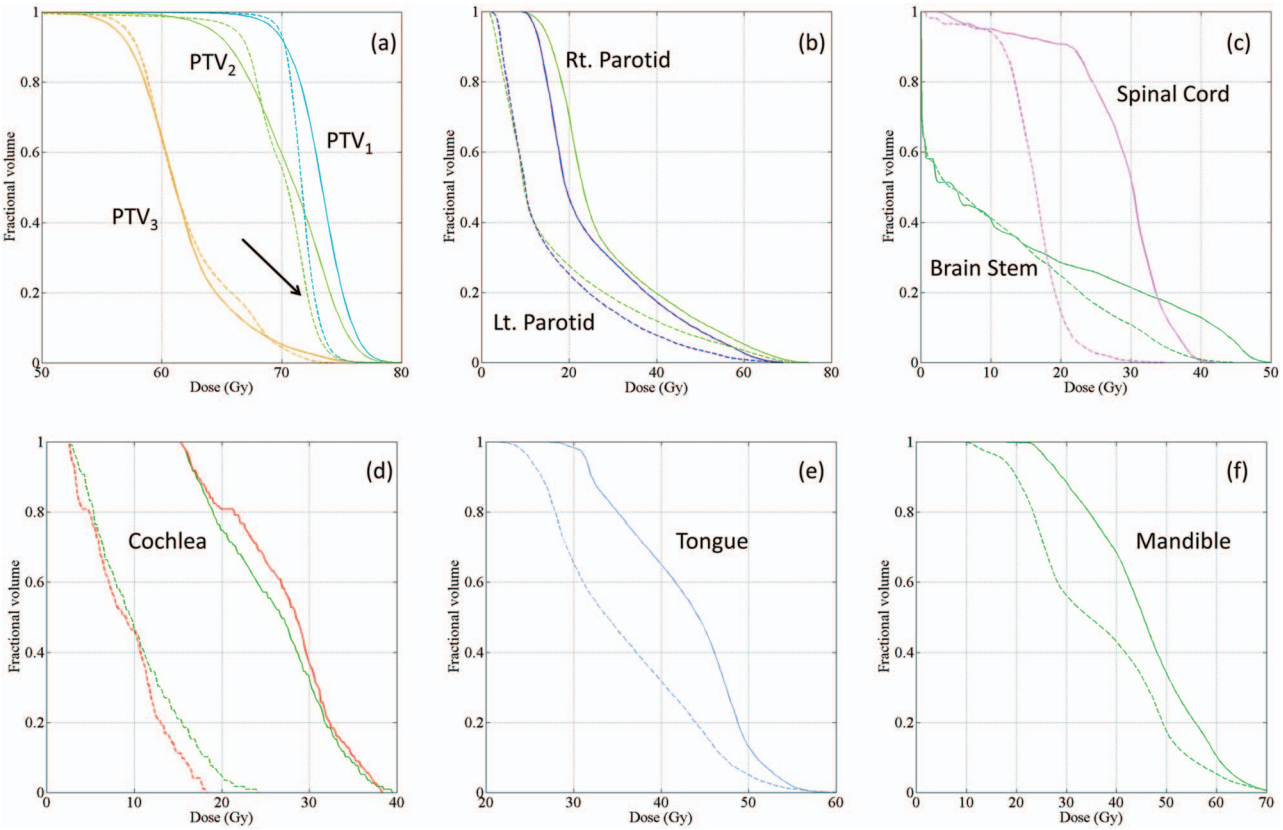

FIG. 10. DVH results for the head-and-neck case: (a) PTV, (b) parotid glands, (c) spinal cord and brain stem, (d) cochlea, (e) tongue, and (f) mandible. Eclipse→CERR plan plots have solid lines and our plan has dashed lines.

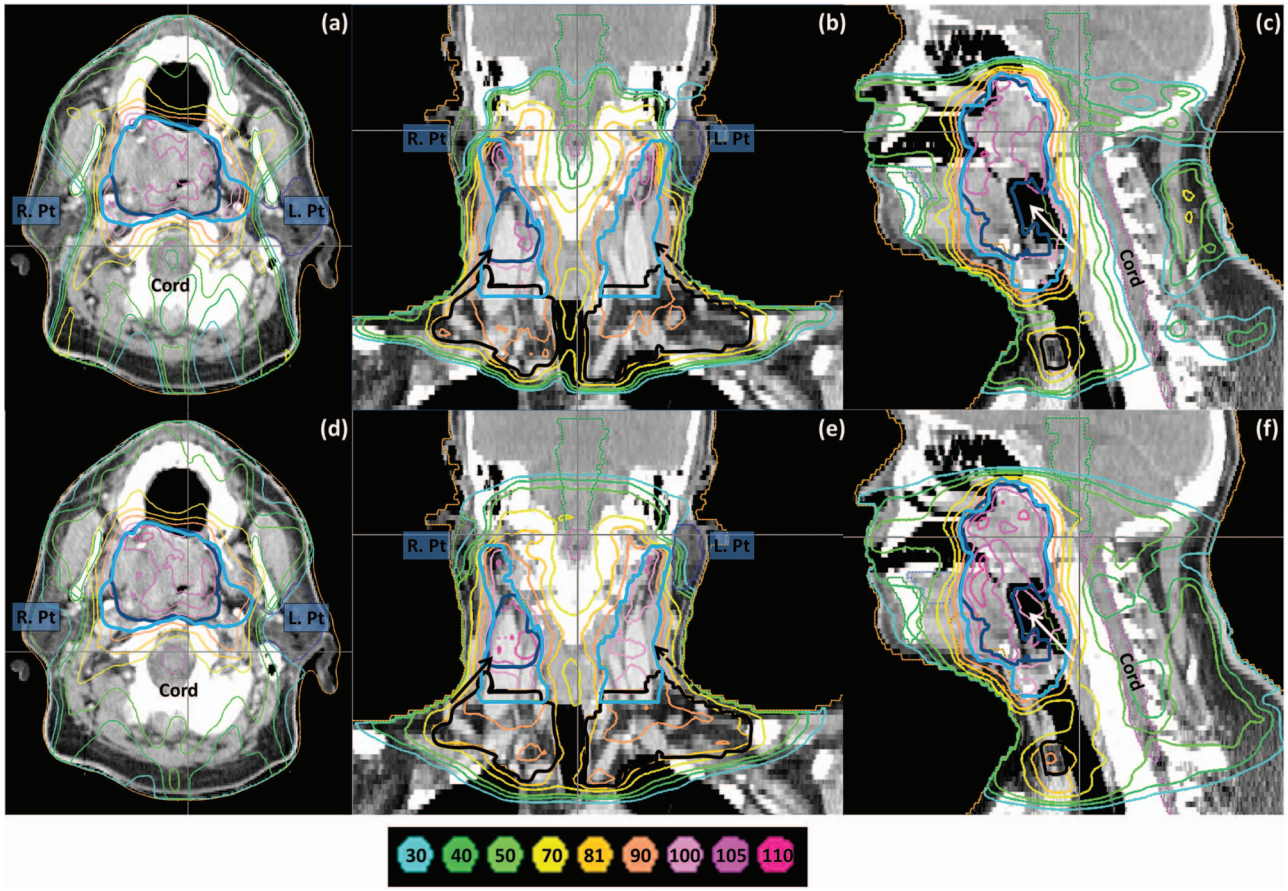

FIG. 11. Selected transaxial, sagittal, and coronal views of the head-and-neck case. Our plan is shown in the top row and the Eclipse→CERR plan is shown in the bottom row. Arrows and textual labels indicate comparisons made in the text. Isodose lines, as a percentage of the PTV_1 prescription dose, are specified above. The 81% line corresponds to the prescription dose of PTV_3 and the 90% line corresponds to the prescription dose of PTV_2. The PTV_1 contour is darkly colored, the PTV_2 contour is lightly colored, and the PTV_3 contour is black.

similar trend: (1) PTV homogeneity, $\sigma_+ = 1.18$ vs 1.77 Gy, and PTV overdose, $RMSE_+ = 2.86$ vs 3.87 Gy, for PTV_1; (2) PTV homogeneity, $\sigma_+ = 2.27$ vs 3.30 Gy, and PTV overdose, $RMSE_+ = 7.91$ vs 8.13 Gy, for PTV_2; and (3) PTV homogeneity, $\sigma_+ = 3.85$ vs 3.78 Gy, and PTV overdose, $RMSE_+ = 7.32$ vs 6.59, for PTV_3. Most importantly, our method could have markedly improved the patient's salivary function, a major quality of life issue, by substantially reducing the mean dose delivered to the left parotid gland (15.55 vs 25.67 Gy) and right parotid gland (16.93 vs 28.73 Gy).[35,36] Large improvements in cochlea-sparing were also evident, as shown in Fig. 10(d) while tongue- and mandible-sparing were maintained.

As with the pancreas case, gains in overall plan quality can be more appreciated when comparing transaxial, sagittal, and coronal views of the Eclipse→CERR plan and our plan. First, improved homogeneity in the PTVs is clearly visualized in our plan in Fig. 11, which exhibits: (1) spatial dose distributions that are much more uniform, as shown by black arrows in Figs. 11(b) and 11(e) for PTV_1 and dotted black arrows in Figs. 11(b) and 11(e) for PTV_2; and (2) a substantial reduction in hot-spot size and magnitude in PTV_1 and PTV_2, as shown by the same arrows. Second, the large improvement in parotid-sparing in our plan, which is clearly seen in the DVH results, may be visualized by comparing Figs. 11(a) and 11(d) and Figs. 11(b) and 11(e). Here, our method was able to carve out a relatively large volume of low dose in and around the parotid glands to substantially increase their sparing. Third, our plan is more pleasing visually and generally has sharper dose falloff away from PTV surfaces where OARs are relatively adjacent, such as near the parotid glands in Figs. 11(a) and 11(b) and the spinal cord in Figs. 11(a) and 11(c). This improved OAR-sparing comes at the small expense of losing target conformity in regions away from OAR overlap and relative adjacency, as indicated by white arrows in Figs. 11(c) and 11(f).

III.C. Plan deliverability

As discussed in Sec. II, ratios of MI_{2D} values of the E_{tot}^{sparse}-optimized plans to their respective Eclipse→CERR plans were used to quantify the relative complexity/deliverability of the E_{tot}^{sparse}-optimized plans. Importantly, deliverability differences between E_{tot}^{sparse}-optimized plans and Eclipse→CERR plans were unlikely to be clinically relevant, because the MI_{2D} ratios of Fig. 12 did not differ much from unity with a range of 0.985–1.075 and a standard deviation 0.03. For reference, the MI_{2D} ratio of the Eclipse→CERR pancreas plan to the

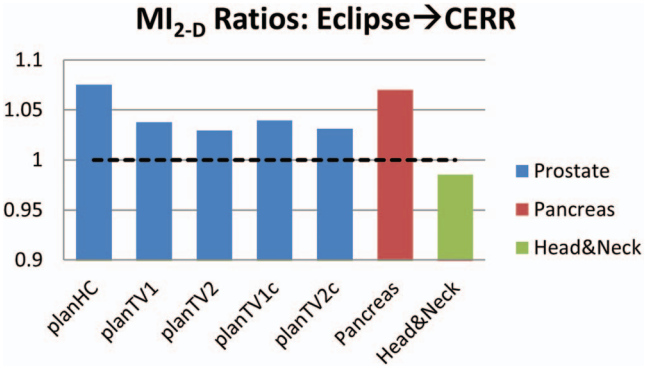

FIG. 12. MI_{2D} ratios between E_{tot}^{sparse}-optimized plans and Eclipse→CERR plans of the prostate, pancreas, and head-and-neck cases. MI_{2D} ratio range = 0.985–1.075. MI_{2D} standard deviation = 0.03.

clinical pancreas plan, which was shown to have good behavioral agreement in Fig. 3, was 1.03.

Inspection of the prostate case results in Fig. 12 shows that plan TV_2 and plan TV_{2c} possessed the smallest MI_{2D} ratios of all plans. Therefore, PTV homogeneity may have been correlated with plan deliverability for plans optimized with our method. However, E_{TVR} could not have been the only contributor to MI_{2D} ratio variation since the pancreas and head-and-neck cases showed similar gains in PTV homogeneity, relative to the prostate case, but deviated from the MI_{2D} ratio of prostate plan TV_2 in opposite directions. The likely reason for this apparent discrepancy was the amount of dose-carving necessary to achieve the desired planning goals. Specifically, dose-carving in and around the parotid glands, two relatively small structures near the exterior of the body, required a smaller amount of isodose surface displacement than dose-carving in and around the spinal cord, an extremely long structure with a relatively central location.

By using the Eclipse plans, rather than Eclipse→CERR plans, to calculate MI_{2D} ratios we observe that our method is far more deliverable than the actual clinical plans for the pancreas case (Eclipse MI_{2D} ratio = 0.1925) and the head-and-neck case (Eclipse MI_{2D} ratio = 0.1548). E_{tot}^{sparse}-optimized plans. A visual comparison depicting the differences in fluence resolution between the clinical (Eclipse) and E_{tot}^{sparse}-optimized plans is shown in Fig. 13, which compares the beams optimized with Eclipse to the beams optimized with E_{tot}^{sparse} for the pancreas case. The E_{tot}^{sparse}-optimized plan appears pixelated because it possesses sixteen-fold fewer beamlets than the clinical plan, but it is more deliverable as a result.

IV. DISCUSSION

In this work, we proposed a novel approach to shape the spatial dose distributions of the PTV and OARs by applying coupled L_1 norms and linear constraints to the dose values in PTVs and OARs and to dose gradients in PTVs. The L_1 norm permits outliers—a sparse number of large deviations from prescription—to arise in solution and improve tradeoff between OAR-sparing and PTV homogeneity.

We performed three major comparisons to validate our method. First, we used our method to generate five Pareto plans for four prostate cases, which differed in how dose-prescription and gradient penalties affected high-dose regions in the PTV, and compared them to the corresponding clinical plan using dose-volume histograms (DVHs), quantitative descriptors, and visual analyses of dose maps. The results showed that the total variation energy (E_{TVR}) homogenized the PTV's spatial dose distribution through an anisotropic smoothing mechanism that led to OAR-sparing when lightly penalized and high PTV homogeneity when heavily penalized. The high-dose leg of E_{PTV} was found to compliment E_{TVR} well by reducing the magnitude and geometric selectivity of hot spots. Properly balancing these two penalties yielded more controlled and accurate dose-shaping than possible with the conventional objective function. Second, we tested the generalizability of our method to more challenging disease sites by repeating the same analyses for a pancreas case and a head-and-neck case, and found that our method produced plans with substantially improved PTV homogeneity and sparing of several OARs, notably the spinal cord in the pancreas case and parotid glands in the head-and-neck case. Third, we showed that our plans could make the aforementioned improvements to dose-shaping while being markedly

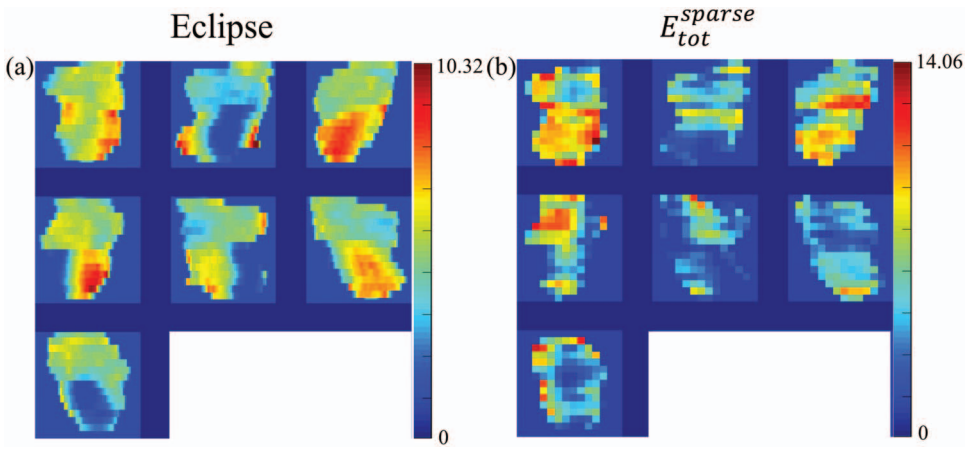

FIG. 13. Visual comparison of the Eclipse beams with E_{tot}^{sparse} beams for the pancreas case. Fluence intensity color bars are shown.

more deliverable than Eclipse plans. As a result of these comparisons, we have gained great confidence that our method can lead to improved planning performance under a wide range of geometric and dose-planning conditions.

In the remainder of this section, we will discuss specifics about: (A) the relative performance of our method in optimizing simple cases, such as the prostate, vs complicated cases, such as the pancreas and head and neck; (B) the relationship between hot-spot distribution and plan geometry; and (C) other considerations and future work.

IV.A. Simple vs complicated cases

Treatment planning for prostate cancer involves relatively few tradeoffs and is substantially less complicated to plan than other tumor sites, such as head-and-neck cancer.[37] Thus, making substantial improvements over conventional plans using our generalized-L_1 method may be difficult, because there are fewer tradeoffs to improve and fewer gains to be had. Conversely, sites with more complicated geometries and added tradeoffs, such as the pancreas and head-and-neck, may benefit more from our method, because there are more opportunities to strategically allocate outliers to regions where they either do less harm or more good for multiple sets of structures. So, the more structures there are, the more effective outlier allocation and our method become. While more studies are needed to ultimately validate this assertion, we believe that our plans for the pancreas and head-and-neck cases provided sufficient gains in DVH-based sparing and dose-shaping to give credence to such an argument.

IV.B. Hot spots and plan geometry

In prostate cases, hot spots that developed in $E_{\text{tot}}^{\text{sparse}}$-optimized plans were of higher magnitude and size than those seen in $E_{\text{tot}}^{\text{sparse}}$-optimized plans of the pancreas and head-and-neck. In fact, the pancreas PTV of the $E_{\text{tot}}^{\text{sparse}}$-opimized plan was nearly uniform in dose, and hot spots that developed in the head-and-neck PTVs, especially when compared with the clinical plan, were minimal. The latter is especially surprising since the head-and-neck case possessed more structures, i.e., tradeoffs, than the prostate cases and the pancreas case, as well as a much more challenging geometry that included adjacent PTVs and complicated structure shapes. By contrast, the prostate geometry was relatively simple. In fact, prostate PTVs are generally spherical or ellipsoidal in shape and the bladder and rectum are also regular, particularly when compared with structures at other sites, such as the small bowel in the pancreas case and the mandible and parotid glands in the head-and-neck case.

The main difference, then, is the amount of overlap between the OARs and the PTV, particularly the fraction of the PTV that is overlapped by OARs. The logic behind this theory can be easily demonstrated by analyzing the physical process of dose-shaping. For a given IMRT case, there are a certain number of radiation beams that are used to deliver the prescription dose to the target while sparing the OARs. As the PTV-OAR overlap fraction increases, the concentration of beams throughout the PTV becomes less uniform as more beams are concentrated into nonoverlapping regions to spare PTV-OAR overlaps of high doses. In other words, as PTV-OAR overlap fraction increase, hot spots forming in nonoverlapping regions increase in size and magnitude.

Thus, for prostate PTVs, which have relatively large overlap with OARs, beams are *naturally* concentrated in a relatively small nonoverlapping region and hot spots are prominent. Conversely, for the pancreas PTV, which is large and negligibly overlapped by OARs, beam delivery is *not constrained* and hot spots are minimal in size and magnitude. The hot spots of the head-and-neck PTVs were slightly more prominent than those of the pancreas, because additional dose-shaping is needed to deliver inhomogeneous prescription to its three PTVs and targeted dose falloff from their surfaces.

IV.C. Other considerations and future work

Since this was a pilot work to introduce a new optimization objective formulation, there are a few aspects of this study that warrant further investigation and discussion. In this study, plans optimized using $E_{\text{tot}}^{\text{sparse}}$that had not been leaf-sequenced were compared with clinical plans that had been leaf-sequenced. The introduction of a leaf-sequencing module would sacrifice the dose distribution, but only slightly, since it has been shown that leaf-sequencing can make delivery acceptably efficient without appreciably degrading plan quality.[32,38] This is especially true for the cases under examination here, as the optimized fluence maps were reasonably smooth (Fig. 3). On the other hand, fluence resolution has a much larger effect on plan quality, because coarse-grid fluence maps are physically limited in delivering steep dose drop-offs from the PTV.[34] Therefore, since the fluence grid used by our method (1.0×1.0 cm) was much coarser than that used by Eclipse (2.5×2.5 mm), the improvements made by our method relative to conventional optimization would have been even more pronounced had they been optimized under equivalent conditions.

A more important issue regarding comparison fairness concerns the potential suboptimality of the clinical plans used for benchmarking. Despite the planner's best effort, clinical plans are obtained in an iterative setting, so optimality is not guaranteed. On the other hand, at the core of the Eclipse system is a weighted quadratic cost, which is convex. Therefore, given the final objective parameters, this optimization problem is expected to be solved well by the commercial inverse treatment-planning engine. From this perspective, the clinical plans can be considered as the optimal solution to a specific weighted quadratic objective, while consistency in clinical parameter selection remains an issue that is outside the scope of this study. However, since plans optimized with $E_{\text{tot}}^{\text{sparse}}$ possessed better PTV coverage, PTV homogeneity, OAR-sparing, and deliverability than their clinical counterparts, it can be concluded that $E_{\text{tot}}^{\text{sparse}}$: (1) explores a different Pareto surface than conventional optimization methods; and (2) enables more efficient tradeoff between planning goals than conventional optimization methods.

Relatedly, while a more rigorous determination of E_{tot}^{sparse} alpha/beta values with Pareto navigation may be possible, the manual determination of their values in this work was made on a site-specific basis using dose-volume goals as a guide. So, while further improvement is possible, substantial gains are not anticipated.

On the application end, we will plan additional pancreas and head-and-neck cases to further validate the performance of our approach and apply our strategy to a wider range of anatomical sites. Upon further validation, we will develop a treatment planning platform with more clinically oriented resolution and dose calculation to achieve the full dose-shaping ability of our method.

V. CONCLUSIONS

We have investigated and developed a sparse objective function (E_{tot}^{sparse}) to: (1) encourage better tradeoff between competing objectives; (2) shift hot spots to insensitive regions; and (3) provide extra flexibility in balancing PTV homogeneity, coverage, with OAR-sparing. We compared the results of our method against the clinical plans of four prostate cases, one pancreas case, and one head-and-neck case, and showed substantial improvements to overall plan quality for complicated disease sites, i.e., the pancreas and head and neck, despite using a much coarser fluence grid for optimization than the clinical system. Finally, E_{tot}^{sparse}. is linear program, so it converges rapidly to the global solution and is easily implementable on various software platforms. Therefore, it may be quickly translated to the clinic where we expect a substantial impact on challenging cases.

ACKNOWLEDGMENTS

This work is supported in part by an AACR career development award and a TRDRP exploratory grant. The authors thank Dr. J. DeMarco and the physics group at UCLA for their assistance with data collection and Weber Shao for technical help. They are grateful to Dr. D. Low, Dr. P. Kupelian, and Dr. M. Steinberg for departmental support.

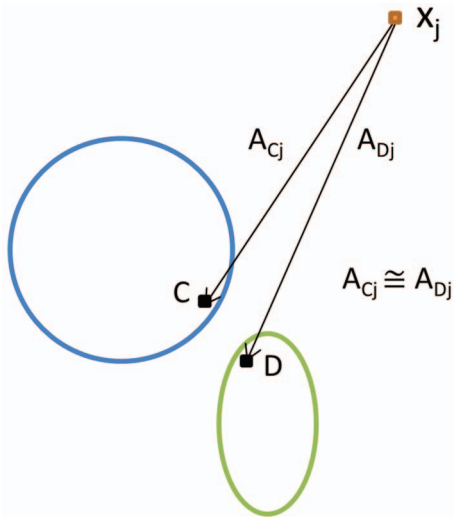

FIG. 14. Coupling bixel. Voxels C and D are coupled by the *j*th column of *A*, because the dose bixel x_j deposits to voxel *C* is nearly the same as the dose it deposits to voxel *D*. Each set of relatively adjacent points/regions is likely be coupled by many such bixels.

APPENDIX: RELATIVE ADJACENCY

Given a structured dose deposition matrix (A)[39] that maps influence bixel values (x) to the dose values (d), i.e., $d = Ax$, tradeoff naturally exists between adjacent regions of dramatically different dose goals. This tradeoff is carried out by the bixels, which couple adjacent regions together by the corresponding column of A (Fig. 14).

[a]Author to whom correspondence should be addressed. Electronic mail: druan@mednet.ucla.edu

[1]D. M. Shepard *et al.*, "Optimizing the delivery of radiation therapy to cancer patients," SIAM Rev. **41**(4), 721–744 (1999).
[2]J. M. Martin, J. Frantzis, T. Eade, and P. Chung, "Clinician's guide to prostate IMRT plan assessment and optimization," J. Med. Imag. Radiat. On. **54**(6), 569–575 (2010).
[3]L. B. Marks, "The impact of organ structure on radiation response," Int. J. Radiat. Oncol., Biol., Phys. **34**(5), 1165–1171 (1996).
[4]S. Benati and R. Rizzi, "A mixed integer linear programming formulation of the optimal mean/Value-at-Risk portfolio problem," Eur. J. Oper. Res. **176**(1), 423–434 (2005).
[5]Q. Wu and R. Mohan, "Multiple local minima in IMRT optimization based on dose-volume criteria," Med. Phys. **29**(7), 1514–1527 (2002).
[6]L. Xing *et al.*, "Optimization of importance factors in inverse planning," Phys. Med. Biol. **44**(10), 2525–2536 (1999).
[7]M. Alber *et al.*, "On the degeneracy of the IMRT optimization problem," Med. Phys. **29**(11), 2584–2589 (2002).
[8]K. L. Moore *et al.*, "Experience-based quality control of clinical intensity-modulated radiotherapy planning," Int. J. Radiat. Oncol., Biol., Phys. **81**(2), 545–551 (2011).
[9]G. A. Ezzell *et al.*, "Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT subcommittee of the AAPM radiation therapy committee," Med. Phys. **30**(8), 2089–2115 (2003).
[10]C. Cotrutz and L. Xing, "IMRT dose shaping with regionally variable penalty scheme," Med. Phys. **30**(4), 544–551 (2003).
[11]Y. Yang and L. Xing, "Inverse treatment planning with adaptively evolving voxel-dependent penalty scheme," Med. Phys. **31**(10), 2839–2844 (2004).
[12]C. Wu *et al.*, "Treatment plan modification using voxel-based weighting factors/dose prescription," Phys. Med. Biol. **48**(15), 2479–2491 (2003).
[13]Z. Shou *et al.*, "Quantitation of the a priori dosimetric capabilities of spatial points in inverse planning and its significant implication in defining IMRT solution space," Phys. Med. Biol. **50**(7), 1469–1482 (2005).
[14]N. Dogan *et al.*, "Improvement of treatment plans developed with intensity-modulated radiation therapy for concave-shaped head and neck tumors," Radiology **223**(1), 57–64 (2002).
[15]Q. W. Wu and R. Mohan, "Algorithms and functionality of an intensity modulated radiotherapy optimization system," Med. Phys. **27**(4), 701–711 (2000).
[16]E. Thomas *et al.*, "Benefit of using biologic parameters (EUD and NTCP) in IMRT optimization for treatment of intrahepatic tumors," Int. J. Radiat. Oncol. Biol. Phys. **62**(2), 571–578 (2005).
[17]B. B. Wu *et al.*, "Patient geometry-driven information retrieval for IMRT treatment plan quality control," Med. Phys. **36**(12), 5497–5505 (2009).
[18]W. De Neve, Y. Wu, and G. Ezzell, *Image-Guided IMRT* (Springer, Heidelberg, Berlin, 2006), p. 9.
[19]O. L. Mangasarian and D. R. Musicant, "Robust linear and support vector regression," IEEE Trans. Pattern Anal. Mach. Intell. **22**(9), 950–955 (2000).
[20]E. A. Kiountouzis, "Linear programming techniques in regression analysis," Appl. Stat. **22**, 69–73 (1973).
[21]R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressed sensing," in *International Conference on Acoustics, Speech, Signal Processing* (IEEE, Las Vegas, NV, 2008).

[22]E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l(1) minimization," J. Fourier Anal. Appl. **14**(5–6), 877–905 (2008).
[23]D. Wipf and S. Nagarajan, "Iterative reweighted l(1) and l(1) methods for finding sparse solutions," IEEE J. Sel. Top. Signal Process. **4**(2), 317–329 (2010).
[24]H. E. Romeijn *et al.*, "A new linear programming approach to radiation therapy treatment planning problems," Oper. Res. **54**(2), 201–216 (2006).
[25]H. E. Romeijn *et al.*, "A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning," Phys. Med. Biol. **48**(21), 3521–3542 (2003).
[26]Q. J. Wu *et al.*, "On-line re-optimization of prostate IMRT plans for adaptive radiation therapy," Phys. Med. Biol. **53**(3), 673–691 (2008).
[27]D. Craft *et al.*, "An approach for practical multiobjective IMRT treatment planning," Int. J. Radiat. Oncol., Biol., Phys. **69**(5), 1600–1607 (2007).
[28]D. M. Aleman *et al.*, "Interior point algorithms: Guaranteed optimality for fluence map optimization in IMRT," Phys. Med. Biol. **55**(18), 5467–5482 (2010).
[29]B. Murray *et al.*, "Experimental validation of the eclipse AAA algorithm," Med. Phys. **33**(7), 2661–2661 (2006).
[30]J. Deasy *et al.*, "A collaboratory for radiation therapy treatment planning optimization research," Ann. Oper. Res. **148**(1), 55–63 (2006).
[31]S. Webb, "Use of a quantitative index of beam modulation to characterize dose conformality: Illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal unmodulated radiotherapy," Phys. Med. Biol. **48**(14), 2051–2062 (2003).
[32]T. Kim *et al.*, "Inverse planning for IMRT with nonuniform beam profiles using total-variation regularization (TVR)," Med. Phys. **38**(1), 57–66 (2011).
[33]C. K. McGarry *et al.*, "Assessing software upgrades, plan properties and patient geometry using intensity modulated radiation therapy (IMRT) complexity metrics," Med. Phys. **38**(4), 2027–2034 (2011).
[34]G. W. Zhang *et al.*, "Effect of beamlet step-size on IMRT plan quality," Med. Phys. **32**(11), 3448–3454 (2005).
[35]S. M. Bentzen *et al.*, "Quantitative analyses of normal tissue effects in the clinic (QUANTEC): An introduction to the scientific issues," Int. J. Radiat. Oncol., Biol., Phys. **76**(3 Suppl), S3–S9 (2010).
[36]J. O. Deasy *et al.*, "Radiotherapy dose-volume effects on salivary gland function," Int. J. Radiat. Oncol., Biol., Phys. **76**(3 Suppl), S58–S63 (2010).
[37]K. S. C. Chao *et al.*, "Intensity-modulated radiation therapy in head and neck cancers: The Mallinckrodt experience," Int. J. Cancer **90**(2), 92–103 (2000).
[38]M. M. Coselmon *et al.*, "Improving IMRT delivery efficiency using intensity limits during inverse planning," Med. Phys. **32**(5), 1234–1245 (2005).
[39]C. Men *et al.*, "An exact approach to direct aperture optimization in IMRT treatment planning," Phys. Med. Biol. **52**(24), 7333–7352 (2007).