Received: 28 July 2023 | Revised: 22 March 2024 | Accepted: 26 March 2024

DOI: 10.1002/mp.17070

RESEARCH ARTICLE

MEDICAL PHYSICS

Simultaneous reduction of number of spots and energy layers in intensity modulated proton therapy for rapid spot scanning delivery

Anqi Fu[1] | **Vicki T. Taasti**[2] | **Masoud Zarepisheh**[1]

[1]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, New York, USA

[2]Danish Center for Particle Therapy, Aarhus University Hospital, Aarhus, Denmark

Correspondence
Anqi Fu, 321 East 61st Street, New York, NY 10065, USA.
Email: fua@mskcc.org

Funding information
National Institutes of Health, Grant/Award Number: P30 CA008748

Abstract

Background: Reducing proton treatment time improves patient comfort and decreases the risk of error from intrafractional motion, but must be balanced against clinical goals and treatment plan quality.

Purpose: To improve the delivery efficiency of spot scanning proton therapy by simultaneously reducing the number of spots and energy layers using the reweighted l_1 regularization method.

Methods: We formulated the proton treatment planning problem as a convex optimization problem with a cost function consisting of a dosimetric plan quality term plus a weighted l_1 regularization term. We iteratively solved this problem and adaptively updated the regularization weights to promote the sparsity of both the spots and energy layers. The proposed algorithm was tested on four head-and-neck cancer patients, and its performance, in terms of reducing the number of spots and energy layers, was compared with existing standard l_1 and group l_2 regularization methods. We also compared the effectiveness of the three methods (l_1, group l_2, and reweighted l_1) at improving plan delivery efficiency without compromising dosimetric plan quality by constructing each of their Pareto surfaces charting the trade-off between plan delivery and plan quality.

Results: The reweighted l_1 regularization method reduced the number of spots and energy layers by an average over all patients of 40% and 35%, respectively, with an insignificant cost to dosimetric plan quality. From the Pareto surfaces, it is clear that reweighted l_1 provided a better trade-off between plan delivery efficiency and dosimetric plan quality than standard l_1 or group l_2 regularization, requiring the lowest cost to quality to achieve any given level of delivery efficiency.

Conclusions: Reweighted l_1 regularization is a powerful method for simultaneously promoting the sparsity of spots and energy layers at a small cost to dosimetric plan quality. This sparsity reduces the time required for spot scanning and energy layer switching, thereby improving the delivery efficiency of proton plans.

KEYWORDS
optimization, proton treatment planning, regularization

1 | INTRODUCTION

Intensity modulated proton therapy (IMPT) is typically delivered via the pencil beam scanning technique. The patient is irradiated by a sequence of proton spots, arranged laterally to cover the treatment volume, where the depth of penetration of each spot is determined by its energy layer. During IMPT, protons are transmitted

 © 2024 American Association of Physicists in Medicine.

spot-by-spot within every energy layer, and layer-by-layer within every beam over a set of fixed-angle beams. The total plan delivery time is roughly equal to the total switching time between beams (gantry rotation plus beam setup time), switching time between energy layers, travel time between spots, and dose delivery time at each spot.[1,2]

In this study, we seek to reduce IMPT delivery time by reducing the number of proton spots and energy layers. A shorter treatment time is desirable because it improves patient comfort, increases patient throughput (hence lowering treatment costs), and decreases the risk of error due to intrafractional motion.[3–5] Simultaneously, we want to ensure clinical goals are met and the quality of the treatment plan remains uncompromised. This trade-off between delivery time and plan quality has been the subject of abundant research.

One thread of research focuses on greedy algorithms for energy layer assignment. These algorithms combine a variety of techniques for control point sampling, energy layer distribution, energy layer filtration, and spot optimization.[6] The goal is to reduce energy layer switching time by pruning the number of energy layers and sequencing them so layer switches only occur from low-to-high energy level.[7,8]

Another strand of research takes a structured optimization-based approach. For example, Müller and Wilkens[9] directly minimize the sum of spot intensities as part of a prioritized optimization routine. Other authors formulate the proton treatment delivery problem as a mixed-integer program (MIP), where each energy layer[10] or path between layers[11,12] is associated with a binary indicator variable. Their objective is to minimize the dose fidelity (e.g., over/underdose to the target) plus some penalty or constraints on the energy layers, which promote a lower switching time. Although mathematically elegant, these MIPs are computationally difficult to solve, as they scale poorly with the number of energy layers due to the combinatorial nature of the problem.

To avoid this issue, researchers turned their attention to continuous optimization models. A proton treatment planning problem in this category only contains continuous variables, like spot intensities and doses. Typically, the objective includes a regularization function that is selected to encourage sparsity (i.e., more zero elements) in the spots and energy layers. The regularizer applies a penalty to the total spot intensity within each layer group. A variety of options have been proposed for this penalty function: logarithm,[13,14] $l_{2,1/2}$ norm,[15] and l_2 norm.[1,11,16,17] The last of these is of particular interest because it is convex and widely used in statistics for promoting group sparsity; the associated regularizer is known as the group lasso.[18–21]

The standard lasso (i.e., l_1 norm penalty) promotes sparsity of the spot intensities, but does not directly penalize the energy layers. The group lasso (i.e., group l_2 norm penalty) promotes sparsity of the energy layers, but actually *increases* the number of nonzero spots. Since IMPT delivery time depends on both the number of spots and energy layers,[1] neither of these regularizers is ideal. In this paper, we propose a new regularization method that simultaneously reduces the number of nonzero spots and energy layers, while upholding treatment plan quality. Our reweighted l_1 method combines the l_1 penalty from standard lasso with a weighting mechanism that differentiates between the spots of different energy layers, similar to the group lasso. We test the proposed method on four head-and-neck cancer patients and demonstrate its ability to (1) reduce the number of spots and energy layers simultaneously, and (2) provide a better trade-off between dosimetric plan quality and plan delivery efficiency than existing regularization methods (i.e., standard l_1 and group lasso).

2 | METHODS AND MATERIALS

2.1 | Problem formulation

We discretize the patient's body into m voxels and the proton beams into n spots. For each spot $j \in \{1, \dots, n\}$, we calculate the radiation dose delivered by a unit intensity of that spot to voxel $i \in \{1, \dots, m\}$ and call this value A_{ij}. The dose influence matrix is then $A \in \mathbf{R}_+^{m\times n}$, where its rows correspond to the voxels and its columns to the spots. Let $p \in \mathbf{R}_+^m$ be the prescription vector, that is, p_i equals the physician-prescribed dose if i is a target voxel and zero otherwise. We concatenate the spot intensities of all beams and all energy layers within each beam, denoting this collective as the vector $x \in \mathbf{R}^n$. The typical treatment planning problem seeks a vector of spot intensities x that minimizes the deviation of the delivered dose, $d = Ax$, from the prescription p. This deviation can be decomposed into a penalty on the overdose, $\overline{d} = (d-p)_+ = \max(d-p, 0)$, and the underdose, $\underline{d} = (d-p)_- = -\min(d-p, 0)$, which we combine to form the *cost function*

$$f(\overline{d}, \underline{d}) = \sum_{i=1}^{m} \overline{w}_i \overline{d}_i^2 + \underline{w}_i \underline{d}_i^2, \tag{1}$$

where $\overline{w}, \underline{w} \in \mathbf{R}_+^n$ are penalty parameters that determine the relative importance of the over/underdose to the treatment plan. (Note the underdose is ignored for non-target voxels i because $p_i = 0$.)

Dose constraints are defined for each anatomical structure. For a given structure s, let $A^s \in \mathbf{R}_+^{m_s\times n}$ be the row slice of A containing only the rows of the m_s voxels in s. A maximum dose constraint takes the form of $A^s x \le d_s^{\max}$, where $d_s^{\max}$ is an upper bound. Similarly, a mean dose constraint is of the

form $\frac{1}{m_s}\mathbf{1}^T A^s x \le d_s^{\text{mean}}$. By stacking the constraint matrices/vectors for all S structures, we can represent the set of dose constraints as a single linear inequality $Bx \le c$, where $B = [A^1, \frac{1}{m_1}\mathbf{1}^T A^1, \dots, A^S, \frac{1}{m_S}\mathbf{1}^T A^S]$ and $c = [d_1^{\max}, d_1^{\text{mean}}, \dots, d_S^{\max}, d_S^{\text{mean}}]$. Then, our treatment planning problem is

$$\begin{aligned} \text{minimize} \quad & f(\overline{d}, \underline{d}) \\ \text{subject to} \quad & \overline{d} = (Ax-p)_+, \quad \underline{d} = (Ax-p)_-, \quad Bx \le c \\ & x \ge 0, \quad \overline{d} \ge 0, \quad \underline{d} \ge 0 \end{aligned} \tag{2}$$

with variables $x \in \mathbf{R}^n$, $\overline{d} \in \mathbf{R}^m$, and $\underline{d} \in \mathbf{R}^m$. Since the objective function f is monotonically increasing in $\overline{d}$ and $\underline{d}$ over the nonnegative reals, we can write this problem equivalently as

$$\begin{aligned} \text{minimize} \quad & f(\overline{d}, \underline{d}) \\ \text{subject to} \quad & Ax - \overline{d} + \underline{d} = p, \quad Bx \le c \\ & x \ge 0, \quad \overline{d} \ge 0, \quad \underline{d} \ge 0. \end{aligned} \tag{3}$$

(The derivation is provided in supplementary material.) Problem 3 is a convex quadratic program (QP), hence can be solved using standard convex methods, for example, the alternating direction method of multipliers (ADMM)[22,23] or interior-point methods.[24,25] The reader is referred to Boyd and Vandenberghe[26] and Nocedal and Wright[27] for a thorough discussion of convex optimization.

2.2 | Common regularizers

The cost function defined in one focuses solely on the difference of the delivered dose from the prescription, that is, the dosimetric plan quality. However, in our treatment scenario, we are also interested in reducing the dose delivery time, that is, increasing the plan delivery efficiency. The delivery time is positively correlated with the number of nonzero spots (spot scanning rate) and nonzero energy layers (energy switching time).[1,28] Thus, we want to augment the objective of problem 3 with a *regularization function* $r : \mathbf{R}^n \to \mathbf{R}$, which penalizes the spot vector x in a way that reduces the number of nonzero spots/layers, while maintaining high plan quality. The regularized treatment planning problem is

$$\begin{aligned} \text{minimize} \quad & f(\overline{d}, \underline{d}) + \lambda r(x) \\ \text{subject to} \quad & Ax - \overline{d} + \underline{d} = p, \quad Bx \le c \\ & x \ge 0, \quad \overline{d} \ge 0, \quad \underline{d} \ge 0 \end{aligned} \tag{4}$$

with respect to x, $\overline{d}$, and $\underline{d}$. Here we have introduced a regularization weight $\lambda \ge 0$ to balance the trade-off between dosimetric plan quality, represented by the cost $f(\overline{d}, \underline{d})$, and plan delivery efficiency, as captured by the regularization term $r(x)$. A larger value of λ places more importance on efficiency.

In the following subsections, we review a few regularization functions that have been suggested in the literature. Let $\mathcal{J} = \{1, \dots, n\}$ and $\mathcal{G} = \{\mathcal{J}_1, \dots, \mathcal{J}_G\}$ be a set of subsets of $\mathcal{J}$, where each $\mathcal{J}_g \subseteq \mathcal{J}$ has exactly $n_g \le n$ elements. Specifically in our setting, $\mathcal{G}$ represents a partition of n spots into G energy layers with $\mathcal{J}_g$ containing the indices of the n_g spots in layer g.

2.2.1 | l_0 regularizer

One method of reducing the delivery time is to directly penalize the number of nonzero spots. This can be accomplished via the l_0 regularizer

$$r_0(x) = \|x\|_0 = \mathbf{card}(\{j : x_j \ne 0\}), \tag{5}$$

which we have defined as the number of nonzero elements in x. (Here $\mathbf{card}(A)$ denotes the cardinality of set A). Unfortunately, the l_0 regularization function is computationally expensive to implement. To solve problem 4 with $r = r_0$, we would need to solve a series of large MIP in order to determine the optimal subset of nonzero spots out of all possible combinations from $\mathcal{J}$.[29] As the number of spots n is typically very large (on the order of 10^3 to 10^4), this quickly becomes computationally intractable.

Another option is to apply the l_0 regularizer to the energy layers:

$$\tilde{r}_0(x) = \left\| \left[\sum_{j\in\mathcal{J}_1} x_j, \dots, \sum_{j\in\mathcal{J}_G} x_j \right] \right\|_0 = \mathbf{card}\left(\left\{ g : \sum_{j\in\mathcal{I}_g} x_j \ne 0 \right\} \right). \tag{6}$$

In this case, $\tilde{r}_0$ returns the number of nonzero energy layers, where a layer g is zero if and only if all its spots are zero, that is, $\sum_{j\in\mathcal{J}_g} x_j = 0$. The associated combinatorial problem or MIP simplifies to finding the optimal subset of nonzero layers, which is more manageable since G is typically on the order of 10^2. Cao et al.[10] developed an iterative method to solve an approximation of this MIP and were able to reduce the number of proton energies in their IMPT plan, while satisfying certain dosimetric criteria. Nevertheless, as combinatorial optimization is still expensive, we turn our attention to a different regularization function.

2.2.2 | l_1 regularizer

A common approximation of the l_0 regularizer is the l_1 norm. Define the l_1 regularization function to be

$$r_1(x) = \|x\|_1 = \sum_{j=1}^{n} |x_j|. \tag{7}$$

This function is closed, convex, and continuous. When used as a regularizer in problem 4, it produces a convex optimization problem – a form of the lasso problem – that promotes sparsity in the solution vector x.[30] The lasso problem is well-studied in the literature,[31,32] and various methods have been developed to solve it quickly and efficiently.[33–36]

One downside of the l_1 regularizer is that it does not differentiate between energy layers and thus is insensitive to the number of layers: since the spot vector is nonnegative, the absolute value of its elements $|x_j| = x_j$, and any sum over energy layers decouples into the sum over all spots $\sum_{g=1}^{G} \sum_{j\in\mathcal{J}_g} |x_j| = \sum_{j=1}^{n} x_j$. As energy layer switching time is typically longer than spot delivery or travel time, l_1 regularization is not the most effective method for improving plan delivery efficiency.

2.2.3 | Group l_2 regularizer

The group l_2 regularizer, also known as the group lasso, provides an alternative method for efficiently implementing group penalties. This regularization function is defined as

$$r_2(x) = \sum_{g=1}^{G} \frac{1}{\sqrt{n_g}} \|\{x_j : j \in \mathcal{J}_g\}\|_2 = \sum_{g=1}^{G} \sqrt{\frac{1}{n_g} \sum_{j\in\mathcal{J}_g} x_j^2}. \tag{8}$$

It is the sum of the l_2 norm of the vector corresponding to each group (i.e., energy layer), weighted by the reciprocal of the square root of the total number of group elements.[18] (The weights $\frac{1}{\sqrt{n_g}}$ may differ across applications; see Simon and Tibshirani[37] for alternatives). Group lasso has been widely researched in the context of statistical analysis and regression,[19–21,38,39] and many algorithms exist for solving the associated optimization problem effectively.[40–42] Jensen et al.[16] employed a version of the group lasso to perform adaptive IMPT energy layer optimization.

In our proton treatment delivery scenario, the group lasso is capable of differentiating between energy layers, and thus is a good regularizer for reducing the number of nonzero layers. However, it also tends to *increase* the number of nonzero spots within the active layers, as the quadratic penalty term in Equation (8) mostly ignores small spot intensities (square of a small x_j is near zero). This failure to generate sparsity in the spot vector due to the characteristics of the l_2 norm make it an inadequate regularizer for our purposes.

2.3 | Reweighted l_1 method

As we have discussed, the l_1 regularization function promotes sparsity of the spots, but not the energy layers. The group l_2 regularization function promotes sparsity of the layers, but not the spots – indeed, it tends to produce dense spot vectors due to the l_2 norm. In this section, we introduce the reweighted l_1 regularization method, which promotes sparsity in both the spots and the energy layers.

The reweighted l_1 method assigns a weight to every spot based on its intensity and energy layer. The weights are chosen to counteract the intensity of each layer, so that all spots, regardless of intensity, contribute roughly equally to the total regularization penalty. This is done to imitate the "ideal" group l_0 regularizer (Equation 6), which counts every nonzero layer as one unit (due to **card**) regardless of intensity.

Formally, we define the *weighted group l_1 regularizer*

$$r_3(x;\beta) = \sum_{g=1}^{G} \beta_g \sum_{j\in\mathcal{J}_g} |x_j| \tag{9}$$

with weight parameters $\beta_g \in \mathbf{R}_+ \cup \{+\infty\}$ for $g = 1, \ldots, G$. An intuitive way to set β_g is to make it inversely proportional to the optimal total intensity of energy layer g, that is,

$$\beta_g = \begin{cases} \frac{1}{\sum_{j\in\mathcal{J}_g} x_j^\star} & \sum_{j\in\mathcal{J}_g} x_j^\star \neq 0 \\ +\infty & \sum_{j\in\mathcal{J}_g} x_j^\star = 0 \end{cases},$$

where $x^\star \in \mathbf{R}_+^n$ is an optimal spot vector. However, we do not know $x^\star$ beforehand. We will approximate this weighting scheme iteratively using the reweighted l_1 method.

The reweighted l_1 method is a type of majorization-minimization (MM) algorithm, which solves an optimization problem by iteratively minimizing a surrogate function that majorizes the actual objective function. MM algorithms have a rich history in the literature,[43–46] and reweighted l_1 in particular has been used to solve problems in portfolio optimization,[47] matrix rank minimization,[48,49] and sparse signal recovery.[50] Research has shown that it is fast and robust, outperforming standard l_1 regularization in a variety of settings.

We now describe the reweighted l_1 method in our treatment planning setting. Let the initial weights $\beta^{(1)} = \mathbf{1}$. At each iteration $k = 1, 2, \ldots$,

1. Set $x^{(k)}$ to a solution of

$$\begin{aligned} &\text{minimize} && f(\overline{d}, \underline{d}) + \lambda r_3(x; \beta^{(k)}) \\ &\text{subject to} && Ax - \overline{d} + \underline{d} = p, \quad Bx \le c \\ &&& x \ge 0, \quad \overline{d} \ge 0, \quad \underline{d} \ge 0. \end{aligned} \tag{10}$$

2. Compute the total intensity of each energy layer $e_g^{(k)} = \sum_{j \in \mathcal{J}_g} x_j^{(k)}$.
3. Lower threshold the solution

$$\tilde{e}_g^{(k)} = \max\left(e_g^{(k)}, \epsilon^{(k)}\right), \quad g = 1, \ldots, G,$$

where $\epsilon^{(k)} = \delta \max_{g'} e_{g'}^{(k)}$ for some small $\delta \in (0, 1)$.
4. Update the weights. First, compute the standardized reciprocals

$$\alpha_g^{(k)} = \left(\frac{1}{\tilde{e}_g^{(k)}}\right) \Bigg/ \left(\sum_{g'=1}^{G} \frac{1}{\tilde{e}_{g'}^{(k)}}\right), \quad g = 1, \ldots, G. \tag{11}$$

Then, calculate the scaling term

$$\mu^{(k)} = \sum_{g=1}^{G} \tilde{e}_g^{(k)} \Bigg/ \left(\sum_{g=1}^{G} \alpha_g^{(k)} \tilde{e}_g^{(k)}\right). \tag{12}$$

The new weights are $\beta_g^{(k+1)} = \mu^{(k)} \alpha_g^{(k)}$.
5. Terminate on convergence of the objective, or when k reaches a user-defined maximum number of iterations K.

Step 3 was introduced to ensure stability of the algorithm, so that a zero energy layer estimate $e_g^{(k)} = 0$ would not preclude the subsequent estimate $e_g^{(k+1)}$ from being nonzero. In our computational experiments, we found that a threshold fraction of $\delta = 0.01$ produced good results. Step 4 was added to ensure the l_1 regularization term ($r_1(x)$) and the reweighted l_1 term ($r_3(x)$) contribute a similar amount to the objective. To this end, the reweighted l_1 term is scaled by $\mu^{(k)}$ so that $r_1(x^{(k)}) = r_3(x^{(k)})$.

The reweighted l_1 method has a number of advantages over the other regularizers we have reviewed here. It encourages sparsity in energy layers by properly grouping the l_1 penalty. It spreads this penalty evenly across all energy layers using its weighting scheme, rather than prioritizing those spots with large magnitudes. Finally, it is easy to implement: each iteration of the algorithm only requires that we solve a simple convex problem with l_1 regularization, which can be done efficiently using many off-the-shelf solvers. Moreover, the number of reweighting iterations needed in practice is typically very low, with most of the improvement coming from the first two to three iterations, so its computational cost is overall low. As we will see in the next section, reweighted l_1 outperforms regular l_1 and group l_2 penalties in sparsifying spots/layers.

TABLE 1 Head-and-neck cancer patient information.

	Patient			
	1	2	3	4
Beam configuration	40°, 90°	74°, 285°	75°, 130°	220°, 290°
PTV volume (cm^3)	162.7	169.7	129.9	12.4
Number of voxels (m)	87012	117907	110869	50728
Number of spots (n)	6378	7011	5257	713
Number of energy layers (G)	56	70	62	38

Abbreviation: PTV, planning target volume.

2.4 | Patient population and computational framework

We compared the reweighted l_1 method with standard l_1 and group l_2 regularization on four head-and-neck cancer patient cases from The Cancer Imaging Archive (TCIA).[51,52] For each patient, we created the dose influence matrix using the proton pencil beam calculation engine in the open-source package MatRad,[53,54] assuming a relative biological effectiveness (RBE) of 1.1. The proton spots were situated on a rectangular grid with a spot spacing of 5 mm, and the grid covered the entire planning target volume (PTV) plus 1 mm out from its perimeter. The voxel resolution and dose grid resolution were both 0.98 mm × 0.98 mm × 2 mm. Every patient plan was generated using two co-planar beams with beam angles selected using a Bayesian optimization algorithm.[55] Table 1 provides more details.

Each patient had a single PTV that was prescribed a dose of $p = 70$ Gy delivered in 35 fractions of 2 Gy per fraction. The mean/max dose bounds for important structures are given in Table 2. We manually adjusted the weights in cost function 1, without exhaustive search, to obtain reasonable treatment plans. In the majority of cases, $\overline{w}_i = 1$, $\underline{w}_i = 10$ for PTV voxels i and $\overline{w}_i = 10^{-3}$ for all other voxels i provided the best results.

We implemented the standard l_1, group l_2, and reweighted l_1 based treatment planning methods in Python using CVXPY[56–58] and solved the associated optimization problems with MOSEK.[59] All computational processes were executed on a 64-bit PC with an AMD Ryzen 9 3900X CPU @ 3.80 GHz/ 12 cores and 128 GB RAM. For reweighted l_1, we ran the algorithm

TABLE 2 Dose constraints for each structure. N/A means the max/mean dose was unbounded, that is, $d^{\max} = +\infty$ or $d^{\text{mean}} = +\infty$.

Structure	Dose Bound (Gy)	
	$d^{\max}$	d^{mean}
PTV	84	N/A
Left parotid	73.5	26
Right parotid	73.5	26
Mandible	70	N/A
Spinal cord	45	N/A
Constrictors	N/A	40
Brainstem	54	N/A

Abbreviation: PTV, planning target volume.

for $K = 3$ iterations due to the diminishing benefits of more iterations.

To facilitate comparisons, we scaled the group l_2 regularizer so it lay in the same range as the standard l_1 regularizer. First, we solved problem 4 with the standard l_1 regularizer (Equation 7) and $\lambda = 1$. Let us call this solution $x^{(1)}$. Then, we computed a scaling term $\eta > 0$ such that $\eta r_2(x^{(1)}) = r_1(x^{(1)})$. When we ran the group l_2 method, we used the scaled regularization function $\tilde{r}_2(x;\eta) := \eta r_2(x)$ as the regularizer $r(x)$ in problem 4. This allowed us to obtain better spot/energy layer comparison plots between the standard l_1 and group l_2 methods. As pointed out earlier, the reweighted l_1 method is similarly scaled via step 4 of the algorithm.

After each method finished, we trimmed the optimal spot vector $x^\star$ further to increase sparsity. First, we zeroed out all elements $x_j^\star$ that fell below a fraction $\gamma \in (0, 1)$ of the maximum spot intensity, that is, we set $x_j^\star = 0$ if $x_j^\star < \gamma \max_{j'} x_{j'}^\star$. We then zeroed out all energy layers of the resulting $\tilde{x}^\star$ that fell below the same fraction of the maximum layer intensity: for each $g \in \{1, \ldots, G\}$, we set $\tilde{x}_j^\star = 0$ for all $j \in \mathcal{J}_g$ if $\sum_{j' \in \mathcal{J}_g} \tilde{x}_{j'}^\star < \gamma \max_{g'} \sum_{j' \in \mathcal{J}_{g'}} x_{j'}^\star$. A choice of $\gamma = 0.01$ provided a reasonable trade-off between sparsity and dose coverage in our computational experiments.

3 | RESULTS

3.1 | Simultaneous reduction of spots and energy layers

We first compared the results of the regularization methods on a single patient. Figure 1 depicts optimal spot intensities of the unregularized model and the l_1, group l_2, and reweighted l_1 regularized models for patient 2. For all three regularizers, a regularization weight of $\lambda = 5$ was used; this choice accentuated the difference between their spot vectors. Without regularization, about one third of the total 7011 spots are nonzero, with individual spot intensities ranging between 10^3 and 10^4. Under l_1 regularization, that fraction is reduced to only 13%, or 918 nonzero spots, as the l_1 penalty encourages further sparsity. By contrast, with group l_2 regularization, the number of active spots increases significantly to 5099 or 72%, while the average intensity drops to a little over 10^3. Reweighted l_1 regularization produced a spot vector with the lowest number of nonzero elements: just 541 or 7.7% of the spots are nonzero. For patient 2, these active spots tend to reside in the first beam, and their maximum intensity exceeds that of the other methods.

Figure 2 shows the optimal intensity of the energy layers for patient 2, using the three regularization models, all with $\lambda = 5$. Over 95% of the total 70 energy layers are nonzero under the unregularized model. These active layers are divided fairly evenly into two clusters, which coincide with the two beams delineated by the vertical red line. The total intensity of each energy layer averages between 10^4 and 10^5. With l_1 regularization, the fraction of nonzero energy layers drops to a modest 80%, where most of that reduction comes from deactivated layers at the edges of the clusters. Group l_2 regularization results in a steeper drop in the fraction of active energy layers, down to 61% with additional sparsity in the middle of both beam clusters. However, the reweighted l_1 method performs better than both of these methods, cutting the number of nonzero energy layers down to only 18 – a reduction of over 75% – with a commensurate increase in the intensity of the active layers.

A summary of the results from the different regularization methods is given in Figure 3. For a fixed λ, it is clear that reweighted l_1 achieves the lowest number of nonzero spots and nonzero energy layers out of all the methods.

3.2 | Trade-off between delivery efficiency and PTV coverage

The regularization weight in the previous section was chosen to highlight the distinctions between the optimal intensity plots. However, λ must be carefully selected to balance the trade-off between the total delivery time (highly correlated with the sparsity of the spots/energy layers) and the quality of the resulting treatment plan. Figure 4 examines this trade-off for reweighted l_1 regularization on patient 2 using two measures of plan quality: D98% and D2% for the PTV. For different values of λ, we solved problem 4 using the reweighted l_1 method, counted up the number of nonzero spots/energy layers, and calculated the optimal dose vector and PTV dose percentiles. We then plotted a point corresponding to this result in each of the subfigures of Figure 4 with the sparsity metric on the vertical axis and the plan quality metric on the horizontal axis (e.g., the unregularized point $\lambda = 0$ is marked by a triangle △). By connecting the points

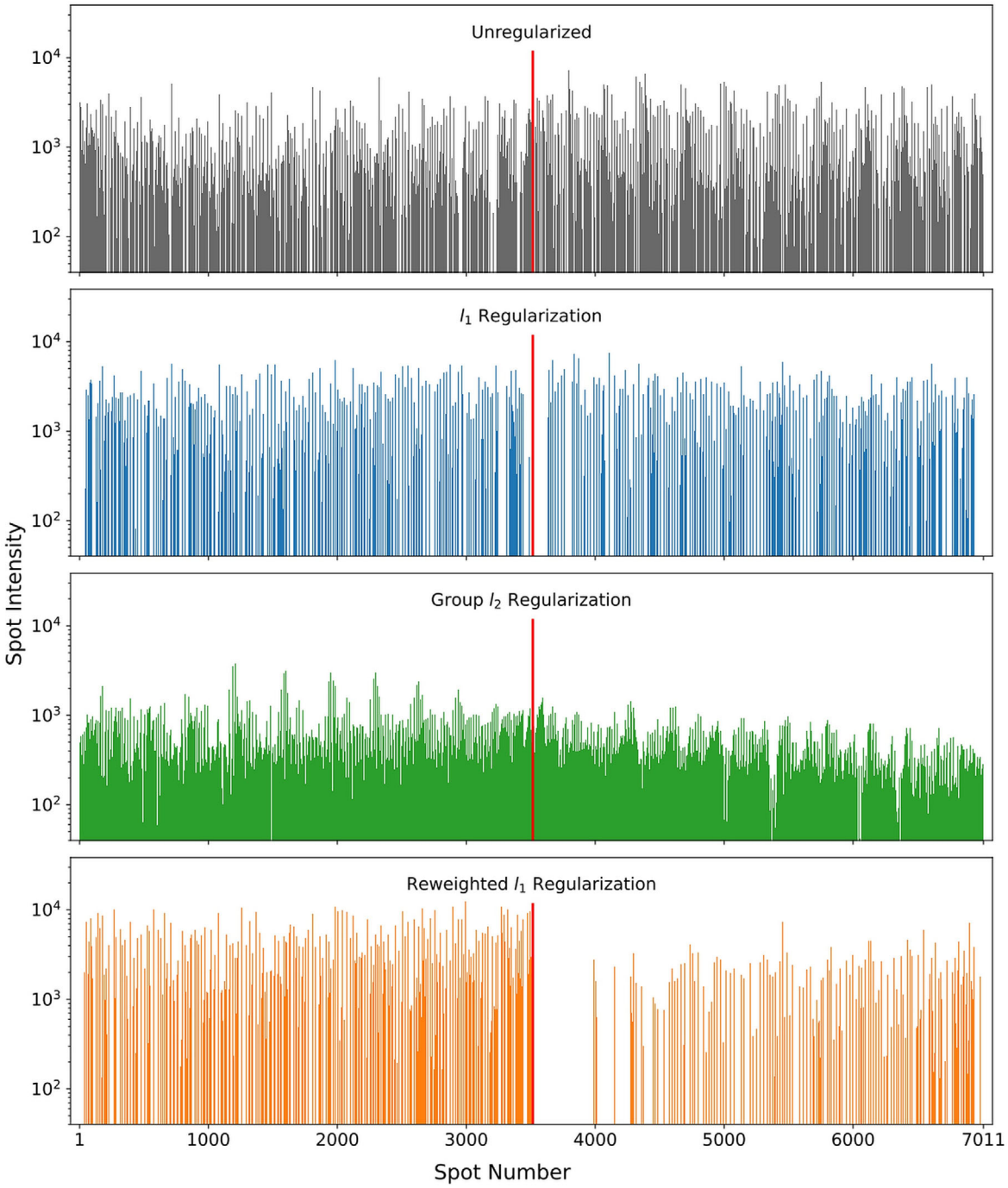

FIGURE 1 Optimal spot intensities resulting from the unregularized model and the l_1, group l_2, and reweighted l_1 regularized models ($\lambda = 5$) for patient 2. The vertical red line divides the spots associated with the first beam (1–3516) from the second beam (3517–7011).

in each subfigure, we obtained a set of Pareto optimal curves, which show the trade-off between plan delivery efficiency and plan quality.

The top left subfigure depicts the number of nonzero spots versus D98% to the PTV for λ ranging from 0 to 6.0 (marked by the square). As λ increases, the number of active spots decreases, but so does D98%. A choice of $\lambda = 0.95$ (marked by the star) achieves the lowest number of nonzero spots ≈ 720, while still maintaining D98% above 95% of the prescription, indicated by the vertical gray dotted line at 66.5 Gy. A similar plot can be seen in the bottom left subfigure, which shows the number of nonzero energy layers versus D98% to the PTV; the same choice of λ yields 27 active layers. On the righthand side, the subfigures display the number of nonzero spots (top) and energy layers (bottom) versus

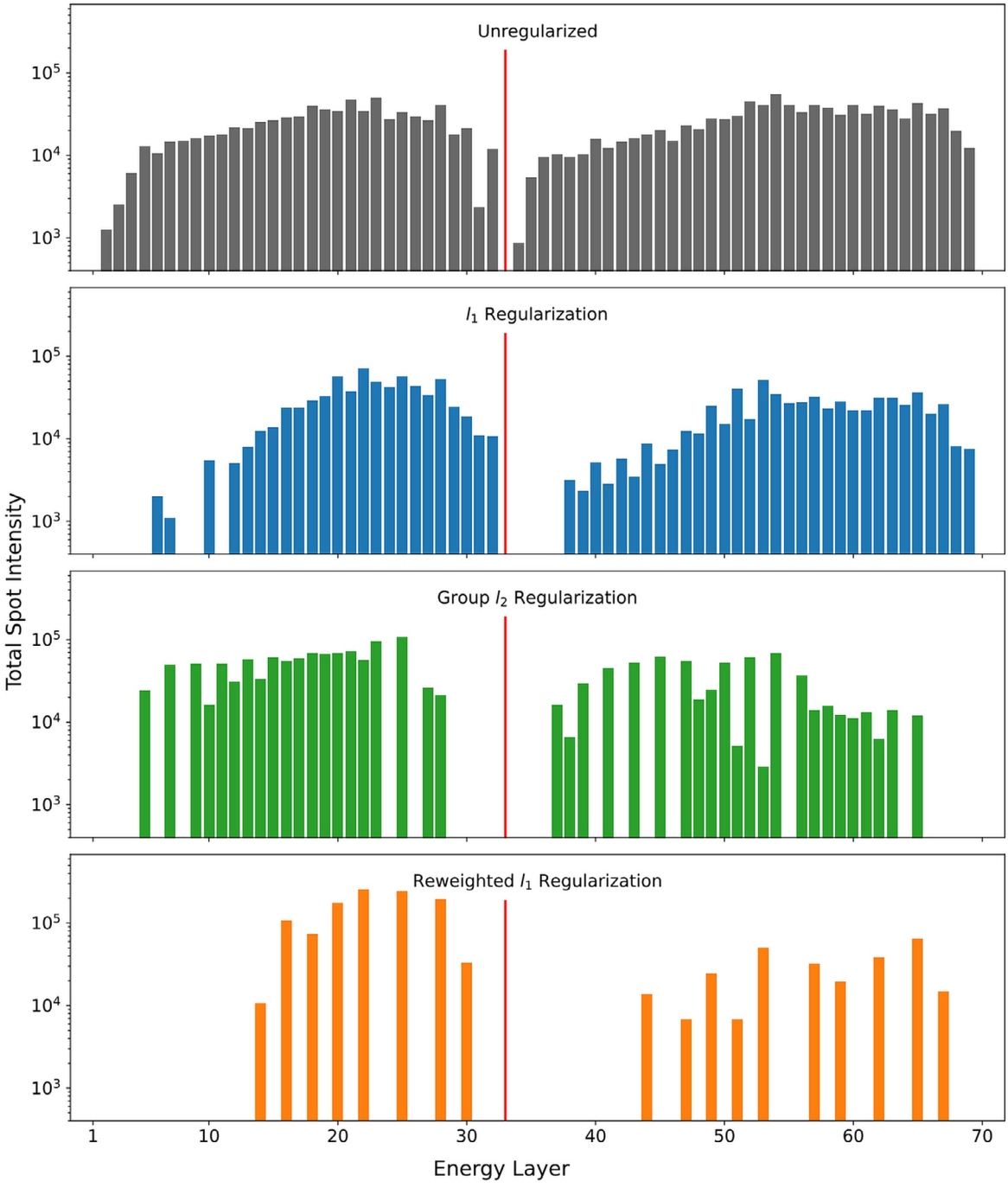

FIGURE 2 Sum of spot intensities in each energy layer (1–70) for the unregularized model and the l_1, group l_2, and reweighted l_1 regularized models ($\lambda = 5$) for patient 2. The vertical red line divides the layers associated with the first beam (1–33) from the second beam (34–70).

D2% to the PTV. As the regularization weight increases, D2% also increases, but never exceeds 108% of the prescription (as indicated by the dotted line at 75.6 Gy) for any $\lambda \leq 0.95$. Thus, out of all the weights, $\lambda = 0.95$ yields a good trade-off between sparsity and PTV coverage: it achieves a reduction of 89% and 61% in the number of spots and energy layers, respectively, while still fulfilling all target dose constraints.

3.3 | Trade-off between delivery efficiency and overall plan quality

This section studies the Pareto optimal trade-off curves between spot/energy layer sparsity and treatment plan quality using different regularizers to determine which regularization method provides the *best* trade-off, that is, the largest increase in sparsity for the least decrease

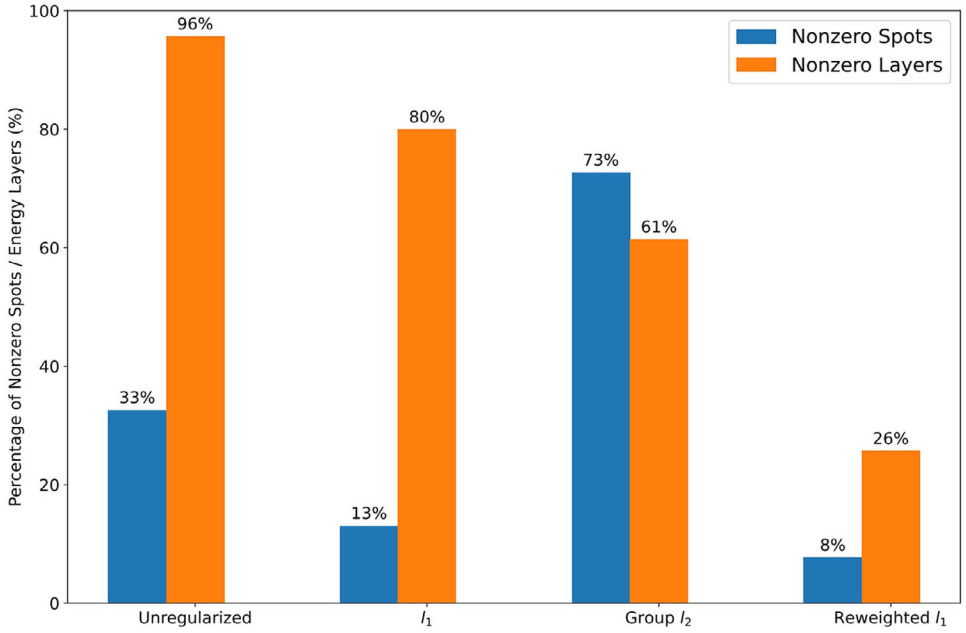

FIGURE 3 Percentage of nonzero spots/energy layers (relative to the total number of spots/layers) for the unregularized model and the l_1, group l_2, and reweighted l_1 regularized models ($\lambda = 5$) for patient 2.

in plan quality. Rather than plotting multiple dose-volume metrics, we focus on a single consolidated quality measure: the plan cost function (Equation 1), which includes dose fidelity terms for the PTV and all organs-at-risk (OARs). A lower value of $f(\overline{d}, \underline{d})$ at the optimum implies a higher quality treatment plan.

To facilitate comparison, we also focus on the *relative* change in sparsity (number of nonzero spots/energy layers) and plan cost with respect to the unregularized solution. Let x_{unreg} be the optimal spots resulting from the unregularized problem 3, and x_{reg} be the optimal spots resulting from a particular regularization method. We evaluate the plan cost function given in Equation (1) on x_{unreg} (with $\overline{d}_{unreg} = \max(Ax_{unreg} - p, 0)$ and $\underline{d}_{unreg} = -\min(Ax_{unreg} - p, 0)$) to obtain c_{unreg}, and similarly on x_{reg} to obtain c_{reg}. Then, the relative percentage change in plan cost for the regularizer is $100(c_{reg} - c_{unreg})/c_{unreg}$. The relative change in the number of nonzero spots (s) and number of nonzero energy layers (l) is defined in a similar fashion as $100(s_{reg} - s_{unreg})/s_{unreg}$ and $100(l_{reg} - l_{unreg})/l_{unreg}$, respectively. Thus, to construct the trade-off curve, we solve the regularized problem for various values of λ and plot the relative change in sparsity versus the relative change in plan cost at each solution point.

For every patient, Figure 5 depicts the relative percentage change in the number of nonzero spots versus the relative percentage change in plan cost for the l_1, group l_2, and reweighted l_1 regularization methods. The origin corresponds to the unregularized plan ($\lambda = 0$). Both the l_1 and reweighted l_1 trade-off curves drop sharply from the origin, attaining on average a 30% to 45% decrease in nonzero spots for a less than 10% increase in plan cost, with reweighted l_1 slightly outperforming l_1 by on average 5 percentage points over all 4 patients. By contrast, the number of nonzero spots rises with group l_2 regularization, increasing up to 140% within the first 10% to 15% increase in plan cost for all except patient 4. This is consistent with our spot intensity plot for patient 2 (Figure 1), which shows the spot distribution is denser under group l_2 than without regularization.

Figure 6 depicts the relative percentage change in the number of nonzero energy layers versus the relative percentage change in plan cost for the three regularization methods. Both l_1 and group l_2 trade-off curves decrease moderately from the origin, with group l_2 averaging about 9.5% lower number of nonzero layers for a given percentage increase in cost. This matches our observations in Figures 2 and 3 that the group l_2 function is more effective at penalizing energy layers than the l_1 norm.

However, the reweighted l_1 method significantly outperforms both these regularizers. For patient 2, it achieves an over 50% decrease in the number of nonzero energy layers for a less than 10% increase in plan cost. For the other patients, it provides a 25% to 35% reduction in active energy layers with a less than 15% increment in plan cost. The average reduction in the number of nonzero layers from reweighted l_1 exceeds the best reduction from group l_2 by 12 percentage points, and the majority of this reduction is realized with only about 10% cost to treatment plan quality, relative to the unregularized plan.

The vertical dotted lines in Figures 5 and 6 for patient 2 correspond to a 10% increase in the plan cost.

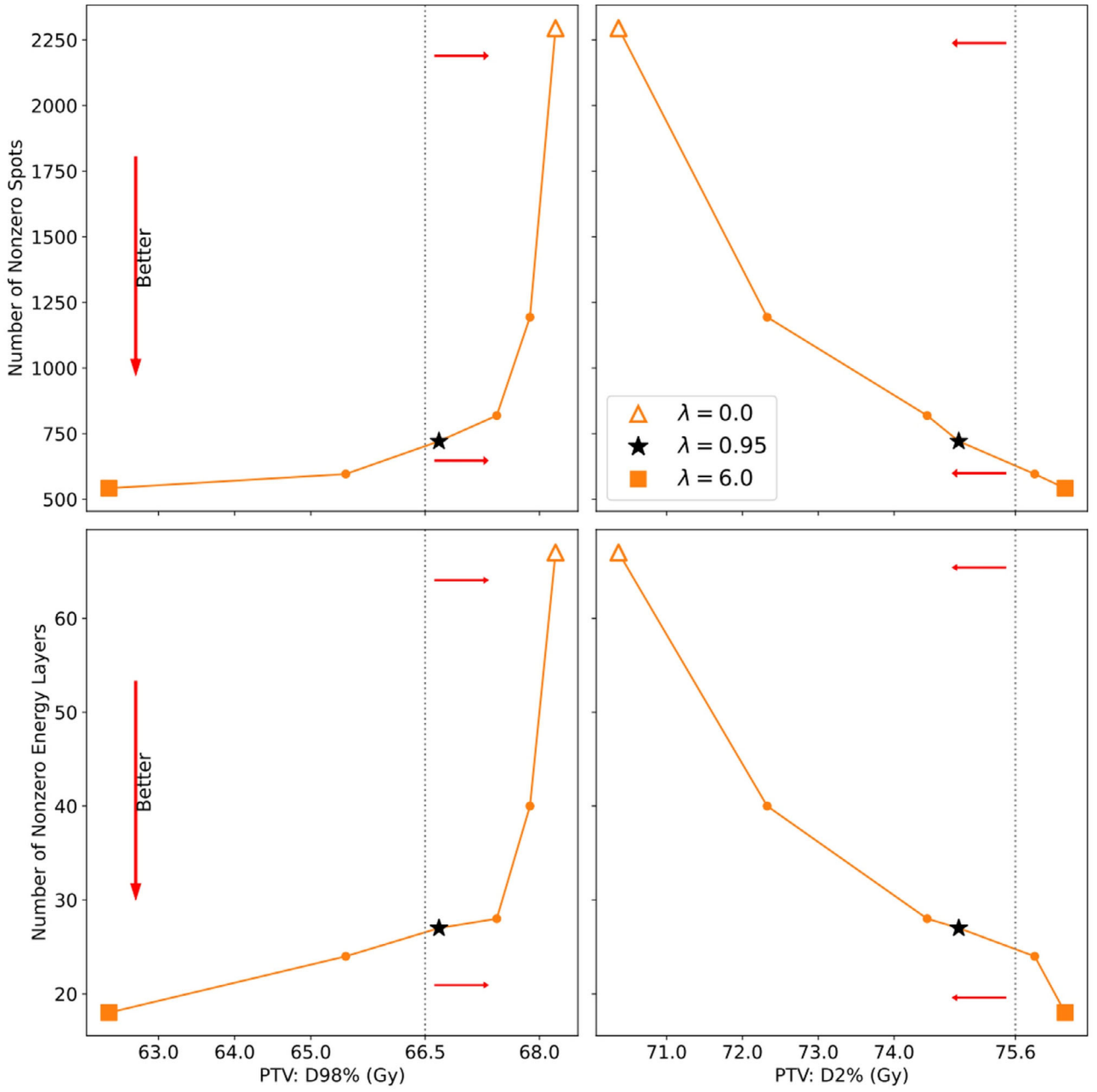

FIGURE 4 Number of nonzero spots (top) and energy layers (bottom) versus PTV dose percentile (left: D98%; right: D2%) for various values of the regularization weight λ, computed using the reweighted l_1 method on patient 2. As λ increases, the curves sweep from the △ marker ($\lambda = 0.0$) to the ■ marker ($\lambda = 6.0$). The vertical gray dotted lines indicate clinical dose constraints on the PTV (D98% $> 0.95p$ and D2% $< 1.08p$, where $p = 70$ Gy is the prescription), and the red arrows indicate the directions of desirable change (increasing D98%, decreasing D2%, and decreasing number of nonzero spots/energy layers). A choice of weight $\lambda = 0.95$, marked by the ★, produces a plan with good sparsity that respects the clinical constraints. PTV, planning target volume.

The intersection of these lines with the Pareto curves of different regularization methods demonstrates the reduction in the number of nonzero spots and energy layers obtained using different regularizers. Figure 7 (top right) depicts the dose-volume histogram (DVH) curves of the unregularized plan, the l_1 regularized plan, and the reweighted l_1 regularized plan at the same 10% relative change in plan cost. (We chose not to show the DVHs of group l_2 regularized plans because they overlap considerably with the DVHs of corresponding l_1 regularized plans, and as we saw earlier, group l_2 results in far worse delivery efficiency than l_1 or reweighted l_1 regularization). Compared to no regularization, the reweighted l_1 method reduces the number of active spots and energy layers by more than 50%, while providing relatively similar DVH curves with

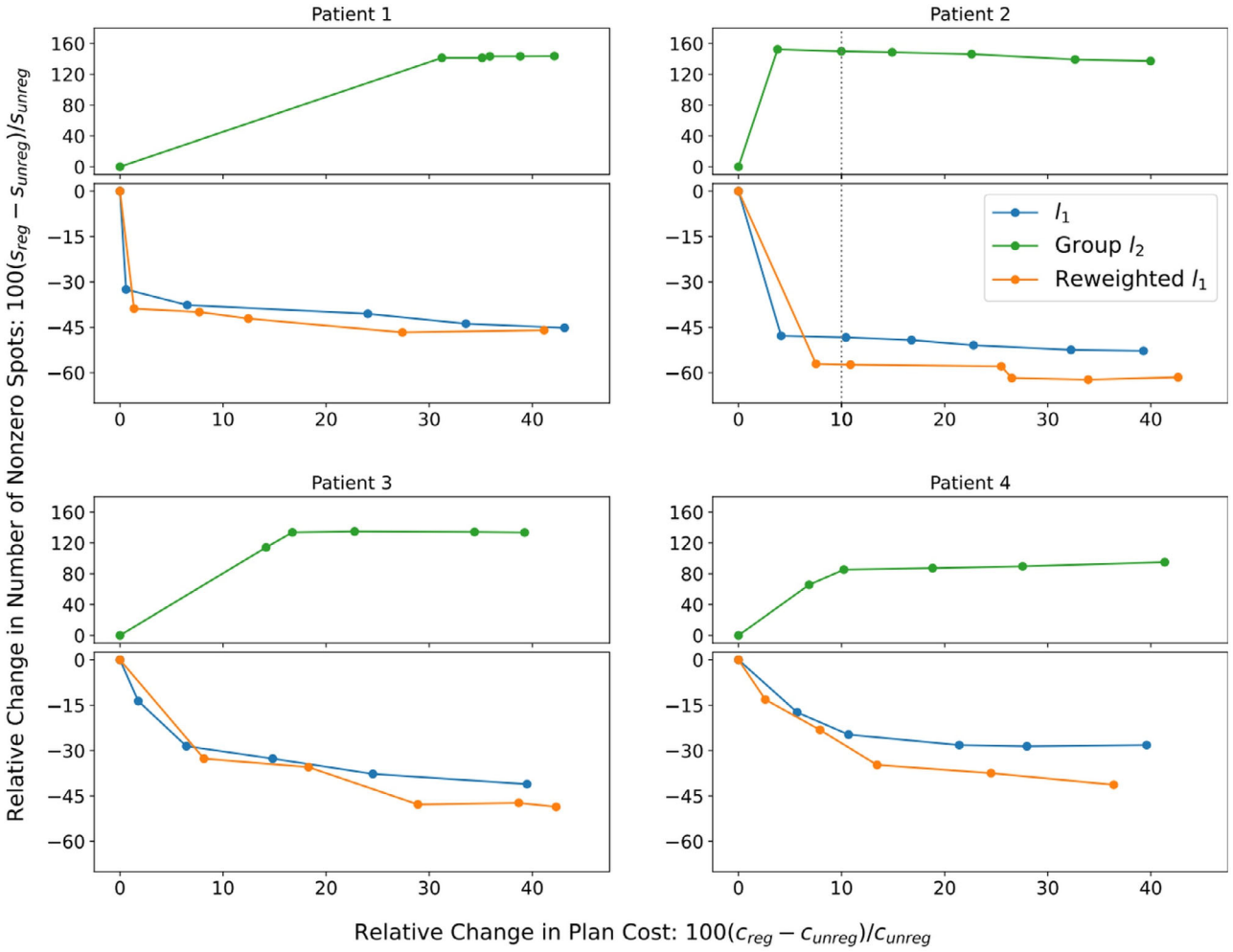

FIGURE 5 Relative change in number of nonzero spots versus relative cost to plan quality with respect to the unregularized model. For patient 2, reweighted l_1 regularization achieves a 57% reduction in the number of nonzero spots at only a 10% cost to overall plan quality, relative to the unregularized model, as indicated by the vertical gray dotted line.

different trade-offs (compromised left parotid and PTV coverage/homogeneity, and improved right parotid and mandible). One can re-adjust the PTV/OAR weights in the plan cost function of the reweighted l_1 problem to achieve more uniform trade-offs. In the same vein, compared to standard l_1 regularization, reweighted l_1 reduces the number of active spots and energy layers by about 10% and 40%, respectively, while producing almost identical DVH curves. The Pareto curves and DVHs of the other patients, also plotted at roughly 10% relative change in plan cost, tell a similar story.

3.4 | Relative improvement in delivery time

The total plan delivery time is dependent on a multitude of machine-specific factors. In this section, we provide an estimate of how regularization directly impacts the delivery time, assuming a specific set of machine parameters. The total delivery time (T) can be approximated by

$$T = T_b \times (h_b(x) - 1)_+ + T_e \times (h_e(x) - 1)_+ + \sum_{g=1}^{G} T_s \times (h_s(x, \mathcal{I}_g) - 1)_+ + \frac{\sum_{j=1}^{n} x_j}{T_d}, \quad (13)$$

where $h_b(x)$ is the number of nonzero beams, $h_e(x)$ is the number of nonzero energy layers, $h_s(x, \mathcal{I}_g)$ is the number of nonzero spots in energy layer g, T_b is the beam switching time (gantry rotation plus beam setup time), T_e is the energy layer switching time, T_s is the spot travel time, and T_d is the proton dose rate.[1,2] Following van de Water et al.,[13] we let $T_b = 30$ s, $T_e = 2$ s, $T_s = 0.01$ s, and $T_d = \frac{4\times10^{11}}{60}$ protons/s. We compute T using the unregularized model and the l_1, group l_2,

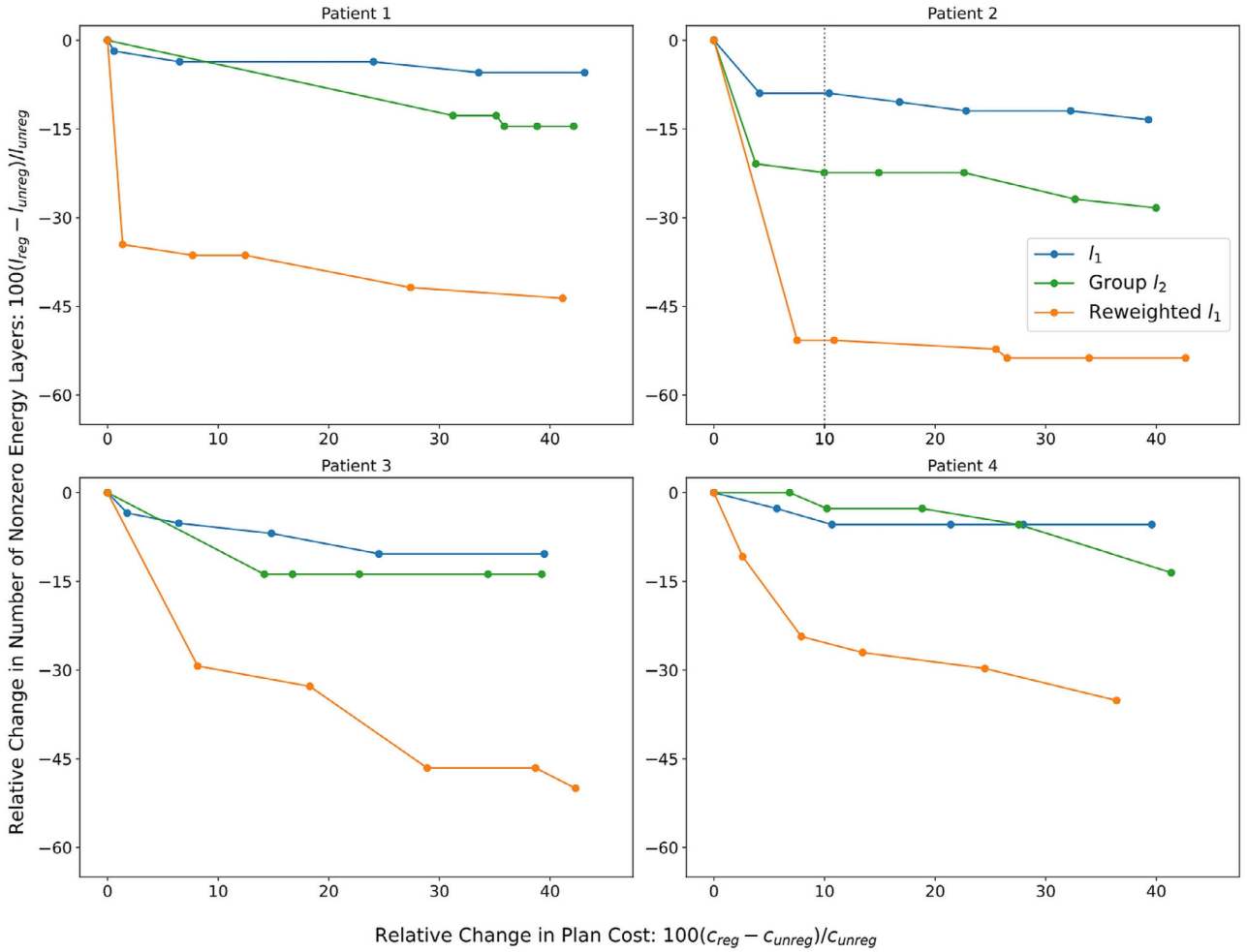

FIGURE 6 Relative change in number of nonzero energy layers versus relative cost to plan quality with respect to the unregularized model. For patient 2, reweighted l_1 regularization achieves a 50% reduction in the number of nonzero layers at only a 10% cost to overall plan quality, relative to the unregularized model, as indicated by the vertical gray dotted line.

and reweighted l_1 regularized models, with regularization weight λ chosen such that each regularized plan achieves a 10% cost to plan quality. Then, we plot the relative percentage change in delivery time of each regularized model with respect to the unregularized model (i.e., $100(T_{reg} - T_{unreg})/T_{unreg}$).

The results for patient 2 are shown in Figure 8. Standard l_1 reduces delivery time by a modest 13%, while group l_2 actually raises delivery time by 2% due to the 149% increase in the number of nonzero spots, which increases the spot delivery and spot travel time. By contrast, reweighted l_1 achieves a 44% reduction in delivery time through its simultaneous reduction of the number of nonzero spots and nonzero energy layers by 57% and 51%, respectively. This reduction comes at only a minor cost to the PTV and mandible – D2% to the PTV goes up by 4% and maximum dose to the mandible goes up by 2%. Results for the other patients reflect similar outcomes, with reweighted l_1 reducing total delivery time by between 20% and 30%; see Table S1 and Figure S1 in supplementary material for additional data and plots.

4 | DISCUSSION

This study proposed a method to improve the delivery efficiency of pencil beam scanning proton plans by simultaneously reducing the number of spots and energy layers using reweighted l_1 regularization. One can exactly model the spot/energy layer reduction problem using the l_0 regularizer, which in principle would improve plan delivery at the smallest possible cost to plan quality, but the l_0-regularized optimization problem is nonconvex and computationally prohibitive to solve. In imaging science and statistics, researchers often employ the l_1 norm as a convex surrogate for the l_0 norm, and in some cases (e.g., compressed sensing), the l_1 norm has proven to be just as effective as the l_0 norm at promoting sparsity.[61] The reweighted l_1 regularization

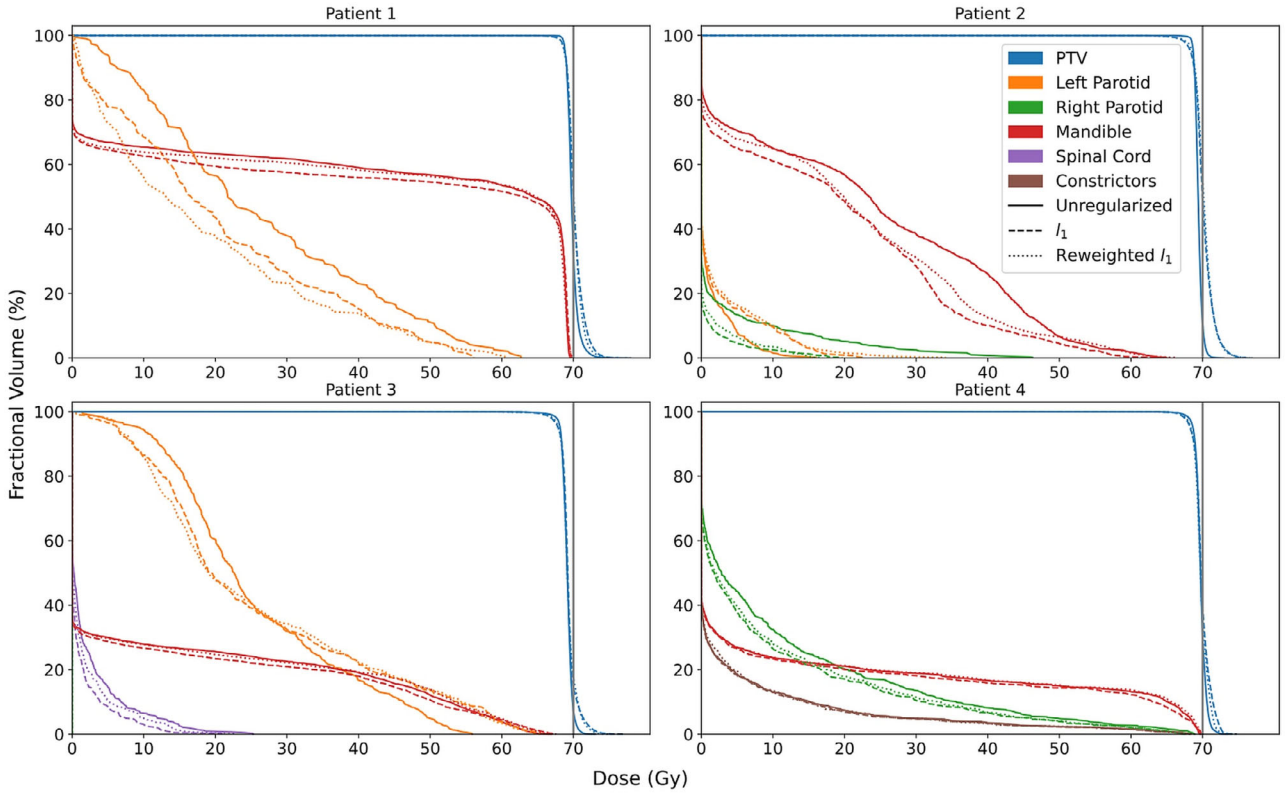

FIGURE 7 DVH curves for each patient obtained from the unregularized model (solid), and the standard l_1 (dashed) and reweighted l_1 (dotted) models regularized to approximately 10% relative cost to plan quality. The vertical gray line indicates the prescription $p = 70$ Gy. DVH, dose-volume histogram.

method was proposed[50] to bridge the gap between the l_0 regularizer and the l_1 regularizer by better approximating the l_0 norm, while retaining the convexity of the l_1 norm. In proton treatment planning, this property translates to improving the plan delivery efficiency at a lower cost to plan quality, which we have demonstrated in this work. Our limited computational experiments on four head-and-neck cancer patients show that, for the same cost to plan quality, the reweighted l_1 method reduced the number of nonzero spots by up to 10 percentage points more than standard l_1 and the number of nonzero energy layers by 25 to 30 percentage points more than group l_2 regularization.

Promoting spot/energy layer sparsity to improve plan delivery in IMPT is analogous to promoting beam profile smoothness to improve plan delivery in IMRT. Prior research has shown that plan delivery efficiency in IMRT can be significantly improved at minimal cost to dosimetric plan quality due to the phenomenon of *degeneracy*.[62,63] The structure of the treatment planning problem results in a multitude of feasible plans with near-equal objective value (i.e., quality). This same phenomenon has been observed in IMPT planning problems,[1,13,14,17] although unlike IMRT, it currently lacks a rigorous mathematical analysis. Our computational experiments demonstrated that with the reweighted l_1 method, one can reduce the number of spots and energy layers by on average 40% and 35%, respectively, without significantly compromising the dosimetric plan quality.

In this study, we have adopted a constrained optimization framework, where the dosimetric plan quality is represented by a quadratic term in the objective, the sparsity promotion is carried out via a regularization penalty term, and the mean/max clinical dose criteria are enforced by hard constraints. However, the proposed reweighted l_1 method is agnostic to the optimization framework and can also be used in conjunction with an automation tool (e.g., hierarchical optimization,[64–66] multiple criteria optimization (MCO),[67,68] knowledge-based planning (KBP)[69,70]). DVH constraints and plan robustness may be integrated into the optimization problem using existing techniques in the literature.[23,66,71–73] To limit the scope of this paper, we have not incorporated robustness into our formulation of the treatment planning problem. A preliminary robustness analysis of 13 range and setup uncertainty scenarios shows that the reweighted l_1 method generates plans with a similar level of robustness to the unregularized model's plans. Moreover, even without accounting for these uncertainties in the optimization, the DVH curves and clinical metrics of the plans lie within an acceptable range

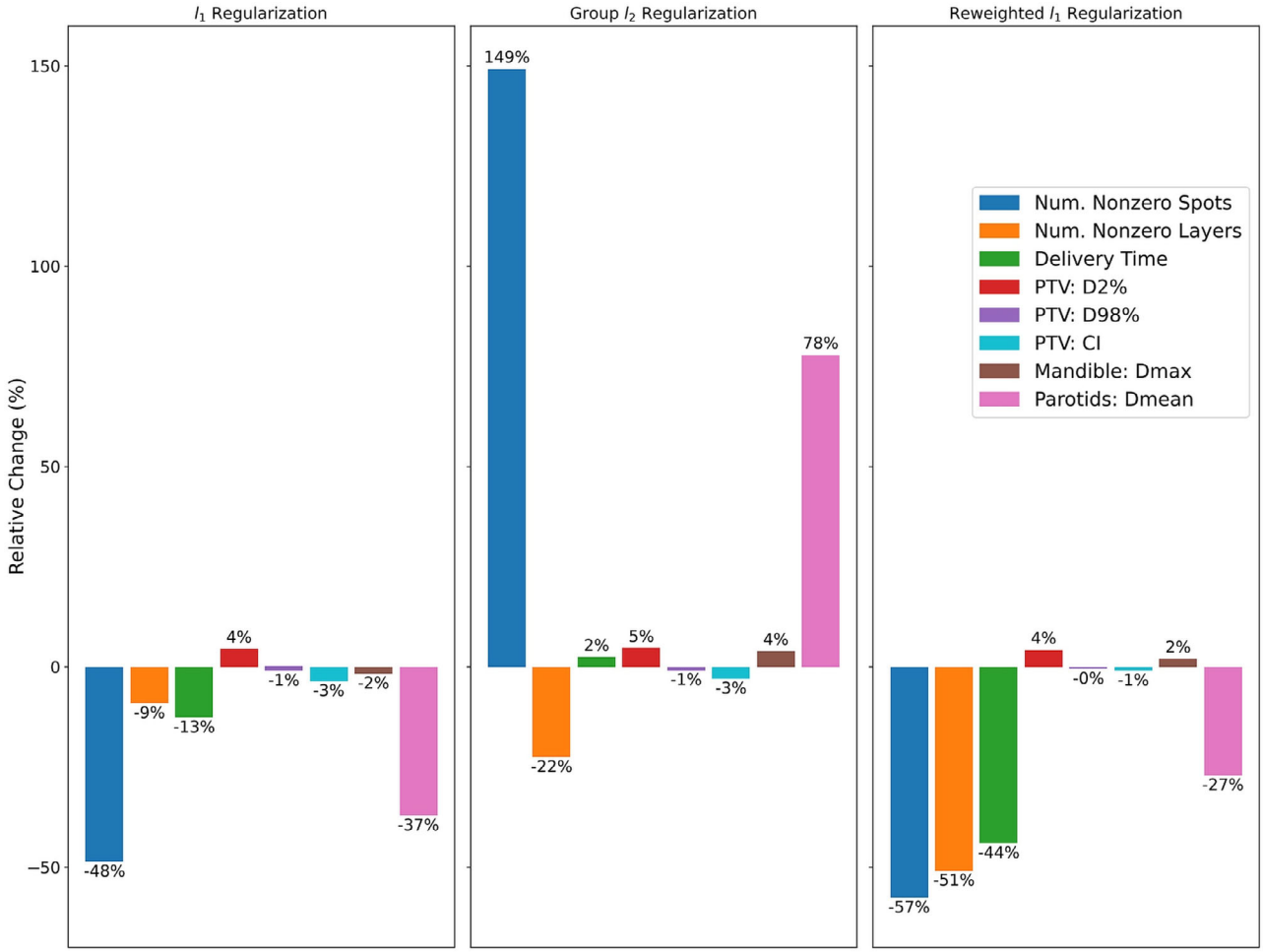

FIGURE 8 Relative change in delivery time and other plan metrics with respect to the unregularized model for l_1, group l_2, and reweighted l_1 regularization for patient 2. For each regularizer, λ was chosen such that the regularized model resulted in about 10% cost to plan quality. Here the CI is defined as the total number of voxels that received at least 95% of the prescribed dose, divided by the number of voxels in the PTV.[60] CI, conformity index; PTV, planning target volume.

across the majority of scenarios. Supplementary material provides details on our analysis and the resulting figures.

Finally, we mention that in this proof-of-concept work, we have not enforced the machine-specific minimum-monitor-unit (min-MU) constraint. One can enforce the min-MU constraint using a two-step optimization method as described in Lin et al.[17]: the first step identifies the active spots/energy layers (i.e., those with intensity greater than a pre-determined value), and the second step removes the inactive spots/energy layers and enforces the min-MU constraint on the remaining spots. Increasing the min-MU threshold also allows for a higher dose rate, which can accelerate the delivery of each spot. This is especially important because the intensities of the active spots usually increase with the overall sparsity of the spots/energy layers in the treatment plan. Gao et al.[1] suggested using different min-MU thresholds for each energy layer to further increase the dose rate and expedite spot delivery.

5 | CONCLUSIONS

The reweighted l_1 regularization method is capable of simultaneously reducing the number of spots and energy layers in a proton treatment plan, while imposing minimal cost to dosimetric plan quality. Moreover, it achieves a better trade-off between delivery efficiency and plan quality than standard l_1 and group l_2 regularization. Thus, reweighted l_1 regularization is a powerful method for improving the delivery of proton therapy.

ACKNOWLEDGMENTS

This work was partially supported by MSK Cancer Center Support Grant/Core Grant from the NIH (P30 CA008748).

CONFLICT OF INTEREST STATEMENT

The authors have no relevant conflicts of interest to disclose.

REFERENCES

1. Gao H, Clasie B, McDonald M, Langen KM, Liu T, Lin Y. Technical note: plan-delivery-time constrained inverse optimization method with Minimum-MU-per-Energy-Layer (MMPEL) for efficient pencil beam scanning proton therapy. *Med Phys*. 2020;47:3892-3897.
2. Zhang G, Shen H, Lin Y, Chen RC, Long Y, Gao H. Energy layer optimization via energy matrix regularization for proton spot-scanning arc therapy. *Med Phys*. 2022;49:5752-5762. doi:10.1002/mp.15836
3. Li H, Zhu XR, Zhang X. Reducing dose uncertainty for spot-scanning proton beam therapy of moving tumors by optimizing the spot delivery sequence. *Int J Radiat Oncol Biol Phys*. 2015;93:547-556. doi:10.1016/j.ijrobp.2015.06.019
4. Suzuki K, Palmer MB, Sahoo N, et al,. Quantitative analysis of treatment process time and throughput capacity for spot scanning proton therapy. *Med Phys*. 2016;43:3975-3986.
5. Mah D, Chen CC, Nawaz AO, et al,. Retrospective analysis of reduced energy switching and room switching times on throughput efficiency of a multi-room proton therapy center. *Brit J Radiol*. 2020;93:20190820. doi:10.1259/bjr.20190820
6. Ding X, Li X, Zhang M, Kabolizadeh P, Stevens C, Yan D. Spot-Scanning Proton Arc (SPArc) therapy: the first robust and delivery-efficient spot-scanning proton arc therapy. *Int J Radiat Oncol Biol Phys*. 2016;96:1107-1116. doi:10.1016/j.ijrobp.2016.08.049
7. Liu G, Li X, Zhao L, et al,. A novel energy sequence optimization algorithm for efficient spot-scanning proton arc (SPArc) treatment delivery. *Acta Oncol*. 2020;59:1178-1185. doi:10.1080/0284186X.2020.1765415
8. Engwall E, Battinelli C, Wase V, et al,. Fast robust optimization of proton PBS arc therapy plans using early energy layer selection and spot assignment. *Phys Med Biol*. 2022;67:065010. doi:10.1088/1361-6560/ac55a6
9. Müller BS, Wilkens JJ. Prioritized efficiency optimization for intensity modulated proton therapy. *Phys Med Biol*. 2016;61:8249-8265. doi:10.1088/0031-9155/61/23/8249
10. Cao W, Lim G, Liao L, et al,. Proton energy optimization and reduction for intensity-modulated proton therapy. *Phys Med Biol*. 2014;59:6341-6354. doi:10.1088/0031-9155/59/21/6341
11. Wuyckens S, Saint-Guillain M, Janssens G, et al,. Treatment planning in arc proton therapy: comparison of several optimization problem statements and their corresponding solvers. *Comput Biol Med*. 2022;148:105609. doi:10.1016/j.compbiomed.2022.105609
12. Wuyckens S, Zhao L, Saint-Guillain M, et al,. Bi-criteria Pareto optimization to balance irradiation time and dosimetric objectives in proton arc therapy. *Phys Med Biol*. 2022;67:245017. doi:10.1088/1361-6560/aca5e9
13. van de Water S, Kooy HM, Heijmen BJM, Hoogeman MS. Shortening delivery times of intensity modulated proton therapy by reducing proton energy layers during treatment plan optimization. *Int J Radiat Oncol Biol Phys*. 2015;92:460-468. doi:10.1016/j.ijrobp.2015.01.031
14. van de Water S, Belosi MF, Albertini F, Winterhalter C, Weber DC, Lomax AJ. Shortening delivery times for intensity-modulated proton therapy by reducing the number of proton spots: an experimental verification. *Phys Med Biol*. 2020;65:095008. doi:10.1088/1361-6560/ab7e7c
15. Gu W, Ruan D, Lyu Q, Zou W, Dong L, Sheng K. A novel energy layer optimization framework for spot-scanning proton arc therapy. *Med Phys*. 2020;47:2072-2084. doi:10.1002/mp.14083
16. Jensen MF, Hoffmann L, Petersen JBB, Møller DS, Alber M. Energy layer optimization strategies for intensity-modulated proton therapy of lung cancer patients. *Med Phys*. 2018;45:4355-4363. doi:10.1002/mp.13139
17. Lin Y, Clasie B, Liu T, McDonald M, Langen KM, Gao H. Minimum-MU and sparse-energy-layer (MMSEL) constrained inverse optimization method for efficiently deliverable PBS, plans. *Phys Med Biol*. 2019;64:205001. doi:10.1088/1361-6560/ab4529
18. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J R Stat Soc, B: Stat Methodol*. 2006;68:49-67. doi:10.1111/j.1467-9868.2005.00532.x
19. Meier L, van de Geer S, Bühlmann P. The group lasso for logistic regression. *J R Stat Soc, B: Stat Methodol*. 2008;70:53-71. doi:10.1111/j.1467-9868.2007.00627.x
20. Lim M, Hastie T. Learning interactions via hierarchical group-lasso regularization. *J Comput Graph Stat*. 2015;24:627-654. doi:10.1080/10618600.2014.938812
21. Ivanoff S, Picard F, Rivoirard V. Adaptive lasso and group-lasso for functional poisson regression. *J Mach Learn Res*. 2016;17:1903-1948.
22. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found Trends Mach Learn*. 2011;3:1-122.
23. Fu A, Taasti VT, Zarepisheh M. Distributed and scalable optimization for robust proton treatment planning. *Med Phys*. 2023;50:633-642. doi:10.1002/mp.15897
24. Wright SJ. *Primal-Dual Interior Point Methods*. Society for Industrial and Applied Mathematics; 1997.
25. Gorissen BL. Interior point methods can exploit structure of convex piecewise linear functions with application in radiation therapy. *SIAM J Optim*. 2022;32:256-275. doi:10.1137/21M1402364
26. Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge University Press; 2004.
27. Nocedal J, Wright SJ. *Numerical Optimization*. Springer-Verlag; 2006.
28. Poulsen PR, Eley J, Langner U, Simone ICB, Langen K. Efficient interplay effect mitigation for proton pencil beam scanning by spot-adapted layered repainting evenly spread out over the full breathing cycle. *Int J Radiat Oncol Biol Phys*. 2018;100:226-234.
29. Bertsimas D, King A, Mazumder R. Best subset selection via a modern optimization lens. *Ann Stat*. 2016;44:813-852. doi:10.1214/15-AOS1388
30. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc, B: Stat Methodol*. 1996;58:267-288.
31. Vidaurre D, Bielza C, Larrañaga P. A survey of L_1 regression. *Int Stat Rev*. 2013;81:361-381.
32. Hastie T, Tibshirani R, Wainwright M. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Monographs on Statistics and Applied Probability. 1st ed. Taylor and Francis; 2015.
33. Park MY, Hastie T. L_1-regularization path algorithm for generalized linear models. *J R Stat Soc, B: Stat Methodol*. 2007;69:659-677. doi:10.1111/j.1467-9868.2007.00607.x
34. Schmidt M, Fung G, Rosales R. Fast optimization methods for L_1 regularization: a comparative study and two new approaches. In: European Conference on Machine Learning (ECML). 2007:286-297. https://link.springer.com/chapter/10.1007/978-3-540-74958-5_28
35. Wu TT, Lange K. Coordinate descent algorithms for lasso penalized regression. *Ann Appl Stat*. 2008;2:224-244. doi:10.1214/07-AOAS147
36. Shi J, Yin W, Osher S, Sajda P. A fast hybrid algorithm for large-scale l_1-regularized logistic regression. *J Mach Learn Res*. 2010;11:713-741.
37. Simon N, Tibshirani R. Standardization and the group lasso. *Stat Sin*. 2012;22:983-1001. doi:10.5705/ss.2011.075
38. Jacob L, Obozinski G, Vert J-P. Group lasso with overlap and graph lasso. In: Proceedings of the 26th Annual International Conference on Machine Learning (ICML). 2009:433-440. doi:10.1145/1553374.1553431
39. Ma S, Song X, Huang J. Supervised group lasso with applications to microarray data analysis. *BMC Bioinform*. 2007;8:1-17. doi:10.1186/1471-2105-8-60

40. Rakotomamonjy A. Surveying and comparing simultaneous sparse approximation (or group-lasso) algorithms. *Signal Process*. 2011;91:1505-1526. doi:10.1016/j.sigpro.2011.01.012
41. Qin Z, Scheinberg K, Goldfarb D. Efficient block-coordinate descent algorithms for the group lasso. *Math Program Comput*. 2013;5:143-169. doi:10.1007/s12532-013-0051-x
42. Yang Y, Zou H. A fast unified algorithm for solving group-lasso penalized learning problems. *Stat Comput*. 2015;25:1129-1141. doi:10.1007/s11222-014-9498-5
43. Figueiredo M, Bioucas-Dias J, Nowak RD. Majorization–minimization algorithms for wavelet-based image restoration. *IEEE Trans Image Process*. 2007;16:2980-2991. doi:10.1109/TIP.2007.909318
44. Schifano ED, Strawderman RL, Wells MT. Majorization-minimizationalgorithms for nonsmoothly penalized objective functions. *Electron J Stat*. 2010;4:1258-1299. doi:10.1214/10-EJS582
45. Mairal J. Incremental majorization-minimization optimization with application to large-scale machine learning. *SIAM J Optim*. 2015;25:829-855. doi:10.1137/140957639
46. Sun Y, Babu P, Palomar DP. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Trans Signal Process*. 2017;65:794-816.
47. Lobo MS, Fazel M, Boyd S. Portfolio optimization with linear and fixed transaction costs. *Ann Oper Res*. 2007;152:341-365. doi:10.1007/s10479-006-0145-1
48. Fazel M. Matrix Rank Minimization with Applications. PhD thesis. Stanford University; 2002. https://faculty.washington.edu/mfazel/thesis-final.pdf
49. Fazel M, Hindi H, Boyd S. Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In: Proceedings of the American Control Conference (ACM). 2003:2156-2162. doi:10.1109/ACC.2003.1243393
50. Candés EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted l_1 minimization. *J Fourier Anal Appl*. 2008;14:877-905. doi:10.1007/s00041-008-9045-x
51. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045-1057.
52. Bejarano T, De Ornelas-Couto M, Mihaylov IB. Head-and-neck squamous cell carcinoma patients with CT taken during pre-treatment, mid-treatment, and post-treatment (HNSCC-3DCT-RT) (2018). [Data set]. The Cancer Imaging Archive. https://www.cancerimagingarchive.net/collection/hnscc-3dct-rt/
53. Wieser H-P, Cisternas E, Wahl N, et al,. Development of the open-source dose calculation and optimization toolkit matRad. *Med Phys*. 2017;44:2556-2568.
54. Wieser H, Wahl N, Gabryś H, et al,. MatRad – an open-source treatment planning toolkit for educational purposes. *Med Phys Int*. 2018;6:119-127.
55. Taasti VT, Hong L, Shim JS, Deasy JO, Zarepisheh M. Automating proton treatment planning with beam angle selection using bayesian optimization. *Med Phys*. 2020;47:3286-3296. doi:10.1002/mp.14215
56. Diamond S, Boyd S. CVXPY: a python-embedded modeling language for convex optimization. *J Mach Learn Res*. 2016;17:1-5.
57. Agrawal A, Verschueren R, Diamond S, Boyd S. A rewriting system for convex optimization problems. *J Control Decis*. 2018;5:42-60.
58. Diamond S, Agrawal A, Murray R, Stellato B, Boyd S. CVXPY: Disciplined Convex Programming in Python, Version 1.3. 2022. https://www.cvxpy.org/index.html
59. ApS M. The MOSEK Optimizer API for Python, Version 10.0. 2022. https://docs.mosek.com/latest/pythonapi/index.html and https://pypi.org/project/Mosek/
60. Petrova D, Smickovska S, Lazarevska E. Conformity index and homogeneity index of the postoperative whole breast radiotherapy. *Open Access Maced J Med Sci*. 2017;5:736-739.
61. Candes EJ, Romberg JK, Tao T. Stable signal recovery from incomplete and inaccurate measurements. *Commun Pure Appl Math*. 2006;59:1207-1223.
62. Alber M, Meedt G, Nüsslin F, Reemtsen R. On the degeneracy of the IMRT optimization problem. *Med Phys*. 2002;29:2584-2589.
63. Llacer J, Agazaryan N, Solberg TD, Promberger C. Degeneracy, frequency response and filtering in IMRT optimization. *Phys Med Biol*. 2004;49:2853-2880.
64. Zarepisheh M, Hong L, Zhou Y, et al,. Automated and clinically optimal treatment planning for cancer radiotherapy. *INFORMS J Appl Anal*. 2022;52:69-89.
65. Breedveld S, Storchi P, Heijmen B. The equivalence of multi-criteria methods for radiotherapy plan optimization. *Phys Med Biol*. 2009;54:7199-7209.
66. Taasti VT, Hong L, Deasy JO, Zarepisheh M. Automated proton treatment planning with robust optimization using constrained hierarchical optimization. *Med Phys*. 2020;47:2779-2790.
67. Craft D, Bortfeld T. How many plans are needed in an IMRT multi-objective plan database? *Phys Med Biol*. 2008;53:2785-2796.
68. Monz M, Küfer K-H, Bortfeld T, Thieke C. Pareto navigation–algorithmic foundation of interactive multi-criteria IMRT planning. *Phys Med Biol*. 2008;53:985-998.
69. Appenzoller LM, Michalski JM, Thorstad WL, Mutic S, Moore KL. Predicting dose-volume histograms for organs-at-risk in IMRT planning. *Med Phys*. 2012;39:7446-7461.
70. Shen C, Nguyen D, Zhou Z, Jiang SB, Dong B, Jia X. An introduction to deep learning in medical physics: advantages, potential, and challenges. *Phys Med Biol*. 2020;65:05TR01.
71. Fu A, Ungun B, Xing L, Boyd S. A convex optimization approach to radiation treatment planning with dose constraints. *Optim Eng*. 2019;20:277-300.
72. Zarepisheh M, Shakourifar M, Trigila G, et al,. A moment-based approach for DVH-guided radiotherapy treatment plan optimization. *Phys Med Biol*. 2013;58:1869-1887.
73. Mukherjee S, Hong L, Deasy JO, Zarepisheh M. Integrating soft and hard dose-volume constraints into hierarchical constrained IMRT optimization. *Med Phys*. 2020;47:414-421.
74. Taasti VT, Bäumer C, Dahlgren CV, et al,. Inter-centre variability of CT-Based stopping-power prediction in particle therapy: survey-based evaluation. *Phys Imaging Radiat*. 2018;6:25-30.
75. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys Med Biol*. 2012;57:R99-R117.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fu A, Taasti VT, Zarepisheh M. Simultaneous reduction of number of spots and energy layers in intensity modulated proton therapy for rapid spot scanning delivery. *Med Phys*. 2024;1-16. https://doi.org/10.1002/mp.17070