Energy layer optimization via energy matrix regularization for proton spot-scanning arc therapy

Gezhi Zhang[1], Haozheng Shen[1], Yuting Lin[2], Ronald C Chen[2], Yong Long[1*], and Hao Gao[2]

[1]University of Michigan-Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai, China

[2]Department of Radiation Oncology, University of Kansas Medical Center, Kansas City, USA

*Email: yong.long@sjtu.edu.cn

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mp.15836.

Abstract.

Purpose: Spot-scanning arc therapy (SPArc) is an emerging proton modality that can potentially offer a combination of advantages in plan quality and delivery efficiency, compared to traditional IMPT of a few beam angles. Unlike IMPT, frequent low-to-high energy layer switching (so called switch-up (SU)) can degrade delivery efficiency for SPArc. However, it is a tradeoff between the minimization of SU times and the optimization of plan quality. This work will consider the energy layer optimization (ELO) problem for SPArc and develop a new ELO method via energy matrix (EM) regularization to improve plan quality and delivery efficiency.

Methods: The major innovation of EM method for ELO is to design an energy matrix that encourages desirable energy-layer map with minimal SU during SPArc, and then incorporate this energy matrix into the SPArc treatment planning to simultaneously minimize the number of SU and optimize plan quality. The EM method is solved by the fast iterative shrinkage-

thresholding algorithm and validated in comparison with a state-of-the-art method, so-called energy sequencing (ES).

Results: EM is validated and compared with ES using representative clinical cases. In terms of delivery efficiency, EM had fewer SU than ES with an average of 35% reduction of SU. In terms of plan quality, compared to ES, EM had smaller optimization objective values and better target dose conformality, and generally lower dose to organs-at-risk and lower integral dose to body. In terms of computational efficiency, EM was substantially more efficient than ES by at least ten-fold.

Conclusion: We have developed a new ELO method for SPArc using energy matrix regularization, and shown that this new method EM can improve both delivery efficiency and plan quality, with substantially reduced computational time, compared to ES.

Key words: IMPT, SPArc, energy layer optimization.

1. Introduction

Proton arc therapy was developed in late 90s with passive scattering [1,2], which has recently revived into spot-scanning arc therapy (SPArc) by Ding et al [3] as an emerging proton modality that can potentially offer a combination of advantages in plan quality and delivery efficiency [4-7], compared to traditional IMPT of a few beam angles. This work concerns about the energy layer optimization (ELO) problem for SPArc.

The need of ELO for SPArc is motivated by the fact that the low-to-high energy layer switching (i.e., switch-up (SU)) time is much longer than the high-to-low energy layer switching (i.e., switch-down (SD)) time, e.g., 5s versus 0.5s. The SU is usually not an issue for IMPT of fixed beam angles, because (1) energy layers are delivered from high to low within each beam angle so that SU does not occur within an angle, and (2) the angle switching takes time anyway so that SU may not require additional time even it occurs between angles. However, the SU can be a limiting factor for delivery efficiency of SPArc, because ideally the beam can be constantly delivering while the gantry angle is continuously changing during SPArc. Without proper reduction of SU, the gantry may need to stop

frequently to wait for beam delivery, or beam delivery may need to be skipped for many arcs in order to wait for SU to complete before resuming beam delivery. Therefore, the minimization of SU is important for SPArc, in terms of both delivery efficiency and plan quality.

While Ding's method [3] was heuristic, Gu et al [8] proposed an ELO method with a rigorous problem formulation and method derivation, using so-called energy sequencing (ES) regularization to minimize SU while optimizing plan quality. However, in order for ES to function, ES imposes the optimization constraint "one energy layer per beam angle" (here each beam angle during SPArc optimization is after discretization and represents an angular segment or control point during SPArc delivery), which seems practically unnecessary and limits optimization degrees of freedom for plan quality. Because the optimization solution space with the constraint "one energy layer per beam angle" is a subset of the solution space without this constraint, the optimal plan quality under this constraint is usually worse than the plan quality that can be achieved with "multiple energy layers per beam angle" [8]. On the other hand, the ES regularization involves the updates of several ES matrices each iteration, which is computationally expensive. Inspired by SPArc [3] and ES [8], this work will

develop a new ELO method for SPArc using energy matrix (EM) regularization. Relative to ES, EM does not need to be restricted by the aforementioned optimization constraint so that more than one energy layer per beam angle is allowed in order to improve plan quality. It will be shown that EM can further reduce number of SU, improve plan quality, and enhance computational efficiency from ES.

2. Methods and Materials

2.1. Notations

Let B be the number of beam angles during SPArc plan optimization, and E_b the number of energy layers in the b^{th} angle. For the convenience of presentation, energy layer distribution for all beam angles (e.g., Fig. 1(a)) is called the energy map.

Note that the modeling of SPArc in many fixed beam angles here is a mathematical approximation or discretization of arc delivery, which does not mean that SPArc is delivered

only at those fixed beam angles. In the actual delivery, the gantry angle varies continuously, for which each beam angle in the presented model corresponds to an angular segment 2π/B or control point.

The vector x denotes intensities or the number of particles per spot for all spots from all angles, x_b all spots for the b^{th} angle, and x_{be} all spots with the e^{th} energy layer in the b^{th} angle. That is

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_b \\ \vdots \\ x_B \end{bmatrix}, x_b = \begin{bmatrix} x_{b1} \\ x_{b2} \\ \vdots \\ x_{be} \\ \vdots \\ x_{bE_b} \end{bmatrix}, x_{be} = \begin{bmatrix} x_{be1} \\ x_{be2} \\ \vdots \\ x_{ben} \\ \vdots \\ x_{beN_{be}} \end{bmatrix} \quad (1)$$

where N_{be} is the length of x_{be}, i.e., the number of spots in the e^{th} energy layer of the b^{th} angle.

To group the spots per energy layer, we introduce the scalar y_{be} as the sum of all elements in x_{be}, the vector y_b as the concatenation of y_{be} corresponding to the b^{th} angle, and y that consists of all y_b's from all angles. That is

$$y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_b \\ \vdots \\ y_B \end{bmatrix}, y_b=\begin{bmatrix} y_{b1} \\ y_{b2} \\ \vdots \\ y_{be} \\ \vdots \\ y_{bE_b} \end{bmatrix}, y_{be}=\sum_{i=1}^{N_{be}} x_{bei} \quad (2)$$

In matrix representation, $y=Wx$, where W is a summation matrix that sums up x per energy layer.

2.2. Energy Layer Optimization

We start the formulation of ELO by giving the dose fidelity term

$$\mathrm{f}(x)=\sum_{S\in\Omega} w_s \left\| A_s x - d_s \right\|_2^2, \quad (3)$$

where Ω is a set of different structures including the treatment planning target PTV and organs-at-risk (OAR), d_s is the constraint dose, A_s is the dose influence matrix, and w_s is the objective weight, corresponding to the s^{th} structure. Note that Ω depends on x and needs to be updated during plan optimization iterations, for dose-volume constraints [9-11]: for

example, for dose-volume constraint $D_{50\%}$<30Gy, if the volume of ≥30Gy is 60% (which is more than 50%), Ω for this constraint will include the 10% dose-violating voxels of the smallest dose for dose minimization.

To benchmark and compare with the ES method, the EM method adopts the similar ELO formulation to ES [8]

$$\begin{aligned} &\min_x \mathrm{F}(x) = \mathrm{f}(x) + \mathrm{GS}(x) + \mathrm{LOG}(y) + \mathrm{EMR}(y) \\ &\text{subject to } x \geq 0, y = Wx \end{aligned} \tag{4}$$

where

$$\begin{aligned} &\mathrm{f}(x) = \left\| Ax - d \right\|_2^2, \\ &\mathrm{GS}(x) = \alpha \sum_{b=1}^{B} \sum_{e=1}^{E_b} w_{be} \left\| x_{be} \right\|_2, \\ &\mathrm{LOG}(y) = -\beta \sum_{b=1}^{B} \log\left(\sum_{e=1}^{E_b} y_{be} \right), \\ &\mathrm{EMR}(y) = \gamma \left\| M\mathcal{S} \right. \end{aligned} \tag{5}$$

The first term f is dose fidelity term Eq. (3), where, without loss of generality, Eq. (3) is denoted by a simplified version in Eq. (5). The second term GS is group-sparsity regularization to minimize the number of energy layers with energy-layer-dependent

weighting w_{be} [12]. The third term LOG uses a log barrier function to promote that each angle has at least one energy layer, since the GS alone may generate the energy map with angles of empty energy layers [8]. The fourth term is the energy matrix regularization (EMR) with the energy matrix M.

The difference between ES and EM methods is that the ES regularization is replaced by the EMR in Eq. (4). In the EMR term Eq. (5), S is a sigmoid function that binarizes y to capture energy SU for the purpose of energy-matrix regularization, i.e., $S(y)=2/(1+e^{-y})-1$.

Note that the necessity of L2 norm (i.e., sum of least squares) in EMR can be explained by comparing Fig. 1(g) and (h). The EM method can minimize the number of SU to improve the delivery efficiency of SPArc, using the properly designed energy matrix, i.e., M in Eq. (5), that will be explained next.

2.3. Energy Matrix Regularization: Principle

This section discusses the principles for designing the energy matrix (P1-P4), and the next section will provide the algorithm for generating the energy matrix.

To mimic the delivery of SPArc, energy layers per beam angle in the energy map (e.g., Fig. 1(a)) are sorted in the descending order of proton energies, i.e., the energy layers are delivered always from high to low energies (e.g., for the 2nd angle A2 in Fig. 1(c)), while the energy SU can occur at the transition of beam angles (e.g., from the 1st angle A1 to the 2nd angle A2 in Fig. 1(d)). Since the energy switch-up (from low to high) takes much longer than the switch-down (from high to low), smaller number of SU often implies faster delivery of SPArc.

For minimizing the number of SU and at the same time optimizing the plan quality, the energy matrix M is designed in such a way with

- (P1) the selection of high-energy layers to minimize SU for the purpose of efficient plan delivery (e.g., by comparing Fig. 1(c) and (d));
- (P2) the selection of high-energy layers to have more degrees of freedom for optimizing plan quality without increasing SU (e.g., by comparing Fig. 1(e) and (f));

- (P3) the penalization of the selection of only high-energy layers to have diversified energy layers for the purpose of optimizing plan quality (e.g., by comparing Fig. 1(g) and (h));
- (P4) the introduction of barrier constraints (e.g., Fig. 2(c)) to minimize SU (e.g., by comparing Fig. 2(d) and (e)).

First, we will illustrate the use of M matrix via two examples (Fig. 1 and 2): P1-P3 will be explained via Fig.1, and P4 will be explained via Fig. 2. Then we will provide the algorithm (Fig. 3) to determine M.

Because the number of SU is determined by *"which energy layers to use"* rather than specific spot values per energy layer, the M matrix concerns about the binarized y: $y_{be}=1$ when the e^{th} energy layer of the b^{th} beam angle is used (i.e., with at least one nonzero spot to be delivered); $y_{be}=0$ otherwise. Throughout this section (including Fig. 1 and 2), EM refers to the value of EMR term in Eq. (5) acting on the binarized y without γ, i.e., $EM=\|My\|^2$ and strictly speaking y corresponds to $S(y)$ after the sigmoid transform.

An example of energy matrix M is provided in Fig. 1(b), which corresponds to the energy map in Fig. 1(a). The optimization degrees of freedom in this example consist of three beam angles (A1-A3), each of which has three energy layers (140MeV, 150MeV, 160MeV). The beam angles are sorted in the order of actual delivery, and the energy layers are arranged in the order of descending energy values.

To illustrate P1, let us consider Fig. 1(c) and Fig. 1(d), where active energy layers (i.e., $y_{be}=1$) are denoted by red and green dots. In terms of SU, since there is no SU in Fig. 1(c), SU=0; since there is a SU from A1 to A2 in Fig. 1(d), SU=1. In terms of EMR, in Fig. 1(c), y=[1 0 0 1 1 0 0 0 1]^T and thus EM=17; in Fig. 1(d), y=[0 0 1 1 1 0 0 0 1]^T and thus EM=41. Therefore, the energy matrix M encourages the selection of high-energy layers (i.e., 160MeV instead of 140MeV for A1 with smaller EM=17 than EM=41) for minimizing SU (i.e., SU=0 instead of SU=1).

To illustrate P2, let us consider Fig. 1(e) and Fig. 1(f). In terms of SU, when 150MeV is active in A2 (Fig. 1(e)), either 150MeV or 140MeV can be active in A3 without increasing SU; when 140MeV is active in A2 (Fig. 1(f)), only 140MeV can be active in A3 without increasing

SU. In terms of EMR, in Fig. 1(e), y=[1 0 0 1 1 0 0 1 0]^T and EM=20 if 150MeV is active in A3, while y=[1 0 0 1 1 0 0 0 1]^T and EM=17 if 140MeV is active in A3; in Fig. 1(f), y=[1 0 0 1 0 1 0 1 0]^T and EM=17 if 150MeV is active in A3, while y=[1 0 0 1 0 1 0 1 0]^T and EM=40 if 140MeV is active in A3. Therefore, M encourages the selection of high-energy layers (i.e., 150MeV instead of 140MeV for A2 in Fig. 1(e)) to have more rooms for optimizing plan quality without increasing SU (i.e., A3 can be filled with either 150MeV or 140MeV without increasing SU in Fig. 1(e), while only 140MeV for A3 does not increase SU for in Fig. 1(f)).

To illustrate P3, let us consider Fig. 1(g) and Fig. 1(h). In terms of EMR, in Fig. 1(g), y=[1 0 0 1 0 0 1 0 0]^T and thus EM=9; in Fig. 1(h), y=[1 0 0 1 0 0 0 1 0]^T and thus EM=8. Therefore, M penalizes the selection of only high-energy layers (i.e., 160MeV for all angles A1-A3 with larger EM in Fig. 1(g)) to have diversified energy layers for the purpose of optimizing plan quality.

Note that the energy matrix in Fig. 1(b) is local in the sense its matrix value only depends on the local ordering of energy layers per beam angle and is independent of the global ordering energy layers from all beam angles. This locality can be problematic, when the

energy map has different energy-layer distribution per beam angle (e.g., Fig. 2(a)). To incorporate the globality (in the ordering energy layers) into energy matrix, barrier constraints are introduced and explained next.

To illustrate P4 on barrier constraints, let us consider Fig. 2. The optimization degrees of freedom (Fig. 2(a)) in this example consist of energy layers (150MeV, 140MeV, 130MeV) for A1, energy layers (170MeV, 160MeV, 150MeV) for A2, and energy layers (160MeV, 150MeV, 140MeV) for A3. The energy matrix M1 (Fig. 2(b)) is similar to the previous one in Fig. 1: M1 is local to each beam angle in the sense that the M1 matrix value (i.e., 1, 2, 3) corresponds to the descending order of energies per angle. On the other hand, the energy matrix M2 (Fig. 2(c)) has the barrier constraints, i.e., the matrix values of M2 equal to a much larger number 30 (here 30 is the barrier constraint value, which could have been set to a different value rather than 30, as long as it is sufficiently larger than the rest of values, e.g., by ten-fold) at 170MeV and 160MeV for A2 and 160MeV for A3; as a result, M2 is non-local in the sense that the M2 matrix value also accounts for the descending order of energies for all angles. In terms of SU, since there is a SU from A1 to A2 in Fig. 2(d), SU=1; since there is no SU in Fig. 2(e), SU=0. In terms of EMR, in Fig. 2(d), y= [0 0 1 0 0 1 0 0 0 0 0 1 0 0 0]^T

(there are 15 elements because there are 5 different energy layers 170MeV, 160MeV, 150MeV, 140MeV and 130MeV, and 3 angles; for each angle, 1 implies the energy available for optimization and 0 otherwise) and thus EM1=3, EM2=1801; in Fig. 2(e), y= [0 0 1 0 0 0 0 1 0 0 0 0 0 1 0]^T and thus EM1=25, EM2=8. Therefore, without barrier constraint, M1 falsely chooses SU=1 over SU=0, because SU=1 has smaller EM1; with barrier constraints, M2 correctly chooses SU=0 over SU=1, because SU=0 has smaller EM2. The purpose of barrier constraint is to de-localize the ordering per angle by giving a sufficiently large weighting (e.g., 30 in this example), so that the EMR correctly penalizes the SU.

2.4. Energy Matrix Regularization: Algorithm

In this section, we will provide the algorithm to determine M for general energy map, which is illustrated in Fig. 3. The algorithm consists of two major steps: (S1) to determine the location of nonzero matrix elements of M; (S2) to calculate the values for these nonzero matrix elements.

Regarding S1, the columns of *M* are arranged for angles in the order of beam scanning, for energy layers per angle in the order of descending energy values. To determine the rows of *M*, first, the leading energy layer (with the largest energy) per beam angle is located (e.g., orange and black circles in Fig. 3). Then, the largest leading energy layer is identified (e.g., the black circle in Fig. 3) and the sub-matrix (e.g., A2 that starts from the first row of *M* in Fig. 3) for this angle (the so-called benchmark angle) is determined. Next, for each of the remaining beam angles (e.g., A1, A3, A4 in Fig. 3), the leading energy layer is aligned to the benchmark angle (e.g., the sub-matrix for A1 starting from the row in A2 of the same energy in Fig. 3); within each sub-matrix per beam angle, the nonzero elements are consecutively placed on the diagonal, starting from the first nonzero element that is already aligned to the benchmark angle. The process of S1 can also be illustrated using Fig. 2: first because A2 has the largest leading energy layer, which is 170MeV, A2 is the benchmark angle; then A1 and A3 are aligned to A2 to form the EM in Fig. 2(c), i.e., to position the leading energy layer 150MeV of A1 in the same row as 150MeV of A2, and i.e., to position the leading energy layer 160MeV of A3 in the same row as 160MeV of A2. Note that the process of S1 is independent of the delivery order of beam angles. For example, since the benchmark angle

is the angle with the largest leading energy layer, the benchmark angle remains the same even though the starting angle during SPArc delivery is changed.

Regarding S2, first, to remove the locality of energy matrix, we identify the elements for barrier constraints (i.e., with substantially larger values for the penalization purpose), which are the energy layers with larger energy (e.g., the black circle in Fig. 3) than the leading energy layer of A1. Then, we determine the rest of diagonal matrix values for each sub-matrix (without barrier constraint) based on the descending energy order (larger matrix value for smaller energy) via the following Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right), \qquad x \le \mu . \tag{5}$$

Here x is the local ordering of energy layers per angle (after Barrier constraints are assigned), e.g., x=1 for the energy layer of the largest energy; μ is set to be the number of energy layers (without barrier constraints) per angle; the barrier constant is set to be 10 times of the largest Gaussian values of all angles; σ=5 in this study.

2.5. Optimization Algorithm

Since ES was solved by the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [13], the optimization algorithm for solving EM here is also based on FISTA, for the purpose of fair comparison with ES, although EM or ES can also be efficiently solved by other methods such as Alternating Direction Method of Multipliers (ADMM) [14,15].

To solve Eq. (4) using FISTA, the optimization problem Eq. (4) is re-formulated as

$$\min_x f(x) + g(x), \tag{6}$$

where

$$\begin{aligned} f(x) &= \| Ax - d \|_2^2 - \beta W_b^T L(W_e x) + \gamma \| MS(Wx) \|_2^2 \\ g(x) &= \alpha \sum_{b=1}^{B} \sum_{e=1}^{E_b} w_{be} \left\| x_{be} \right\|_2 + I_{\geq 0}(x) \end{aligned}. \tag{7}$$

Here W is the summation matrix that sums up x per energy layer, i.e., $Wx=\sum_i x_{bei}$; W_e is the summation matrix that sums up x from all energy layers per beam angle, i.e., $W_e x=\sum_{ei} x_{bei}$; W_b is the summation vector that sums up x from all beam angles, i.e., $W_b^T x=\sum_b x_b$; L and S are element-wise functions, $L(u) = [l(u_1), \dots, l(u_n), \dots, l(u_B)]^T$ with $l(x) = -log(x)$, and

$S(u) = \left[s(u_1), \ldots, s(u_n), \ldots, s\left(u_{N_{be}}\right)\right]^T$, and $s(x) = \frac{2}{1+e^{-x}} - 1$. $I_{\geq 0}$ is an indicator function defined as:

$$I_{\geq 0}(x) = \begin{cases} x & \text{if } x \geq 0 \\ \infty & \text{otherwise} \end{cases} \tag{8}$$

Then the gradient of f is given by

$$\nabla f(x) = 2A^{\mathrm{T}}\left(Ax - d\right) - \beta L'(W_e x)W_e^T + 2\gamma[MS(Wx)]^T[MS'(Wx)]W^T \tag{9}$$

Here *L'* and *S'* are derivatives for *L* and *S* respectively, with $l'(x) = -\frac{1}{x}$, and $s'(x) = \frac{2e^{-x}}{(1+e^{-x})^2}$; W^T is the transpose of *W*, i.e., $x_{bei}=W^T y_{be}$, where the value of y_{be} is projected to all spots on this energy layer; W_e^T is the transpose of W_e, i.e., $x_{bei}=W_e^T y_b$, where the value of y_b is projected to all spots on this beam angle.

The proximal operator of *g(x)* has analytical formula [16]

$$\text{prox}_{tg}(x) = \text{prox}_{t\|\cdot\|_2}(\max(x,0)) = x - x \cdot \min\left(\frac{t}{\|\max(x,0)\|_2}, 1\right) \tag{10}$$

The FISTA algorithm for solving EM is summarized in the following. The parameters α, β, and γ are manually chosen case by case. Their values are determined based on the rule

of thumb to minimize the number of SU while the plan quality is still preserved. For fair comparison, the values of α, β, and s are the same between ES and EM.

Algorithm: FISTA for solving EM

Initialize: $M, \alpha, \beta, \gamma, \mathrm{x}_0, \mathrm{t}_0, \mathrm{s}, \mathrm{N}$

for k = 1, 2, …, N

$$\nabla f(x_{k-1}) = 2A^{\mathrm{T}}\left(Ax_{k-1} - d\right) - \beta L'(W_e x_{k-1})W_e^T + 2\gamma[MS(Wx_{k-1})]^T[MS'(Wx_{k-1})]W^T$$

$$y = \mathrm{prox}_{s\|\|_1}(x_{k-1} - s\nabla f(x_{k-1}))$$

$$t_k = \frac{1+\sqrt{1+4t_{k-1}^2}}{2}$$

$$x_k = y + \left(\frac{t_{k-1}-1}{t_k}\right)(y - x_{k-1})$$

end for

2.6. Materials

The new method ("EM") was validated in comparison with a state-of-the-art method ("ES"), using Abdomen (2.2Gy x 25 fractions), Brain (2Gy x 10 fractions), Head-and-neck (HN) (2.12Gy x 33 fractions) and Lung (2Gy x 30 fractions) cases. For optimizing plan quality, first

we solved ES and EM (Eq.(4)) with all regularization terms for selecting energy layers; then we solved Eq. (4) with the dose-fidelity term only on selected energy layers to get the treatment plan. For simplicity, treatment planning was with respect to PTV, which is sufficient for the purpose of this work. For fair comparison, both EM and ES were solved by FISTA and plans were normalized after optimization with D95=100% at PTV. To simulate SPArc, we divided the full 360º arc into 72 equally segmented beam intervals of at 5º, and each interval was approximated by the beam angle at its center. Then the dose influence matrix A was generated via MatRad [17] using 5 mm lateral spacing, and 3 mm longitudinal spacing for energy discretization, on 3 mm-resolution dose grid.

In Table 1, the conformity index (CI) is defined as $V_{100}^2/(V \times V'_{100})$ (V_{100}: PTV volume receiving at least 100% of prescription dose; V: PTV volume; V'_{100}: total volume receiving at least 100% of prescription dose). The value of CI is between 0 and 1, with its optimal value being 1. The homogeneity index (HI) [23] is defined as $(D_2 - D_{98})/D_p \times 100\%$ (D_2: the min dose for 2% volume receiving the highest dose; D_{98}: the min dose for 98% volume receiving the highest dose; D_p: the prescription dose); smaller value of HI corresponds to a more homogeneous target dose.

The total plan delivery time T is estimated using a simplified beam delivery model that only accounts for energy switching time T_E and spot delivery time T_B. In Table 2, T_E is calculated assuming 5.5 seconds (s) for each SU and 0.6 s for each SD, and T_B is estimated by $T_B=//x//_1/\gamma$, with the dose rate $\gamma=2.6\times10^{10}$ protons/mins (this dose rate corresponds to the minimum MU threshold 1.1×10^6 protons for the Varian ProBeam system).

3. Results

3.1. EM had fewer SU than ES

The energy-layer delivery trajectories optimized with ES and EM are presented in Fig. 4, which shows that EM had fewer SU than ES. Specifically, compared to ES, EM reduced SU from 36 to 13 for Abdomen, from 31 to 26 for Brain, from 35 to 24 for HN, from 32 to 24 for Lung, with an average of 35% reduction of SU. Also note that ES was strictly one energy layer per angle, while EM removed such striction and allowed more than one energy layers per angle. This flexibility in the number of energy layers per angle can allow for the improvement in plan quality, which will be presented in the next section.

3.2. EM had better plan quality than ES

For plan quality comparison, dosimetric parameters are summarized in Table 1, dose volume histogram (DVH) plots and dose plots are presented in Fig. 6 and 7 respectively. As shown in Table 1, EM had smaller optimization objective values than ES, which indicates that EM had better plan quality than ES. Moreover, EM provided better target dose conformality, indicated by larger CI values, e.g., an increase from 0.53 to 0.84 for Brain in Table 1; EM provided better target dose homogeneity, indicated by smaller HI values, e.g., a decrease from 15.14 to 9.34 for lung in Table 1; EM had generally lower dose to OAR and lower integral dose to body. The improved plan quality via EM is also demonstrated by DVH plots in Fig. 6 and dose plots in Fig. 7: EM had steeper target DVH curve and lower OAR DVH curve in Fig. 6, and tighter 80% isodose line in Fig. 7.

3.3. EM was more efficient than ES

As shown in Table 2, EM had shorter energy switching time T_E than ES, due to reduced number of SU, despite of increased number of SD. On the other hand, EM also had shorter spot delivery time T_B than ES. As a result, EM had shorter total plan delivery time T and thus was more efficient than ES. Specifically, compared to ES, EM reduced total plan delivery time T from 307s to 140s (54% reduction) for Abdomen, from 252s to 217s (14% reduction) for Brain, from 460s to 359s (22% reduction) for HN, from 415s to 339s (18% reduction) for Lung, with an average of 23% reduction of total delivery time.

3.4. EM was more computationally efficient than ES

Regarding the number of iterations, ES converged in 1000-1500 iterations to meet the one-energy-layer-per-beam requirement, while EM converged in 500 iterations or fewer (e.g., Fig. 7). Regarding computational cost per iteration, one EM iteration was one tenth of one ES iteration. The average run time of 500 ES iterations took about 6000 seconds, while that

of 500 EM iterations took about 600 seconds, on a desktop computer with i5-11400 6-core 12-thread CPU. Therefore, EM was more computationally efficient than ES by at least ten-fold.

4. Discussion

It is an intrinsic tradeoff between plan quality and delivery efficiency, even for SPArc, regardless of ES or EM. However, as demonstrated in the Result section, compared to ES, EM improved both delivery efficiency (with reduced number of SU) and plan quality.

Therefore, EM has a better tradeoff than ES, i.e., lower Pareto surface in terms of multi-criteria optimization.

In term of algorithm difference for SU reduction, ES requires the calculation of several regularization matrices each FISTA iteration, which substantially slows down the ES method; on the other hand, ES has to impose the condition “one energy layer per angle” in order to for ES to function. In contrast, the M matrix in EM is computed only once, and thus its computational cost is negligible overall; EM does not need to be constrained by the ES condition and can have more than one energy layer per angle, and the removal of the constraint allows EM to have more optimization degrees of freedom to improve plan quality. On the other hand, there is a flaw for EM that the M matrix can have unnecessary penalization to some energy layers. For example, considering the scenario that SU occurs at the current angle (i.e., the leading energy layer at this angle has larger energy than the smallest energy layer of the previous angle), although the selection of other energy layers at this angle does not further increase SU with respect to the previous angle, the M matrix still penalizes the selection of remaining energy layers, which is unnecessary. However, this flaw seems practically minor, as EM still outperformed ES in plan quality.

We have tried both ADMM and FISTA and found no major difference in results. However, since FISTA was used in the original ES paper [8], we also adopt FISTA here. On the other hand, we found the difference between ½ norm and L1 norm for GS is negligible; given the L1 norm is more common for GS, we also use L1-norm based GS term in this work.

Without complicating the matter of this work on ELO, we have not considered the minimum monitor-unit (MMU) constraint, i.e., MMU optimization problem, that may be needed for SPArc plans to be deliverable. It should be technically straightforward to add MMU to ELO for either ES and EM, e.g., to incorporate our previous MMU methods [18-21]. However, MMU-based SPArc treatment planning can be practically challenging, as the MMU constraint is nonconvex and more restrictive for SPArc than IMPT. That is, since SPArc has more optimization degrees of freedom and thus smaller monitor units per spot, the MMU constraint is harder to meet for SPArc, for which a recently developed MMU method [22] for relatively large MMU threshold may help.

6. Conclusion

We have developed a new ELO method for SPArc using energy matrix regularization, and shown that this new method EM can improve both delivery efficiency and plan quality, with substantially enhanced computational efficiency, compared to a state-of-the-art method ES.

Acknowledgment

The authors are very thankful to the valuable comments from anonymous reviewers. This research is partially supported by NIH grants No. R37CA250921, R01CA261964, and a KUCC physicist-scientist recruiting grant.

COI Statement

The authors have no conflicts to disclose.

Figure 1. P1-P3 principles behind energy matrix regularization. (a) Energy map of 3 beam angles A1-A3; (b) energy matrix *M* corresponding to this energy map. P1 is demonstrated by comparing (c) and (d): the energy matrix encourages the selection of high-energy layers (i.e., 160MeV instead of 140MeV for A1 with smaller EM in (c)) for minimizing SU (i.e., smaller SU in (c)). P2 is illustrated by comparing (e) and (f): *M* encourages the selection of high-energy layers (i.e., 150MeV instead of 140MeV for A2 in (e)) to have more degrees of freedom for optimizing plan quality without increasing SU (i.e., A3 can be filled with either 150MeV or 140MeV without increasing SU in (e), while only 140MeV for A3 does not increase SU for in (f)). P3 is explained by comparing (g) and (h): *M* penalizes the selection of only high-energy layers

(i.e., 160MeV for all angles A1-A3 with larger EM in (g)) to have diversified energy layers for the purpose of optimizing plan quality.

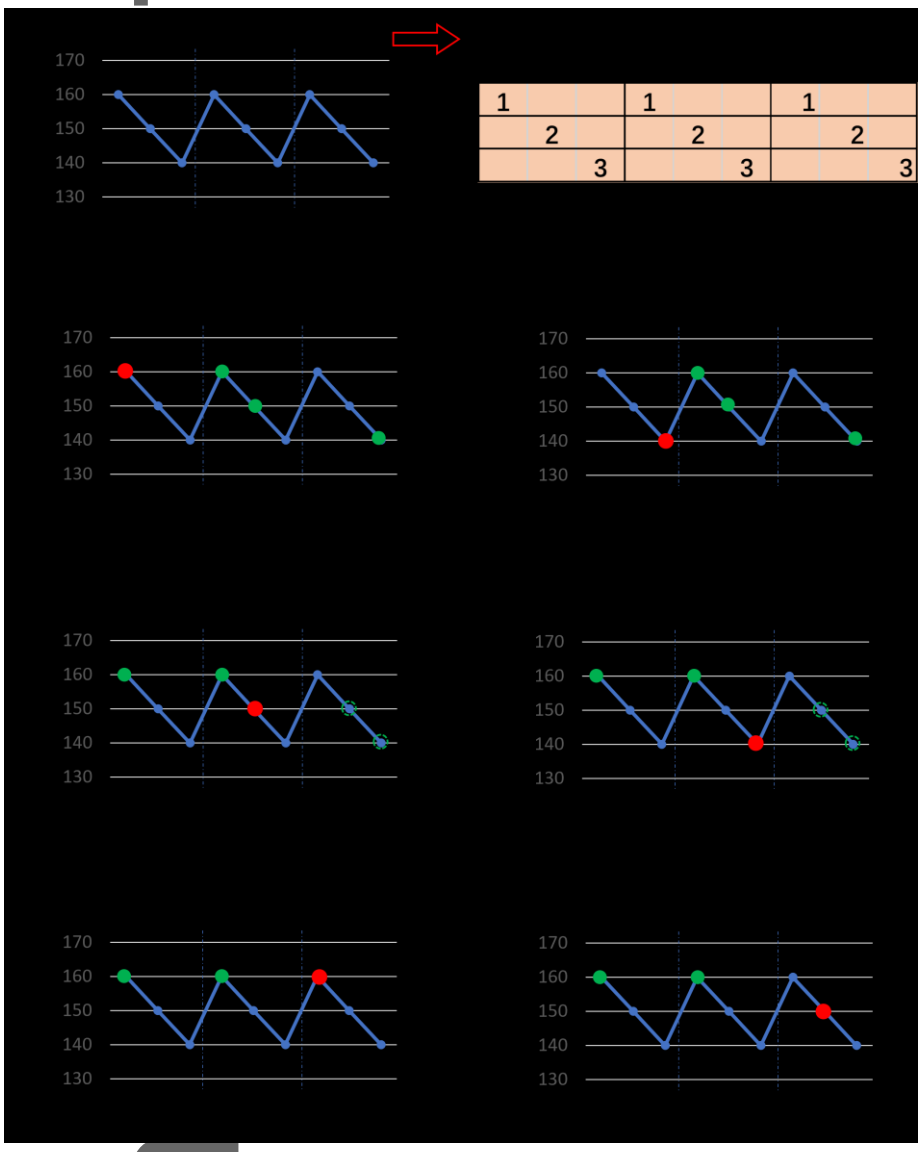


Figure 2. P4 principle behind energy matrix regularization. (a) Energy map. (b) Energy matrix M1 (without barrier constraint); M1 is local to each beam angle in the sense that the M1 matrix value (i.e., 1, 2, 3) corresponds to the descending order of energies per angle. (c) Energy matrix M2 (with barrier constraints); owing to the barrier constraints (i.e., the matrix values of M2 equal to a much larger number 30 at 170MeV and 160MeV for A2 and 160MeV for A3), M2 is non-local in the sense that the M2 matrix value is also determined by the descending order of energies for all angles. P4 is demonstrated by comparing (d) and (e):

P4 introduces barrier constraints to minimize SU; without barrier constraint, M1 falsely chooses SU=1 (d) over SU=0 (e), because SU=1 has smaller EM1; with barrier constraints, M2 correctly chooses SU=0 (e) over SU=1 (d), because SU=0 has smaller EM2. Note that the barrier constraint value 30 could have been set to a different value rather than 30, as long as it is sufficiently larger than the rest of matrix values.

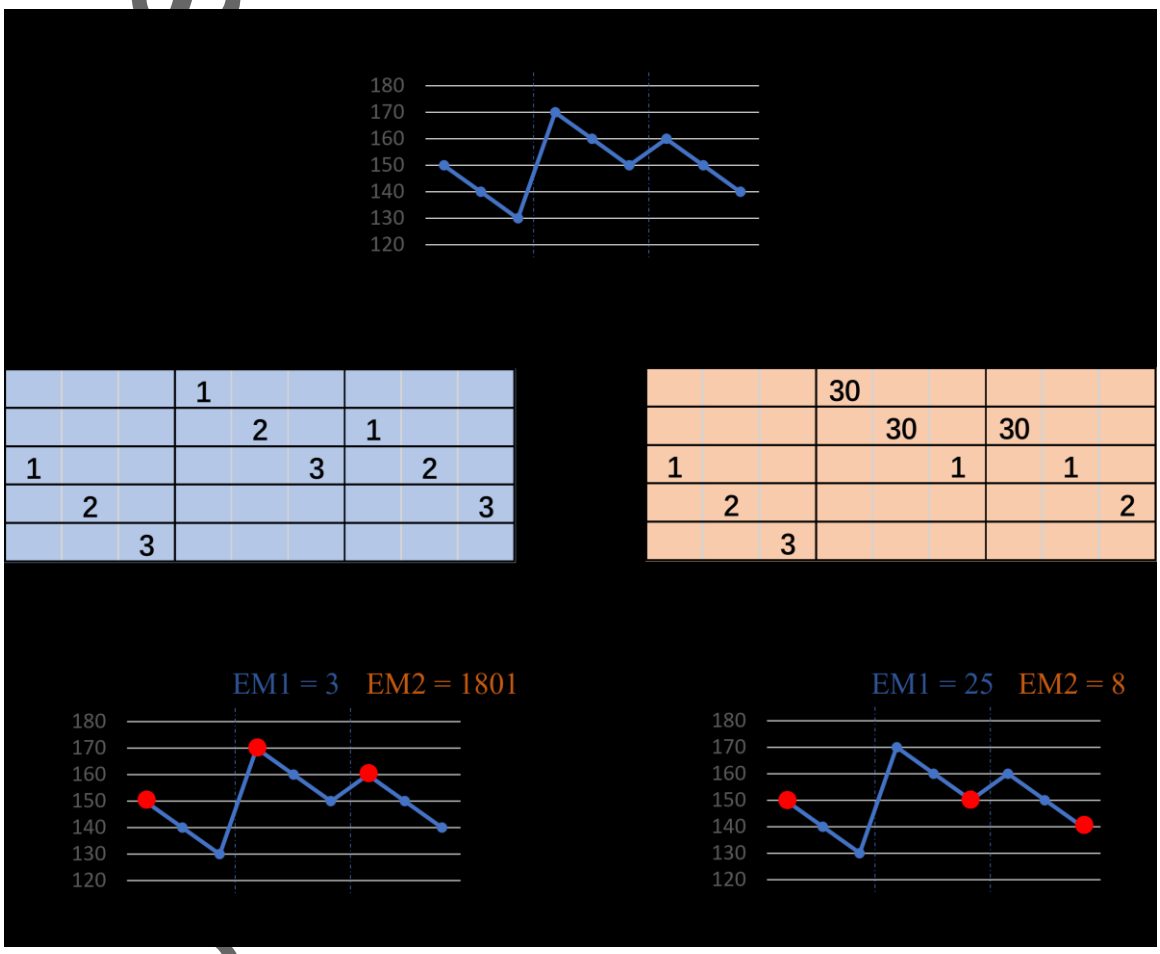


Figure 3. Generation of energy matrix. First, the leading energy layer (with the largest energy) per beam angle is located (orange and black circles in the figure). Second, the largest leading energy layer is identified (the black circle) and the sub-matrix (i.e., A2 that starts from the first row of *M*) for this angle (the so-called benchmark angle) is determined. Third, for each of the rest of beam angles (A1, A3, A4), the leading energy layer is aligned to A2 (e.g., the sub-matrix for A1 starting from the row in A2 of the same energy). Fourth, the barrier constraints (with substantially larger values for the penalization purpose) are assigned for the layers with larger energy (the black circle) than the leading energy layer of A1. Last, diagonal matrix values for each sub-matrix (without barrier constraint) are calculated in the descending energy order (larger matrix value for smaller energy) via the Gaussian distribution.

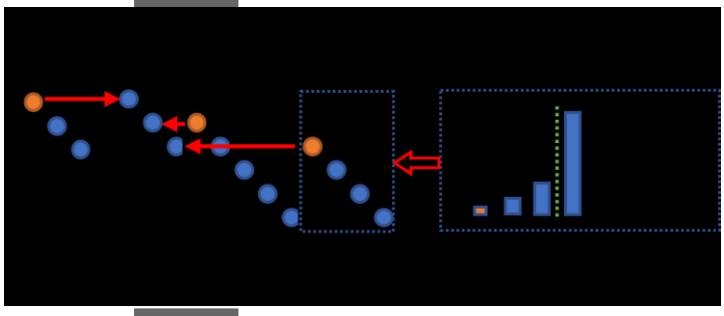


Figure 4. Comparison of energy-layer delivery trajectories between ES (blue) and EM (red).

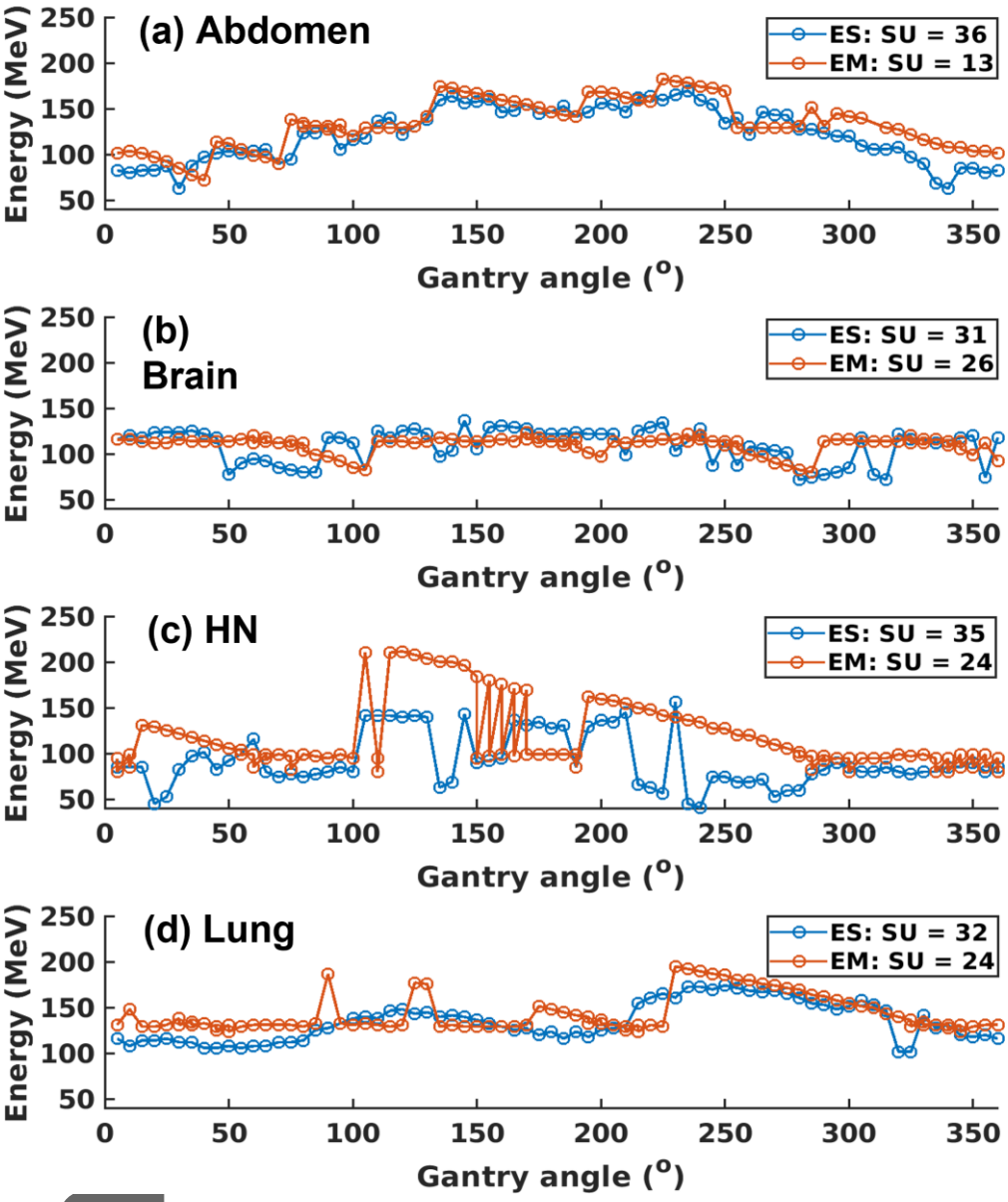


Figure 5. DVH plots. Blue solid line: PTV of ES; blue dashed line: OAR of ES; red solid line: PTV of EM; red dashed line: OAR of EM.

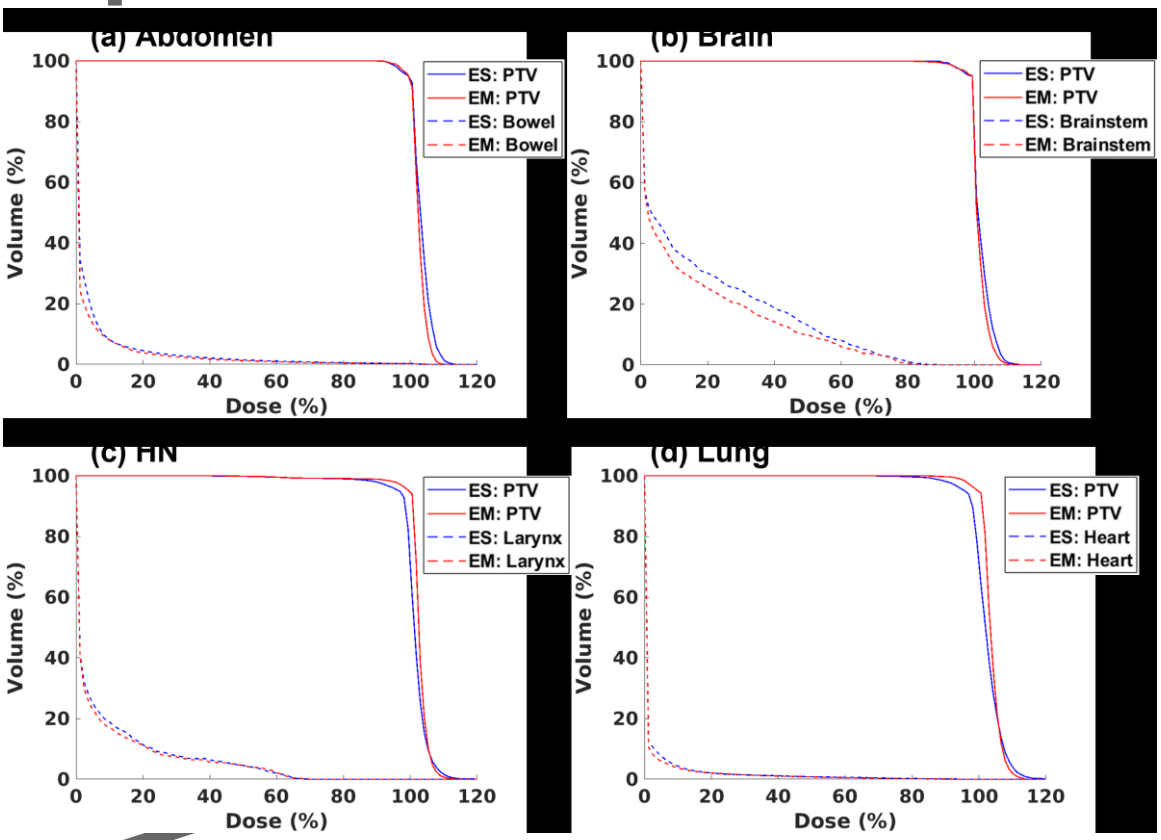


Figure 6. Dose plots. (a-d): ES; (e-h): EM. (a, e): abdomen; (b, f) Brain; (c, g): HN; (d, h): lung. The dose plot window is [0%, 110%]. PTV, 100% isodose line and 80% isodose line are highlighted.

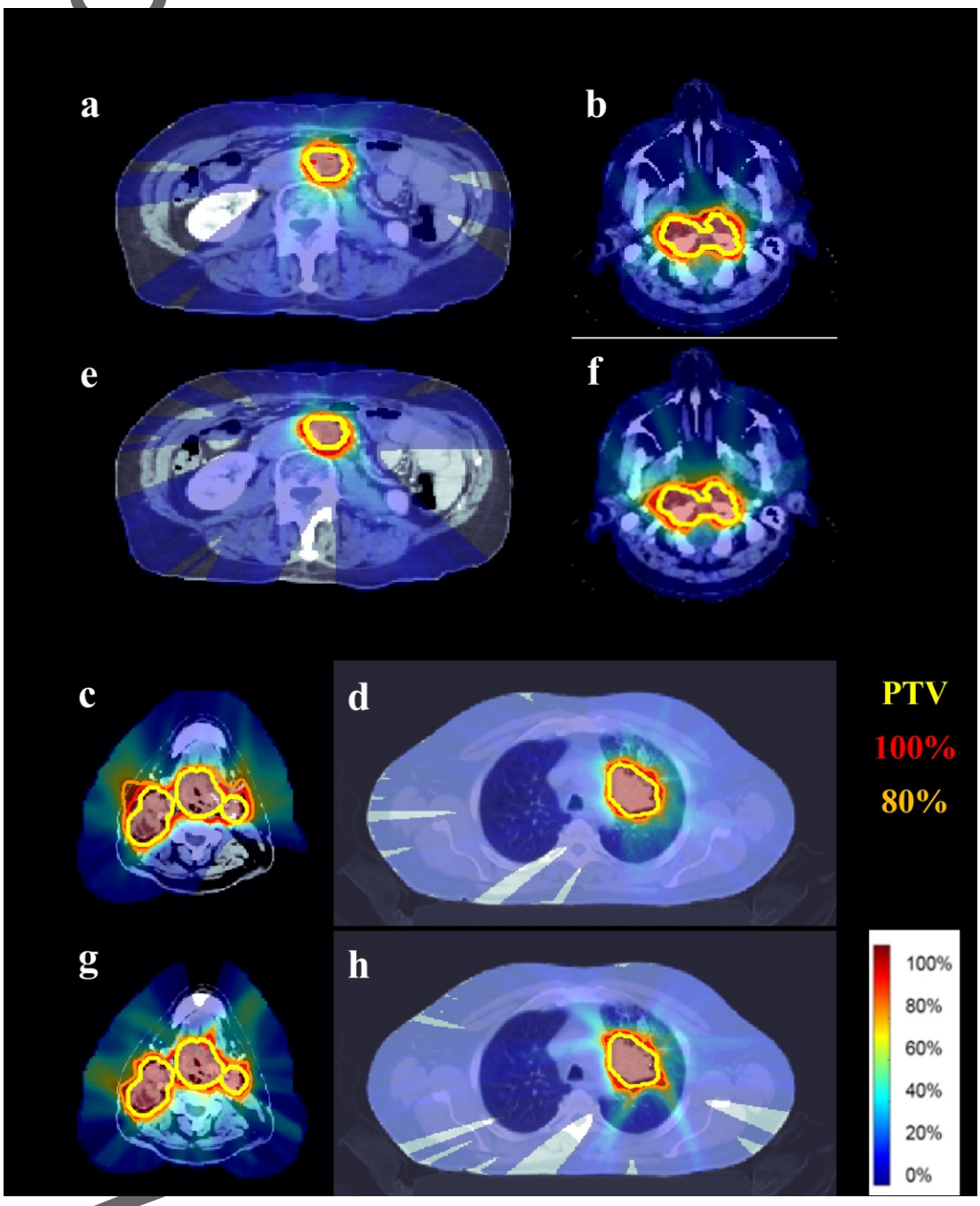
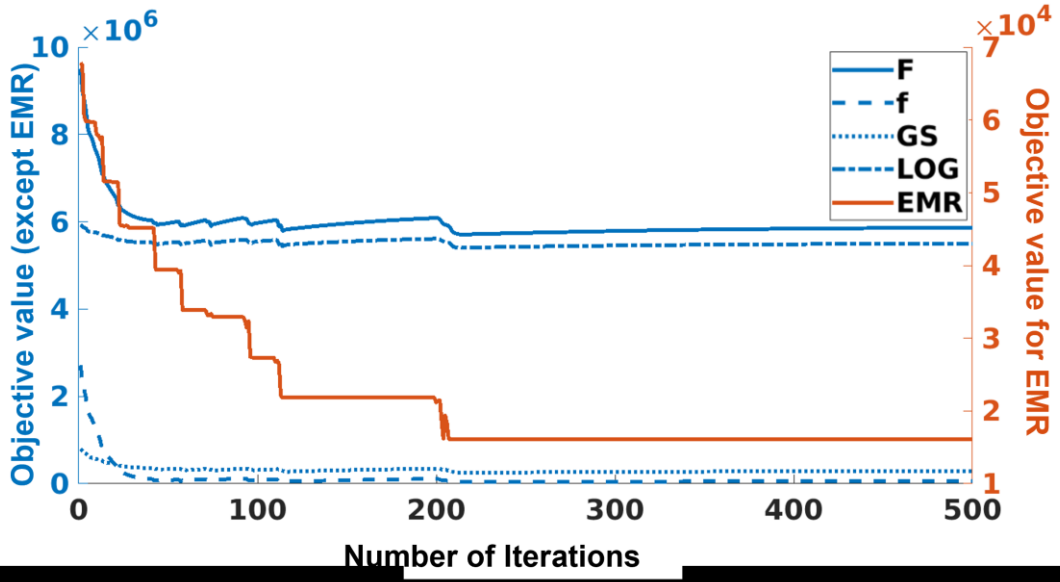


Figure 7. Convergence plot of EM for the Brain case. F: total objective; f: dose fidelity; GS: group sparsity term; LOG: log barrier term; EMR: energy matrix term. Note that the y-axis for EMR is different from the rest terms.



References

[1] Deasy J, Mackie T, DeLuca P. Method and apparatus for proton therapy. 1997.

[2] Sandison GA, Papiez E, Block C, et al., 1997. Phantom assessment of lung dose from proton arc therapy. Int J Radiat Oncol Biol Phys. 38, 891–897.

[3] Ding X, Li X, Zhang JM, et al., 2016. Spot-Scanning Proton Arc (SPArc) Therapy: The First Robust and Delivery-Efficient Spot-Scanning Proton Arc Therapy. Int J Radiat Oncol Biol Phys. 96, 1107–1116.

[4] Li X, Kabolizadeh P, Yan D, et al., 2018. Improve dosimetric outcome in stage III non-small-cell lung cancer treatment using spot-scanning proton arc (SPArc) therapy. Radiat Oncol. 13, 35.

[5] Ding X, Li X, Qin A, et al., 2018. Have we reached proton beam therapy dosimetric limitations?–A novel robust, delivery-efficient and continuous spot-scanning proton arc (SPArc) therapy is to improve the dosimetric outcome in treating prostate cancer. Acta Oncol (Madr). 57, 435-437.

[6] Ding X, Zhou J, Li X, et al., 2019. Improving dosimetric outcome for hippocampus and cochlea sparing whole brain radiotherapy using spotscanning proton arc therapy. Acta Oncol (Madr). 58, 483-490.

[7] Li X, Liu G, Janssens G, et al., 2019. The first prototype of spot-scanning proton arc treatment delivery. Radiother Oncol. 137, 130-136.

[8] Gu Wenbo, et al., 2020. A novel energy layer optimization framework for spot-scanning proton arc therapy. Med Phys. 47.5, 2072-2084.

[9] Gao H, 2019. Hybrid proton-photon inverse optimization with uniformity-regularized proton and photon target dose. Phys Med Biol. 64, 105003.

[10] Gao H, Lin B, Lin Y, et al., 2020. Simultaneous dose and dose rate optimization (SDDRO) for FLASH proton therapy. Med Phys. 47, 6388-6395.

[11] Gao H, Liu J, Lin Y, et al., 2022. Simultaneous dose and dose rate optimization (SDDRO) of the FLASH effect for pencil-beam-scanning proton therapy. Med Phys. 49, 2014-2025.

[12] Gu W, O'Connor D, Nguyen D, et al., 2018. Integrated beam orientation and scanning-spot optimization in intensity-modulated proton therapy for brain and unilateral head and neck tumors. Med Phys. 45, 1338–1350.

[13] Beck A, Teboulle M, 2009. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. SIAM J Imaging Sci. 2, 183–202.

[14] Boyd S, Parikh N, Chu E, et al., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning. 3, 1-122.

[15] Goldstein T, and Osher S, 2009. The split Bregman algorithm for l1 regularized problems. SIAM J Imaging Sci. 2, 323-343.

[16] Parikh, Neal, and Stephen Boyd, 2014. "Proximal algorithms." Foundations and Trends in optimization. 1.3, 127-239.

[17] Wieser HP, Cisternas E, Wahl N, et al., 2017. Development of the open-source dose calculation and optimization toolkit matRad. Med Phys. 44, 2556-2568.

[18] Lin Y, Kooy H, Craft D et al., 2016. A Greedy reassignment algorithm for the PBS minimum monitor unit constraint. Phys Med Biol. 61, 4665-4678.

[19] Gao H, Clasie BM, Liu T et al., 2019. Minimum MU optimization (MMO): an inverse optimization approach for the PBS minimum MU constraint. Phys Med Biol. 64, 125022.

[20] Lin Y, Clasie BM, Liu T et al., 2019. Minimum-MU and sparse-energy-level (MMSEL) constrained inverse optimization method for efficiently deliverable PBS plans. Phys Med Biol. 64, 205001.

[21] Gao H, Clasie BM, McDonald M et al., 2020. Plan-delivery-time constrained inverse optimization method with minimum-MU-per-energy-layer (MMPEL) for efficient pencil beam scanning proton therapy. Med Phys. 47, 3892-3897.

[22] Cai JF, Chen RC, Fan J, and Gao H, 2022. Minimum-monitor-unit optimization via a stochastic coordinate descent method. Phys Med Biol. 67, 015009.

[23] Wu Q, Mohan R, Morris M, et al., 2003. Simultaneous integrated boost intensity-modulated radiotherapy for locally advanced head-and-neck squamous cell carcinomas. I: dosimetric results. Int J Radiat Oncol Biol Phys. 56, 573-585.

Table 1. Comparison of dosimetric parameters between ES and EM. Optimized planning objective value *f*, CI and HI are unitless.

	Quantity (Unit)	ES	EM
Abdomen	CI	0.83	0.86
	HI	9.38	7.30
	$D_{mean,bowel}$ (Gy)	2.14	1.69
	$D_{mean,body}$ (Gy)	0.47	0.42
	f	0.04	0.01
Brain	CI	0.53	0.84
	HI	9.27	8.59
	$D_{mean,brainstem}$ (Gy)	3.32	2.77
	$D_{mean,body}$ (Gy)	0.93	0.95

	f	1.41	1.29
HN	CI	0.59	0.79
	HI	13.65	9.94
	$D_{mean,larynx}$ (Gy)	4.89	4.62
	$D_{mean,body}$ (Gy)	2.00	1.86
	f	0.04	0.04
Lung	CI	0.63	0.83
	HI	15.14	9.34
	$D_{mean,heart}$ (Gy)	1.02	0.89
	$D_{mean,body}$ (Gy)	2.39	2.24
	f	0.03	0.02

Table 2. Comparison of plan delivery time between ES and EM. The values of energy switching time T_E, spot delivery time T_B, and total plan delivery time $T=T_E+T_B$ have the unit of second.

	Quantity	ES	EM
Abdomen	SU	36	13
	SD	28	48
	T_E	215	100
	T_B	92	40
	T	307	140
Brain	SU	31	26
	SD	31	43
	T_E	189	169
	T_B	63	48
	T	252	217
HN	SU	35	24
	SD	26	58
	T_E	208	167
	T_B	252	192
	T	460	359

Lung	SU	32	24
	SD	32	47
	T_E	195	160
	T_B	220	179
	T	415	339