An adaptive spot placement method on Cartesian grid for pencil beam scanning proton therapy

Bowen Lin[1], **Shujun Fu**[1], **Yuting Lin**[2], **Ronny L Rotondo**[2], **Weizhang Huang**[3], **Harold H Li**[2], **Ronald C Chen**[2], **Hao Gao**[2]

[1]School of Mathematics, Shandong University, Jinan, China

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

[3]Department of Mathematics, University of Kansas, Lawrence, USA

Abstract

Pencil beam scanning (PBS) proton radiotherapy (RT) offers flexible proton spot placement near treatment targets for delivering tumoricidal radiation dose to tumor targets while sparing organs-at-risk (OAR). Currently the spot placement is mostly based on a non-adaptive sampling (NS) strategy on a Cartesian grid. However, the spot density or spacing during NS is a constant for the Cartesian grid that is independent of the geometry of tumor targets, and thus can be suboptimal in terms of plan quality (e.g., target dose conformality) and delivery efficiency (e.g., number of spots). This work develops an adaptive sampling (AS) spot placement method on the Cartesian grid that fully accounts for the geometry of tumor targets. Compared with NS, AS places (1) a relatively fine grid of spots at the boundary of tumor targets to account for the geometry of tumor targets and treatment uncertainties (setup and range uncertainty) for improving dose conformality, and (2) a relatively coarse grid of spots in the interior of tumor targets to reduce the number of spots for improving delivery efficiency and robustness to the minimum-minitor-unit (MMU) constraint. The results demonstrate that (1) AS achieved comparable plan quality with NS for regular MMU and substantially improved plan quality from NS for large MMU, using merely about 10% of spots from NS, where AS was derived from the same Cartesian grid as NS; (2) on the other hand, with similar number of spots, AS had better plan quality than NS consistently for regular and large MMU.

Keywords

pencil beam scanning (PBS); intensity modulated proton therapy (IMPT); adaptive sampling; adaptive spot placement; minimum-minitor-unit optimization

shujunfu@163.com .

1. Introduction

Pencil beam scanning (PBS) proton radiotherapy (RT) can provide flexible proton spot placement for intensity modulated proton therapy (IMPT). Currently the spot placement mostly utilizes a non-adaptive sampling (NS) strategy on a Cartesian grid (e.g., Page 359 of [1]). However, NS does not fully account for the geometry of tumor targets, as the spot density or spacing is a constant that is independent of the geometry of tumor targets. Two exceptions are so-called contour-scanning spot placement method [2] and mesh-based spot placement method [3]. In the contour-scanning method, driven by penumbra improvement of dose conformality to tumor targets, Meier et al placed spots along concentric paths based on the target contour at the boundary of tumor targets [2]. However, the improved conformality via the contour-scanning method came at the expense of degraded spot uniformity inside the target volume or close to the tumor boundary, which was subsequently addressed by the mesh-based method [3]. In the mesh-based method, to improve both lateral dose falloff and uniformity, ur Rehman et al [3] utilized Delaunay triangulation followed by Lloyd's algorithm to generate a non-Cartesian mesh grid of proton spots. However, the mesh triangulation method for uniform sampling has its own limitation such as suffering from poor mesh quality for complex geometry [4]. Plus, delivery of spots on the non-Cartesian grid is technically challenging, which can be inefficient or/and subject to spot positioning error.

Given the simplicity and universality of the Cartesian grid to place spots for tumor targets that can be arbitrarily shaped, this work will be based on the Cartesian grid and develop an adaptive sampling (AS) method for spot placement, which will be highly adaptive to the geometry of tumor targets. The first consideration of AS, which is the same as for [2,3], is to improve the plan quality, particularly at the boundary of tumor targets to improve target dose conformality. The second consideration is to minimize the number of spots for improving delivery efficiency and also the robustness to the minimum-monitor-unit (MMU) constraint. Specifically, AS will place (1) a relatively fine grid of spots at the boundary of tumor targets to account for the geometry of tumor targets and treatment uncertainties (setup and range uncertainty) for improving dose conformality, and (2) a relatively coarse grid of spots in the interior of tumor targets to reduce the number of spots and/or increase the spot weights for improving delivery efficiency.

2. Methods and Materials

2.1 Overview of Adaptive Sampling

For the purpose of this work, the PBS proton treatment planning process can be divided into three steps: planning setup, spot placement and spot weight optimization. The contribution of this work is on the spot placement via AS (Section 2.2 and 2.3). The planning setup (e.g., beam angles and plan objectives) is provided in Section 2.5. After the spot placement, the robust inverse optimization for IMPT is utilized to optimize spot weights while considering the MMU constraint for plan deliverability (Section 2.4).

During AS of spot placement, the spots are grouped first by beam angles and then by energy layers, which is the same as NS. However, within a specific energy layer for a particular

beam angle, AS is drastically different from NS in the sense that the spot placement in AS is highly adaptive to the geometry of tumor targets with variable spot spacing. Examples of AS are provided in Fig. 1, where the spot placement in AS is relatively fine at the boundary of tumor targets and coarse in the interior of tumor targets.

For presentation clarity, we will first introduce the 2D undersampling algorithm as the AS spot placement engine per energy layer in Section 2.2, and then provide the overall 3D AS algorithm for all spots per beam angle in Section 2.3. Note that in 2D undersampling algorithm, 2D refers to the 2D coordinate in beam's eye view (BEV), while in 3D AS algorithm, 3D refers to the 2D BEV coordinate plus the proton energy.

2.2 2D Undersampling Algorithm

This section of 2D undersampling algorithm serves as the AS spot placement engine per energy layer, which will be an essential component of the overall 3D AS algorithm for all spots per beam angle in Section 2.3. We will first introduce the notations and spot grouping in Section 2.2.1, and then provide the details of 2D undersampling algorithm in Section 2.2.2 and its analysis in Section 2.2.3.

2.2.1 Notations and Spot Grouping—Let us consider a set of $X \times Y$ spots

$$\mathbf{\Omega_{2D}} = \{(x, y) : 1 \leq x \leq X, 1 \leq y \leq Y\} \tag{1}$$

with the corresponding set of indexes

$$\mathbf{\Pi_{2D}} = \{l, 1 \leq l \leq |\mathbf{\Omega_{2D}}|\}, \tag{2}$$

where the operator $|\cdot|$ represents the counting measure of the set, i.e. the number of spots in the set.

For the purpose of spot grouping, we define the center spot $I = (x_0, y_0)$ of $\mathbf{\Omega_{2D}}$, which is determined through the following process: first $Iso = \mathbf{mean}(\mathbf{\Omega_{2D}})$, where $\mathbf{mean}$ is the averaging operator, and then one of the nearest spots to Iso in $\mathbf{\Omega_{2D}}$ is taken as the center spot I.

Then for spot grouping, we start from the center spot I, and expand the 2D coordinates simultaneously with an equal step $= 1$, i.e., the generation of concentric squares Ω_c with the side length $2(c-1)$ (Fig. 2), where $c \in [1, C]$, $C = max\{|X - x_0| + 1, |Y - y_0| + 1\}$. The number of spots of Ω_c satisfies

$$|\Omega_c| \leq \begin{cases} 1, & c = 1 \\ 8(c-1), & c \geq 2 \end{cases}. \tag{3}$$

Thus

$$|\mathbf{\Omega_{2D}}| = \sum_{c=1}^{C} |\Omega_c| \leq 4C^2 - 4C + 1. \tag{4}$$

Each subset Ω_c ($c \in [2, C]$) and its corresponding index subset Π_c are denoted as

$$\Omega_c = \{(m_c^i, n_c^i): 1 \le i \le |\Omega_c|\}, \tag{5}$$

$$\Pi_c = \{l_c^i: 1 \le i \le |\Omega_c|\}, \tag{6}$$

And $\Omega_c \subset \Omega_{2D}, \Pi_c \subset \Pi_{2D}; \Omega_{2D} = \cup_{c=1}^{C} \Omega_c, \Pi_{2D} = \cup_{c=1}^{C} \Pi_c$.

For the convenience of undersampling, the polar coordinate values of the Cartesian coordinates are computed. This goes as follows: after subtracting the coordinate of the center spot I from all spots in Ω_c, the coordinates of these spots are normalized to [0, 1], and then the following polar coordinate values are computed

$$\Omega_c^p = \{(\rho_c^i, \theta_c^i): 0 \le \rho_c^i \le 1, -\pi < \theta_c^i \le \pi, 1 \le i \le |\Omega_c|\}. \tag{7}$$

Note that we do not actually transform Cartesian coordinates into polar coordinates, but only use polar coordinate values for the purpose of grouping Cartesian coordinates. That is, the spots in each subset Ω_c are sorted in non-descending order with respect to the θ coordinate. For the notion convenience, we will still denote the reordered spot subset and its corresponding index subset as Ω_c and Π_c respectively in the following.

2.2.2 2D Undersampling Steps—The 2D undersampling for each subset Ω_c ($c \in [2, C]$) consists of three steps.

2.2.2.1 Step 1: The first undersampling step has two scenarioes.

Case 1.1: $mod(c, 2) = 1$, where mod is the modulus operator. The undersampling starts from the first spot of Ω_c, and the undersampling step size is 2, i.e., the alternative undesampling during which all spots with odd indexes are removed and all spots with even indexes are kept. The set of removed spots is denoted by $\Omega_{c,1}$ with the corresponding index set $\Pi_{c,1}$, and the set of remianing spots is denoted by $\widetilde{\Omega}_{c,1}$ with the corresponding index set $\widetilde{\Pi}_{c,1}$.

Case 1.2: $mod(c, 2) = 0$. The undersampling starts from the second spot of Ω_c, while the remaining process is the same as Case 1.1.

2.2.2.2 Step 2: The second undersampling step is peformed with respect to $\widetilde{\Omega}_{c,1}$, $c \in [2, C]$ obtained from the first undersampling. With two-to-one grouping of consecutive concentric sets $\widetilde{\Omega}_{c,1}$, the second undersampling step has two scenarioes.

Case 2.1: $mod(C-1, 2) = 0$. First the spots are re-combined into $\mathscr{J} = \frac{C-1}{2}$ groups by

$$G = \begin{cases} [2,3] & 1 \\ [4,5] & 2 \\ \vdots & \vdots \\ [c, c+1] & i' \\ \vdots & \vdots \\ [C-1, C] & \mathcal{J} \end{cases}, 1 \leq i' \leq \mathcal{J} \tag{8}$$

Then the alternative sampling similar to Step 1 is applied onto the re-combined groups, i.e.,

$$F = \begin{bmatrix} 1 & 1 \\ 0 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 0 & i' \\ 1 & i'+1 \\ \vdots & \vdots \\ mod(\mathcal{J}, 2) & \mathcal{J} \end{bmatrix}, \tag{9}$$

where $F(i') = 0$ means that all evenly-indexed spots from the i' group are kept, and $F(i') = 1$ means that all oddly-indexed spots from the i' group are kept. The set of removed spots is denoted by $\Omega_{c,2}$ with the corresponding index set $\Pi_{c,2}$, and the set of remianing spots is denoted by $\widetilde{\Omega}_{c,2}$ with the corresponding index set $\widetilde{\Pi}_{c,2}$.

Case 2.2: $mod(C-1, 2) = 1$. Now the spots are re-combined into $\mathcal{J} = ceil\left(\frac{C-1}{2}\right)$ groups, where $ceil$ is a rounding operator towards positive infinity, i.e.,

$$G = \begin{cases} [2,3] & 1 \\ [4,5] & 2 \\ \vdots & \vdots \\ [c, c+1] & i' \\ \vdots & \vdots \\ [C] & \mathcal{J} \end{cases}, 1 \leq i' \leq \mathcal{J}, \tag{10}$$

while the remaining process is the same as Case 2.1.

2.2.2.3 Step 3: The third undersampling step is peformed with respect to $\widetilde{\Omega}_{c,2}$, $c \in [2, C]$ obtained from the second undersampling, with the following two scenarioes.

Case 3.1 $mod(c, 2) = 0$. All spots are removed. Following prior notations, we have $\Omega_{c,3} = \widetilde{\Omega}_{c,2}, \Pi_{c,3} = \widetilde{\Pi}_{c,2}, \widetilde{\Omega}_{c,3} = \varnothing$, and $\widetilde{\Pi}_{c,3} = \varnothing$.

Case 3.2 $mod(c, 2) = 1$. All spots are kept. That is, $\Omega_{c,3} = \varnothing, \Pi_{c,3} = \varnothing, \widetilde{\Omega}_{c,3} = \widetilde{\Omega}_{c,2}$ and $\widetilde{\Pi}_{c,3} = \widetilde{\Pi}_{c,2}$.

2.2.3 Algorithm Summary and Analysis

The 2D undersampling algorithm has up to three aforementioned undersampling steps, i.e., the number of undersampling steps N 3. The total set of removed spots and its corresponding index set are

$$\overline{\Omega} = \cup_{c=1}^{C} \cup_{1}^{N} \Omega_{c,n} \tag{11}$$

$$\overline{\Pi} = \cup_{c=1}^{C} \cup_{1}^{N} \Pi_{c,n}, \tag{12}$$

and the set of remaing spots and its corresponding index set are

$$\widetilde{\Omega} = \cup_{c=1}^{C} \widetilde{\Omega}_{c,N} \tag{13}$$

$$\widetilde{\Pi} = \cup_{c=1}^{C} \widetilde{\Pi}_{c,N}. \tag{14}$$

Thus

$$\Omega_{2D} = \overline{\Omega} \cup \widetilde{\Omega}, \ \Pi_{2D} = \overline{\Pi} \cup \widetilde{\Pi}. \tag{15}$$

To further illstruate our 2D undersampling algorithm, an example based on Fig. 2 is provided in Appendix A. Note that the undersampling is not applied to the set with no more than β spots, i.e., when $|\Omega_c| \leq \beta$ or $|\widetilde{\Omega}_{c,n}| \leq \beta$. For the convenience of implementation, the overall 2D undersampling algorithm is summarized as follows.

Algorithm 1:

2D Undersampling

1:    Inputs and parameters: $\Omega_{2D}, \Pi_{2D}, N, \beta$

2:    Initialization: $\overline{\Omega} = \varnothing$, $\overline{\Pi} = \varnothing$, $\widetilde{\Omega} = \varnothing$, $\widetilde{\Pi} = \varnothing$, $c = 1$

3:    Calculate the isocenter *Iso* of Ω_{2D}

4:    Determine the center spot I

5:    Calculate C and determine spot groups $\Omega_c, c \in [1, C]$, and Π_c

6:    **while** c C

7:      Normalize the spot coordinates in Ω_c to [0,1]

8:      Compute polar coordinate values of Ω_c

9:      Sort the spots in Ω_c and the corresponding index set Π_c by θ

10:      $n = 1, \Omega_0 = \Omega_c, \Pi_0 = \Pi_c$

11:      **while** n N

12:       **if** $|\Omega_0| \leq \beta$

13:        $\widetilde{\Omega}_{c,n} = \Omega_0, \widetilde{\Pi}_{c,n} = \Pi_0$

14:        **break**

15:       **else**

16:          Perform the n-th undersampling processing to get $\Omega_{c,n}, \Pi_{c,n}, \widetilde{\Omega}_{c,n}$ and $\widetilde{\Pi}_{c,n}$

17:          $\Omega_0 = \widetilde{\Omega}_{c,n},\ \Pi_0 = \widetilde{\Pi}_{c,n}$

18:       **end if**

19:       $n = n + 1$

20:       $\overline{\Omega} = \overline{\Omega} \cup \Omega_{c,n},\ \overline{\Pi} = \overline{\Pi} \cup \Pi_{c,n}$

21:    **end while**

22:    $\widetilde{\Omega} = \widetilde{\Omega} \cup \widetilde{\Omega}_{c,N},\ \widetilde{\Pi} = \widetilde{\Pi} \cup \widetilde{\Pi}_{c,N}$

23:    $c = c + 1$

24: **end while**

25: Outputs: $\overline{\Omega}, \overline{\Pi}, \widetilde{\Omega}$ and $\widetilde{\Pi}$

Note that Line 12 of Algorithm 1 indicates that the undersampling step is not applied to the subset with no more than β spots, i.e., when $|\Omega_0| \leqslant \beta$; otherwise, the spot density may be too sparse, which can degrade the plan quality. For the same reason, the number of undersampling steps N is an input variable to Algorithm 1. That is, only when the number of spots derived from a given Cartesian grid is still sufficiently large, we set $N=3$.

Next we will provide a complexity analysis for the proposed 2D undersampling algorithm. Let us start from the concentric squares Ω_c (Fig. 2), for which the number of spots is determined by Eq. (3).

After the first undresamping,

$$\left|\widetilde{\Omega}\right| = \sum_{c=1}^{C} |\widetilde{\Omega}_{c,1}| = 1 + 2C(C-1). \tag{16}$$

After the second undresamping,

$$\left|\widetilde{\Omega}\right| = \sum_{c=1}^{C} |\widetilde{\Omega}_{c,2}| = 1 + C(C-1). \tag{17}$$

After the third undresamping,

$$|\widetilde{\Omega}| = \sum_{c=1}^{C} |\widetilde{\Omega}_{c,3}| = \begin{cases} \dfrac{(C-1)^2 + 1}{2}, & mod(C,2) = 0 \\[2ex] \dfrac{C^2 + 1}{2}, & mod(C,2) = 1 \end{cases}. \tag{18}$$

Thus, for a general 2D Cartesian grid after AS, which will be covered by a subset of the concentric squares Ω_c, the total number of remaining spots satisfies:

$$|\widetilde{\boldsymbol{\Omega}}| \leq \begin{cases} 1 + 2C(C-1), & N = 1 \\ 1 + C(C-1), & N = 2 \\ [(2\mathbf{\mathit{ceil}}(C/2) - 1)^2 + 1]/2 & N = 3 \end{cases} \tag{19}$$

2.3 3D AS Algorithm

Assuming there are J proton beams, the entire coordinate set of proton spots is

$$\boldsymbol{\Omega}_{3D} = \cup_{j=1}^{J} \ \boldsymbol{\Omega}_j, \tag{20}$$

and the corresponding index set is

$$\boldsymbol{\Pi}_{3D} = \cup_{j=1}^{J} \ \boldsymbol{\Pi}_j, \tag{21}$$

where $\boldsymbol{\Omega}_j$ and $\boldsymbol{\Pi}_j$ correspond to the jth beam.

3D AS algorithm is performed for each beam. The notations and 3D spot grouping are discussed in Section 2.3.1; the adaptive separation of spots into boundary and interior target sub-regions for each energy layer is described in Section 2.3.2; the oerall 3D AS algorithm is provided in Section 2.3.3.

2.3.1 Notations and 3D Spot Grouping—In the BEV, the 3D spots are axially organized in energy layers, while spots on each energy layer can be regarded as a group of 2D concentric squares (Fig. 2) or its subset. To overlay the proton spots with tumor targets, for each energy layer, the proton spots are projected into the patient volume, i.e., to determine the radiological depth of the Bragg peak for each spot. Then these depths are averaged for all spots for this energy layer, and the contours of tumor targets at this mean radiological depth are overlayed together with the spots for this energy layer. When there is no tumor target at this radiological depth, the nearest tumor slice to this radiological depth is used.

Let K_j be the number of energy layers for the jth beam. Then,

$$\boldsymbol{\Omega}_j = \cup_{k=1}^{K_j} \ \boldsymbol{\Omega}_{j,k} \,. \tag{22}$$

$$\boldsymbol{\Pi}_j = \cup_{k=1}^{K_j} \ \boldsymbol{\Pi}_{j,k} \,. \tag{23}$$

where $\boldsymbol{\Omega}_{j,k}$ and $\boldsymbol{\Pi}_{j,k}$ are respectively the coordinate set and the index set for the kth energy layer in the jth beam.

Moreover, we account for multiply connected regions from tumor targets, i.e., to define the target boundary set as

$$\boldsymbol{B}_{j,k} = \cup_{s=1}^{S_{j,k}} \boldsymbol{B}_{j,k,s},$$ (24)

where, the subset $\boldsymbol{B}_{j,k,s}$ represents the set of boundary coordinates of the s-th target region and $S_{j,k}$ is the number of connected regions, for the kth energy layer in the jth beam.

2.3.2 Classification of Boundary and Interior Spots—In this section, the proton spots are classified into boundary and interior spots, so that different undersampling rates can be applied to achieve the sharp dose falloffs at tumor boundaries and also the overall efficient utility of proton spots.

For the purpose of spot classfication, one needs to calculate the spot-to-boundary distance, which is via the L_1 norm[1]. The appropriateness of L1 norm here is because the iso-value curves of L1 norm are concentric squares, which are naturally formed on the Cartesian grid (Fig. 2). In contract, the iso-value curves of L2 norm are concentric circles, which is not a natural choice for the Cartesian grid. Specifically, the spot-to-boundry distance is the minimum distance between the spot and all boundary points in $\boldsymbol{B}_{j,k}$. The z coordinate is skipped here as it is a constant for each pair of $\boldsymbol{B}_{j,k}$ and $\boldsymbol{\Omega}_{j,k}$.

For each spot in $\boldsymbol{\Omega}_{j,k}$, we first determine which connected target region it belongs to. For example, given a proton spot $A \in \boldsymbol{\Omega}_{j,k}$, and the boundary point $B \in \boldsymbol{B}_{j,k,s}$ that has the smallest L_1 distance to A, the spot A is classified into the s-th proton subgroup $\boldsymbol{\Omega}_{j,k,s}$ that is associated with $\boldsymbol{B}_{j,k,s}$.

Then we determine whether this proton spot is boundary or interior to the s-th connnected region. For this purpose, we specify the coordinates for $\boldsymbol{B}_{j,k,s}$ and $\boldsymbol{\Omega}_{j,k,s}$

$$\boldsymbol{B}_{j,k,s} = \{(x_p, y_p) : 1 \leq p \leq |\boldsymbol{B}_{j,k,s}|\},$$ (25)

$$\boldsymbol{\Omega}_{j,k,s} = \{(x_q, y_q) : 1 \leq q \leq |\boldsymbol{\Omega}_{j,k,s}|\},$$ (26)

and the index set $\boldsymbol{\Pi}_{j,k,s}$ for $\boldsymbol{\Omega}_{j,k,s}$

$$\boldsymbol{\Pi}_{j,k,s} = \{l_q : 1 \leq q \leq |\boldsymbol{\Omega}_{j,k,s}|\},$$ (27)

which satisfy

$$\boldsymbol{\Omega}_{j,k} = \cup_{s=1}^{S_{j,k}} \boldsymbol{\Omega}_{j,k,s},$$ (28)

$$\boldsymbol{\Pi}_{j,k} = \cup_{s=1}^{S_{j,k}} \boldsymbol{\Pi}_{j,k,s}.$$ (29)

---

[1]The L_1-norm distance between two points $A(x_1, y_1)$ and $B(x_2, y_2)$ is defined as $L_1(A, B) = |A - B|$, i.e., $L_1(A, B) = |x_1 - x_2| + |y_1 - y_2|$.

The classification of proton spots into boundary and interior spots (Fig. 3) goes as follows.

1. Average the coordinates of all points and spots in $B_{j,k,s} \cup \Omega_{j,k,s}$, and then round to the nearest integer to determine the center $I_{j,k,s}$, $I_{j,k,s} \in B_{j,k,s} \cup \Omega_{j,k,s}$.

2. Subtract the coordinate of $I_{j,k,s}$ from the coordinates of all points and spots in $B_{j,k,s} \cup \Omega_{j,k,s}$.

3. Compute polar coordinate values for Cartesian coordinates, i.e.,

$$B_{j,k,s}^{\mathscr{P}} = \{(\rho_p, \theta_p) : 1 \le q \le |B_{j,k,s}|\}, \tag{30}$$

$$\Omega_{j,k,s}^{\mathscr{P}} = \{(\rho_q, \theta_q) : 1 \le q \le |\Omega_{j,k,s}|\}. \tag{31}$$

4. Calculate the spot-to-boundary distance d_q

$$d_q = \min_{p}\{|\rho_p - \rho_q| : 1 \le p \le |B_{j,k,s}|\} \tag{32}$$

and the spot-to-boundary distance set

$$\mathscr{D}_{j,k,s} = \{d_q : 1 \le q \le |\mathscr{D}_{j,k,s}|\}, \tag{33}$$

where $|\mathscr{D}_{j,k,s}| = |\Omega_{j,k,s}|$.

5. Classify the spots into boundary and interior spots based on the signed distance

$$\delta\rho_q = \rho_p - \rho_q, \tag{34}$$

where p is the boundary point determined from Eq. (32), and the threshold

$$\epsilon = \min\mathscr{D}_{j,k,s}^{+} + \frac{\gamma}{2}\left(max\ \mathscr{D}_{j,k,s}^{+} - min\ \mathscr{D}_{j,k,s}^{+}\right), \tag{35}$$

where $\mathscr{D}_{j,k,s}^{+}$ consists of the spots with positive signed distance, i.e.,

$$\mathscr{D}_{j,k,s}^{+} = \{\delta\rho_q : \delta\rho_q > 0\}. \tag{36}$$

To interpret $\delta\rho_q$, the spot is inside the boundary if $\delta\rho_q > 0$, at the boundary if $\delta\rho_p = 0$, and outside the boundary if $\delta\rho_q < 0$. γ is a classification parameter that controls the proportion into boundary or interior spots. Specifically, when $\delta\rho_q \ \epsilon$, the spot q is an interior spot; otherwise, it is a boundary spot.

The coordinate sets of boundary spots and interior spots are denoted as $\Omega_{j,k,s}^{b}$ and $\Omega_{j,k,s}^{in}$ respectively,

$$\mathbf{\Omega}_{j,k,s}^{b} = \{(x_q^b, y_q^b) : 1 \le q \le |\mathbf{\Omega}_{j,k,s}^{b}|\}, \tag{37}$$

$$\mathbf{\Omega}_{j,k,s}^{in} = \{(x_q^{in}, y_q^{in}) : 1 \le q \le |\mathbf{\Omega}_{j,k,s}^{in}|\}. \tag{38}$$

with the corresponding index sets $\mathbf{\Pi}_{j,k,s}^{b}$ and $\mathbf{\Pi}_{j,k,s}^{in}$,

$$\mathbf{\Pi}_{j,k,s}^{b} = \{l_q^b : 1 \le q \le |\mathbf{\Omega}_{j,k,s}^{b}|\}, \tag{39}$$

$$\mathbf{\Pi}_{j,k,s}^{in} = \{l_q^{in} : 1 \le q \le |\mathbf{\Omega}_{j,k,s}^{in}|\}. \tag{40}$$

Thus we have

$$\mathbf{\Omega}_j = \cup_{k=1}^{K_j} \ \cup_{s=1}^{S_{j,k}} \left(\mathbf{\Omega}_{j,k,s}^{b} \cup \mathbf{\Omega}_{j,k,s}^{in}\right), \tag{41}$$

$$\mathbf{\Pi}_j = \cup_{k=1}^{K_j} \ \cup_{s=1}^{S_{j,k}} \left(\mathbf{\Pi}_{j,k,s}^{b} \cup \mathbf{\Pi}_{j,k,s}^{in}\right). \tag{42}$$

and

$$\mathbf{\Omega}_{3D} = \cup_{j=1}^{J} \ \mathbf{\Omega}_j, \tag{43}$$

$$\mathbf{\Pi}_{3D} = \cup_{j=1}^{J} \ \mathbf{\Pi}_j. \tag{44}$$

2.3.3 Algorithm Summary—For every

$\mathbf{\Omega}_{j,k} = \cup_{s=1}^{S_{j,k}} \left(\mathbf{\Omega}_{j,k,s}^{b} \cup \mathbf{\Omega}_{j,k,s}^{in}\right), k \in [1, K_j], j \in [1, J]$, Algorithm 1 is applied to each subset $\mathbf{\Omega}_{j,k,s}^{b}$ and $\mathbf{\Omega}_{j,k,s}^{in}$ respectively, with the set of removed spots and its index set via Eq. (11) and (12)

$$\overline{\mathbf{\Omega}}_j = \cup_{k=1}^{K_j} \ \cup_{s=1}^{S_{j,k}} (\overline{\mathbf{\Omega}}_{j,k,s}^{b} \cup \overline{\mathbf{\Omega}}_{j,k,s}^{in}), \tag{45}$$

$$\overline{\mathbf{\Pi}}_j = \cup_{k=1}^{K_j} \ \cup_{s=1}^{S_{j,k}} (\overline{\mathbf{\Pi}}_{j,k,s}^{b} \cup \overline{\mathbf{\Pi}}_{j,k,s}^{in}), \tag{46}$$

and the set of remaining spots and its index set via Eq. (13) and (14)

$$\widetilde{\mathbf{\Omega}}_j = \cup_{k=1}^{K_j} \ \cup_{s=1}^{S_{j,k}} \left(\widetilde{\mathbf{\Omega}}_{j,k,s}^{b} \cup \widetilde{\mathbf{\Omega}}_{j,k,s}^{in}\right), \tag{47}$$

$$\widetilde{\boldsymbol{\Pi}}_j = \cup_{k=1}^{K_j} \ \cup_{s=1}^{S_{j,k}} \left(\widetilde{\boldsymbol{\Pi}}_{j,k,s}^b \cup \widetilde{\boldsymbol{\Pi}}_{j,k,s}^{in} \right). \tag{48}$$

To summarize, the entire set of removed spots and its index set are

$$\overline{\boldsymbol{\Omega}} = \cup_{j=1}^{J} \ \overline{\boldsymbol{\Omega}}_j, \tag{49}$$

$$\overline{\boldsymbol{\Pi}} = \cup_{j=1}^{J} \ \overline{\boldsymbol{\Pi}}_j, \tag{50}$$

and the entire set of remaining spots and its index set are

$$\widetilde{\boldsymbol{\Omega}} = \cup_{j=1}^{J} \ \widetilde{\boldsymbol{\Omega}}_j, \tag{51}$$

$$\widetilde{\boldsymbol{\Pi}} = \cup_{j=1}^{J} \ \widetilde{\boldsymbol{\Pi}}_j. \tag{52}$$

We have $\boldsymbol{\Omega}_{3D} = \overline{\boldsymbol{\Omega}} \cup \widetilde{\boldsymbol{\Omega}}, \ \boldsymbol{\Pi}_{3D} = \overline{\boldsymbol{\Pi}} \cup \widetilde{\boldsymbol{\Pi}}.$

To quantify the undersampling ratio, we define

$$\zeta_{in} = \frac{\left| \cup_{j=1}^{J} \ \cup_{k=1}^{K_j} \ \cup_{s=1}^{S_{j,k}} \ \widetilde{\boldsymbol{\Omega}}_{j,k,s}^{in} \right|}{\left| \cup_{j=1}^{J} \ \cup_{k=1}^{K_j} \ \cup_{s=1}^{S_{j,k}} \ \boldsymbol{\Omega}_{j,k,s}^{in} \right|} \times 100\%, \tag{53}$$

$$\zeta_b = \frac{\left| \cup_{j=1}^{J} \ \cup_{k=1}^{K_j} \ \cup_{s=1}^{S_{j,k}} \ \widetilde{\boldsymbol{\Omega}}_{j,k,s}^b \right|}{\left| \cup_{j=1}^{J} \ \cup_{k=1}^{K_j} \ \cup_{s=1}^{S_{j,k}} \ \boldsymbol{\Omega}_{j,k,s}^b \right|} \times 100\%, \tag{54}$$

$$\zeta = \frac{\left| \widetilde{\boldsymbol{\Omega}} \right|}{\left| \boldsymbol{\Omega}_{3D} \right|} \times 100\%, \tag{55}$$

for interior, boundary, and entire spots respectively.

The overall 3D AG algorithm is summarized as follows.

Algorithm 2:

3D Adpative Sampling

1:   Inputs and parameters: $\boldsymbol{\Omega}_{3D}, \boldsymbol{\Pi}_{3D}, N_b, N_{in}, \beta_b, \beta_{in}, \gamma$

2:   Initialization: $\overline{\boldsymbol{\Omega}} = \varnothing, \ \overline{\boldsymbol{\Pi}} = \varnothing, \ \widetilde{\boldsymbol{\Omega}} = \varnothing, \ \widetilde{\boldsymbol{\Pi}} = \varnothing$

3:   **for** $j = 1 : J$ **do**

4:      **for** $k = 1 : K_j$ **do**

5:         Determine spot sets $\mathbf{\Omega}_{j,k}$ and $\mathbf{\Pi}_{j,k}$

6:         Determine target boundary set $\boldsymbol{B}_{j,k} = \cup_{s=1}^{S_{j,k}} \boldsymbol{B}_{j,k,s}$

7:         Determine $\mathbf{\Omega}_{j,k} = \cup_{s=1}^{S_{j,k}} \mathbf{\Omega}_{j,k,s}$ and $\mathbf{\Pi}_{j,k} = \cup_{s=1}^{S_{j,k}} \mathbf{\Pi}_{j,k,s}$ asociated to each connected target region

8:         Classification of boundary and interior spots
$$\mathbf{\Omega}_{j,k} = \cup_{s=1}^{S_{j,k}} (\mathbf{\Omega}_{j,k,s}^b \cup \mathbf{\Omega}_{j,k,s}^{in}), \quad \mathbf{\Pi}_{j,k} = \cup_{s=1}^{S_{j,k}} \left(\mathbf{\Pi}_{j,k,s}^b \cup \mathbf{\Pi}_{j,k,s}^{in}\right)$$

9:         **for** $s = 1 : S_{j,k}$ **do**

10:            $\left[\overline{\mathbf{\Omega}}_{j,k,s}^b, \overline{\mathbf{\Pi}}_{j,k,s}^b, \widetilde{\mathbf{\Omega}}_{j,k,s}^b, \widetilde{\mathbf{\Pi}}_{j,k,s}^b\right] = \textbf{Algorithm\_1}(\mathbf{\Omega}_{j,k,s}^b, \mathbf{\Pi}_{j,k,s}^b, N_b, \beta_b)$

11:            $\left[\overline{\mathbf{\Omega}}_{j,k,s}^{in}, \overline{\mathbf{\Pi}}_{j,k,s}^{in}, \widetilde{\mathbf{\Omega}}_{j,k,s}^{in}, \widetilde{\mathbf{\Pi}}_{j,k,s}^{in}\right] = \textbf{Algorithm\_1}(\mathbf{\Omega}_{j,k,s}^{in}, \mathbf{\Pi}_{j,k,s}^{in}, N_{in}, \beta_{in})$

12:            $\overline{\mathbf{\Omega}} = \overline{\mathbf{\Omega}} \cup (\overline{\mathbf{\Omega}}_{j,k,s}^b \cup \overline{\mathbf{\Omega}}_{j,k,s}^{in}), \overline{\mathbf{\Pi}} = \overline{\mathbf{\Pi}} \cup (\overline{\mathbf{\Pi}}_{j,k,s}^b \cup \overline{\mathbf{\Pi}}_{j,k,s}^{in})$

13:            $\widetilde{\mathbf{\Omega}} = \widetilde{\mathbf{\Omega}} \cup (\widetilde{\mathbf{\Omega}}_{j,k,s}^b \cup \widetilde{\mathbf{\Omega}}_{j,k,s}^{in}), \widetilde{\mathbf{\Pi}} = \widetilde{\mathbf{\Pi}} \cup (\widetilde{\mathbf{\Pi}}_{j,k,s}^b \cup \widetilde{\mathbf{\Pi}}_{j,k,s}^{in})$

14:         **end for**

15:      **end for**

16:   **end for**

17:   Outputs: $\overline{\mathbf{\Omega}}, \overline{\mathbf{\Pi}}, \widetilde{\mathbf{\Omega}}, \widetilde{\mathbf{\Pi}}$

Note that different numbers of undersampling steps N_b and N_{in} are used for boundary and interior region of the target respectively with N_b N_{in}, to achieve fine sampling in the boundary region, and coarse sampling inside the interior region (Fig. 1) for balanced plan quality and spot placement efficiency.

2.4   MMU Optimization

For either NS or AS, the treatment planning of IMPT is to solve the following MMU optimization problem

$$\min_{x} f(x) \\ \text{s.t. } x \in \{0\} \cup [g, +\infty).$$

(56)

In Eq. (56), x denotes the spot weight (unit: number of protons) to be optimized, g is the planning MMU, and f is the sum of planning objectives. The plan objectives are based on dose-volume constraints (e.g., our previous works [12,13,17]). The MMU constraint is enforced in Eq. (56) to generate deliverable spots, which states that each entry of spot weight vector x is nonnegative, and no less than g if positive.

The MMU optimization problem Eq. (56) is nonconvex due to the nonconvexity of the MMU constraint. And the nonconvexity of the MMU constraint increases, as the MMU threshold g increases. Note that the plan with larger g has higher dose rate and thus shorter plan delivery time [13]. However, the nonconvexity from large g can severely degrade

the plan quality. For example, as g increases, the target dose conformality can deteriorate rapidly, which will be demonstrated in the result section. Therefore, we will consider the MMU optimization with regular and large values of g respectively in the result section.

Various methods have been developed to solve the MMU optimization problem for deliverable PBS plans, including postprocessing methods [5–7] and optimization methods [8–13]. For example, Varian Eclipse treatment planning system (TPS) currently uses the round method [5], while its research version of TPS use the optimization approach [10]. In this work, the MMU optimization is based on our previously developed iterative convex relaxation (ICR) algorithm [12,13] with inner loops solved by alternating direction method of multipliers [14–16], which has been shown to be effective for solving a variety of treatment planning problems besides MMU optimization, including energy-layer minimization [12], dose-rate optimization [13,17–19], and hybrid proton-photon optimization [20].

Specifically, the ICR handles the MMU constraint rigorously with the analytic formula, which is derived analytically from solving proximal operator with respect to the MMU constraint and then applied to the dual variable of spot weights that effectively sets this dual variable to be zero if it is value is less than $g/2$. Other methods may also be capable of solving the MMU problem for large g, e.g., the spot-reduction method [8], which can handle the MMU constraint empirically by iteratively setting certain percent of smallest weights to zero. It will be interesting to compare these methods in a future work.

2.5 Materials

AS was validated using head-and-neck (HN), liver and pancreas cases (Table 1), in comparison with NS, with results presented in Table 2–5 and Fig. 4–8. Specifically, we compare NS on a Cartesian grid of 3mm spot spacing (FINE), NS on a Cartesian grid of 9mm spot spacing (COARSE), and AS based on the same Cartesian grid of 3mm spot spacing as FINE (ADAPTIVE). Note that ADAPTIVE and FINE have the same (smallest) spot spacing, while ADAPTIVE and COARSE have the similar number of spots (Table 2–4). The choice that ADAPTIVE originates from the same Cartesian grid as FINE is because the spot spacing can have a great impact on the plan quality [21], and plan comparison using the same spacing eliminates the factor owing to the difference in spot spacing.

Robust optimization was considered with range and setup uncertainty. All plans in this study were based on clinically used DVH planning objectives, beam angles, and uncertainty level (Table 1). The dose influence matrix (the dose distribution per unit weight for each spot with the unit Gy/protons; e.g., in Chapter 15 of [1]) was generated using MatRad [22] with 5mm spot width (in both directions) and 3mm longitudinal spacing on $3mm^3$ grid. Here the spot width is defined as the full width at half maximum (FWHM). Although MatRad allows for different FWHM values for different energy, a constant 5mm was used for all energies for simplicity, i.e., use of generic beam model as provided in MatRad [22]. For fair comparison, the same plan objectives, the same ICR-based optimization algorithm, and the same plan normalization (D98=100% to CTV for the worst case) were used, while the only difference is dose influence matrix from NS or AS. The results (Table 2–4 and Fig. 4–8) are from probabilistically weighted proton dose from all uncertainty scenarios.

3. Results

3.1 ADAPTIVE v.s. FINE

Regarding the efficiency of spot utility, (1) the ratio of number of spots (i.e., N_x in Table 2–4) of ADAPTIVE over that of FINE was 10.5%, 11.1%, 11.0% for HN, liver, and pancreas respectively; (2) the ratio of number of nonzero spots after optimization (i.e., #x_g in Table 2–4) of ADAPTIVE over that of FINE was 12.2%, 10.9%, 11.9% for HN, liver, and pancreas respectively for g=10, and 19.5%, 14.3%, 27.2% for HN, liver, and pancreas respectively for g=60.

Regarding the plan quality, (1) for g=10, ADAPTIVE had comparable plan quality with FINE; (2) for g=60, ADAPTIVE had substantially better plan quality than FINE. These are indicated by the comparison of (1) total planning objective value f and DVH values (e.g., the maximum target dose D_{max}) in Table 2–4, (2) dose plots in Fig. 4–6, (3) target DVH plots in Fig. 7, and (4) dose profiles in Fig. 8.

In paticular, ADAPTIVE was more robust to large g than FINE. For example, for FINE, the target dose conformality was severely degraded from g=10 to g=60, as indicated by target DVH plots in Fig. 7, and target dose profiles in Fig. 8; in contrast, for ADAPTIVE, the target dose conformality was preserved from g=10 to g=60.

The robustness of ADAPTIVE for large g suggests that ADAPTIVE allows for the delievery of high dose rate [13]. Note that (1) ADAPTIVE had the same number of energy layers as FINE, (2) total number of protons to be delievered (i.e., $\|x\|_1$ in Table 2–4) for ADAPTIVE was simiar to that for FINE, and (3) $\|x\|_1$ for g=60 was simiar to that for g=10 for ADAPTIVE, while $\|x\|_1$ for g=60 was slightly larger than g=10 for FINE (e.g., 1.44 *v.s.* 1.38 in Table 3, 0.81 *v.s.* 0.75 in Table 4). Therefore, ADAPTIVE with larger g allows for shorter dose delivery time in higher dose rate, with preserved plan quality. In contrast, FINE takes longer time to deliver more protons, and moreover its plan quality was severely degraded.

The undersampling ratios from FINE to ADAPTIVE is summarized in Table 5, in which the interior spots had more undersampling than the boundary spots, as indicated by larger values of ζ_b than ζ_{in}. This is consistent with the methdology of AS (e.g., Fig. 1).

3.2 ADAPTIVE v.s. COARSE

ADAPTIVE had similar number of spots (i.e., N_x in Table 2–4) with COARSE, with ADAPTIVE being slightly less than COARSE. To match N_x of ADAPTIVE, the spot spacing for COARSE was chosen to be 9mm. The purpose of comparing ADAPTIVE and COARSE is to compare the plan quality between AS and NS, when both have similar number of spots to be optimized.

In terms of plan quality, ADAPTIVE had better plan quality than COARSE. This is indicated by the comparison of (1) total planning objective value f and DVH values (e.g., the maximum target dose D_{max}) in Table 2–4, (2) dose plots in Fig. 4–6, (3) target DVH plots in Fig. 7, and (4) dose profiles in Fig. 8.

3.3 Summary

To summarize, (1) with the same (smallest) spot spacing, ADAPTIVE had comparable plan quality with FINE for g=10, and substantially better plan quality than FINE for g=60, using ~10% of spots; (2) with the similar number of spots, ADAPTIVE had better plan quality than COARSE. These two summaries demonstrate the advantages of AS compared to NS, in terms of plan quality (e.g., target dose conformality) and delivery efficiency (e.g., number of spots).

4. Disscussion

Note that all spots of the ADAPTIVE exist in the FINE, because ADAPTIVE originates from FINE with the same spot spacing. As a result, the comparison between FINE and ADAPTIVE indeed shows FINE had slightly better plan quality than ADAPTIVE for g=10, which is as expected. However, this is not the case for g=60, in which FINE had substantially worse plan quality than ADAPTIVE. This is because the MMU optimization is nonconvex and FINE with more spots is more sensitive to the nonconvexity from the MMU constraint, especially for large *g*. In contrast, ADAPTIVE was still able to preserve the plan quality for large g. Therefore, AS is more robust to the MMU constraint than NS, besides efficient spot utility (e.g., ADAPTIVE using ~10% of spots from FINE).

Large spot spacing can compromise plan quality [21], e.g., COARSE. For NS, the spot spacing has to be uniformly refined to improve plan quality, e.g., from COARSE to FINE. Moreover, such uniform refinement may not lead to the improved plan quality in the presence of strong nonconvexity from large MMU threshold, e.g., FINE for g=60 in this study. In contrast, AS allows for variable spot spacing to increase the sampling density at the locations (e.g., the boundary of tumor targets) that need smaller spot spacing and decrease the sampling density at the locations (e.g., the interior of tumor targets) that larger spot spacing is sufficient for, so that the total number of spots remains the same. This has been shown to be efficient in spot utility and robust to the MMU constraint, e.g., ADAPTIVE in this result section.

The Cartesian grid in this work is a square grid. Another commonly used Cartesian grid is the hexagonal grid, which differs from the square gird by alternating line-by-line shifts. However, they are essentially equivalent in terms of the uniformity of spot placement concerning the plan quality, although hexagonal grid can be slightly more efficient in boundary representation. In this study, without loss of generality, we choose the square grid, which is relatively easy to adapt with. Note that adaptive grid via AS is drastically different from Cartesian grid (regardless of being hexagonal or square) in the sense that the spot spacing is a constant for Cartesian grid, but a variable for adaptive grid.

A limitation of AS is that the tumor targets need to be sufficiently large compared to the spot width and the spot spacing, to allow AG to place fine grid at the boundary and coarse grid in the interior of the targets.

While this work and others [2,3] only consider the AS in the lateral directions, the AS in the longitudinal direction (i.e., the proton energy) or the optimization of energy layer

distribution was previously studied by Kang et al [23], i.e., to determine an efficient longitudinal arrangement of Bragg peaks in a target volume using the Nyquist-Shannon sampling theorem.

Besides spot placement methods in this and prior works [2,3], the lateral dose falloffs can also be sharpened using dynamic collimation system [24] and spot size reduction [25]. However, dynamic collimation system may come at the cost of prolonged treatment time due to the mechanical movement of collimators, increased cost, and compatibility issues with existing PBS systems, while the spot size is a machine parameter that depends on the mechanical design and it is usually not a feasible option to reduce the spot size too much for existing delivery systems.

In this work, the AS is with respect to the spots on the same energy layer, instead of on the same radiological depth. In order for spots to be at the same radiological depth in heterogenous tissues, these spots are mostly likely to have different energies, although the energy variation can be small. As a result, although this should improve the target coverage in terms of evenly-distributed spot placement in radiological depth, the delivery of spots with different energies can be technically challenging and practically inefficient, as the switching of proton energies can take a significant amount of time [23]. Therefore, we have arranged the spots on the same energy layer instead of the same radiological depth, although the AS algorithm is also applicable to the spots of the same radiological depth. Moreover, the AS can be combined with our previously developed methods using sparse-energy-layer regularization [12] and plan-delivery-time constraint [13] to further improve the delivery efficiency of PBS plans, with further reduced number of spots and/or increased spot weights, which will be a future work.

On the other hand, the AS has equal step sizes in the lateral directions, which may not be ideal for the elongated target volume. AS with variable step sizes in the lateral directions will be addressed in a future work.

5. Conclusion

We have developed an adapative spot placement method on the Cartesian grid that is fully compatible with existing treatment planning convention and systems. Compared with NS, AS can improve both plan quality (e.g., target dose conformality) and delivery efficiency (e.g., number of spots). The key for AS is to place (1) a relatively fine grid of spots at the boundary of tumor targets to fully account for the geometry of tumor targets to improve target dose conformality, and (2) a relatively coarse grid of spots in the interior of tumor targets to reduce the number of spots for improving delivery efficiency and also the roboustness to the MMU constraint.

Acknowledgment

Appendix A

To fully illstruate the 2D undersampling algorithm (Algorithm 1), An example based on Fig. 2 is provided in Fig. A1 with the remaining spots after each undersampling step.
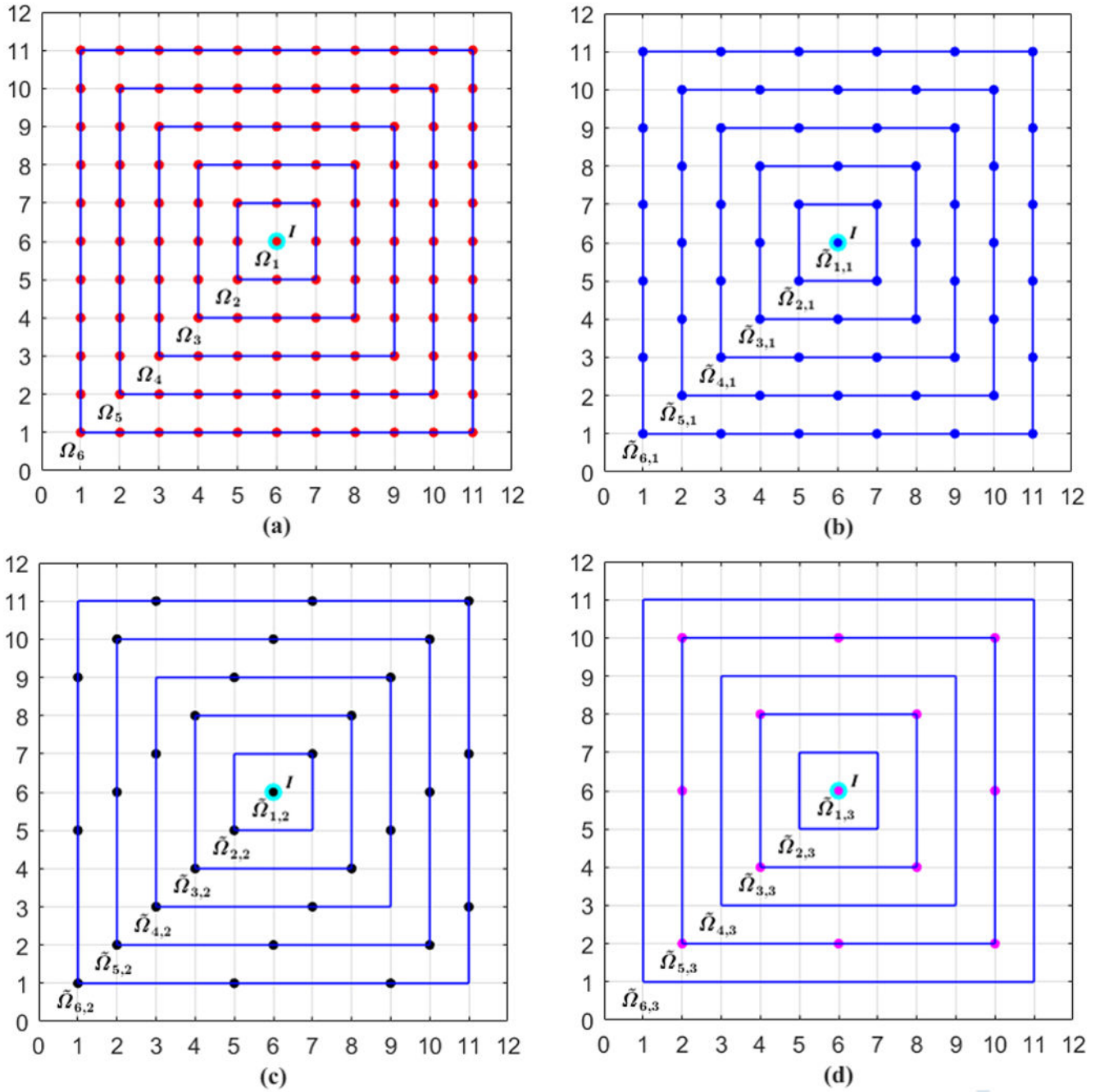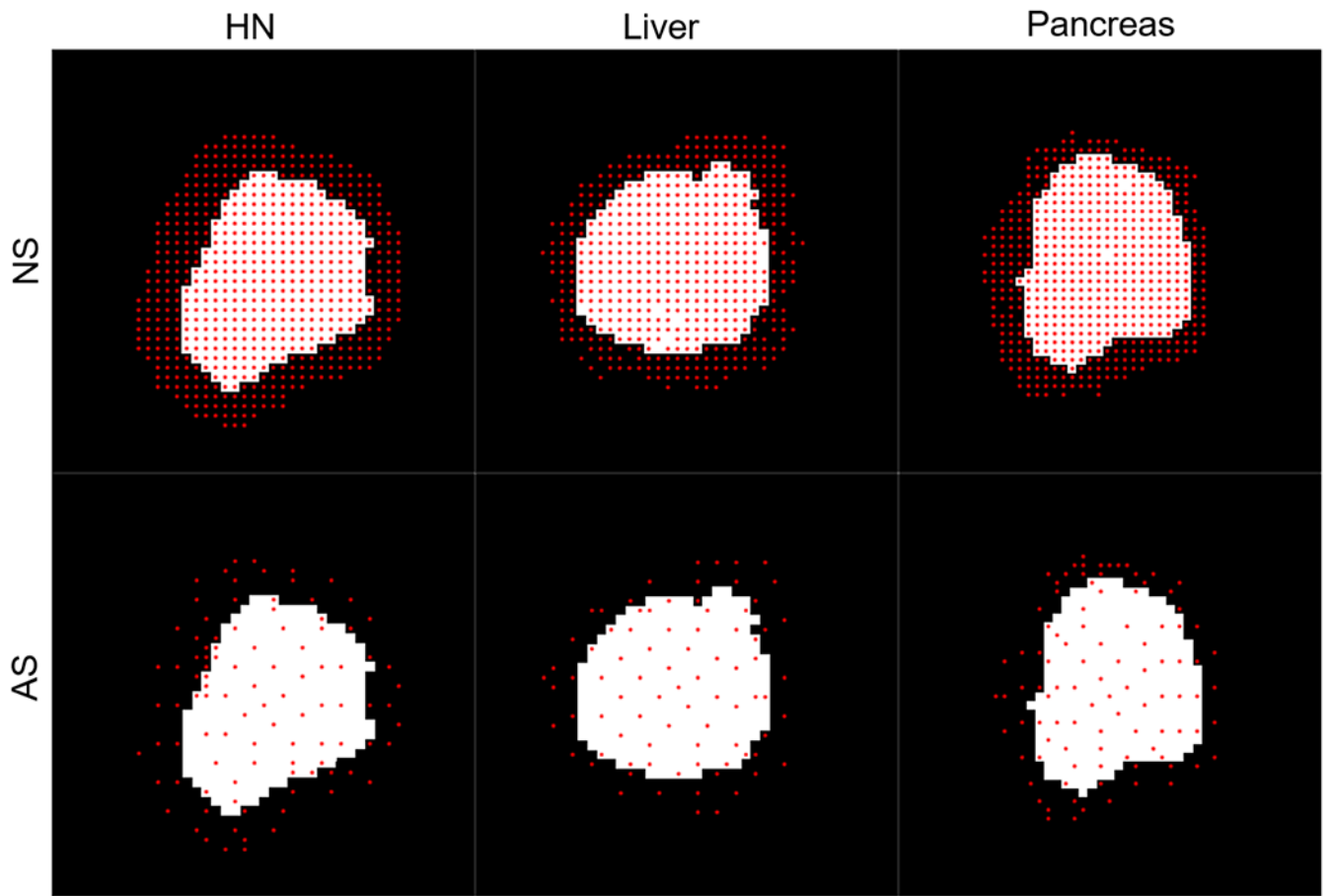

Figure A1.
An example of 2D undersampling for Fig. 2. (a) The red dots represent proton spots, and the blue concentric squares denote spot grouping into Ω_c, $c \in [1,6]$. (b) The blue dots represent the remaining spots after the first undersamplingg. (c) The dark dots represent the remaining

spots after the second undersampling. (d) The magenta dots represent the remaining spots after the third undersampling.

References

[1]. Paganetti H ed., 2016. Proton Therapy Physics. CRC Press.

[2]. Meier G, Leiser D, Besson R et al. , 2017. Contour scanning for penumbra improvement in pencil beam scanned proton therapy. Phys Med Biol. 62, 2398–2416. [PubMed: 28151727]

[3]. ur Rehman M, Erhart K, Kielbasa J et al. , 2019. An optimized approach for robust spot placement in proton pencil beam scanning. Phys Med Biol. 64, 235016. [PubMed: 31618722]

[4]. Freitag LA and Ollivier-Gooch C, 1997. Tetrahedral mesh improvement using swapping and smoothing. Int J Numer Methods Eng. 40, 3979–4002.

[5]. Zhu XR, Sahoo N, Zhang X et al. , 2010. Intensity modulated proton therapy treatment planning using single-field optimization: the impact of monitor unit constraints on plan quality. Med Phys. 37, 1210–1219. [PubMed: 20384258]

[6]. 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. [PubMed: 27245098]

[7]. 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. [PubMed: 31082813]

[8]. Albertini F, Gaignat S, Bosshardt M et al., 2009. Planning and optimizing treatment plans for actively scanned proton therapy. In: Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning, and Inverse Problems. Madison, WI: Medical Physics Publishing. 1–18.

[9]. Cao W, Lim G, Li X et al. , 2013. Incorporating deliverable monitor unit constraints into spot intensity optimization in intensity-modulated proton therapy treatment planning. Phys Med Biol. 58, 5113–5125. [PubMed: 23835656]

[10]. Howard M, Beltran C, Mayo CS et al. , 2014. Effects of minimum monitor unit threshold on spot scanning proton plan quality. Med Phys. 41, 091703. [PubMed: 25186378]

[11]. Shan J, An Y, Bues M et al. , 2018. Robust optimization in IMPT using quadratic objective functions to account for the minimum MU constraint. Med Phys. 45, 460–469. [PubMed: 29148570]

[12]. 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. [PubMed: 31530746]

[13]. 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. [PubMed: 32614472]

[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]. Gao H, 2016. Robust fluence map optimization via alternating direction method of multipliers with empirical parameter optimization. Phys Med Biol. 61, 2838–2850. [PubMed: 26987680]

[17]. 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. [PubMed: 33068294]

[18]. Lin Y, Lin B, Fu S, et al. , 2021. SDDRO-Joint: simultaneous dose and dose rate optimization with the joint use of transmission beams and Bragg peaks for FLASH proton therapy. Phys Med Biol. 66, 125011.

[19]. Gao H, Liu J, Lin Y, et al. , 2021. Simultaneous dose and dose rate optimization (SDDRO) of the FLASH effect. Med Phys. Accepted.

[20]. Gao H, 2019. Hybrid proton-photon inverse optimization with uniformity-regularized proton and photon target dose. Phys Med Biol. 64, 105003. [PubMed: 30978714]

[21]. Liu C, Schild SE, Chang JY, et al. , 2018. Impact of spot size and spacing on the quality of robustly optimized intensity modulated proton therapy plans for lung cancer. Int J Rad Oncol Biol Phys. 101, 479–489.

[22]. 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. [PubMed: 28370020]

[23]. Kang JH, Wilkens JJ, Oelfke U, 2008. Non-uniform depth scanning for proton therapy systems employing active energy variation. Phys Med Biol. 53, N149–N155. [PubMed: 18401060]

[24]. Hyer DE, Hill PM, Wang D, et al. , 2014. Effects of spot size and spot spacing on lateral penumbra reduction when using a dynamic collimation system for spot scanning proton therapy. Phys Med Biol 59, N187–N196. [PubMed: 25330783]

[25]. Moteabbed M, Yock TI, Depauw N, et al. , 2016. Impact of spot size and beam-shaping devices on the treatment plan quality for pencil beam scanning proton therapy. Int J Rad Oncol Biol Phys. 95, 190–198.
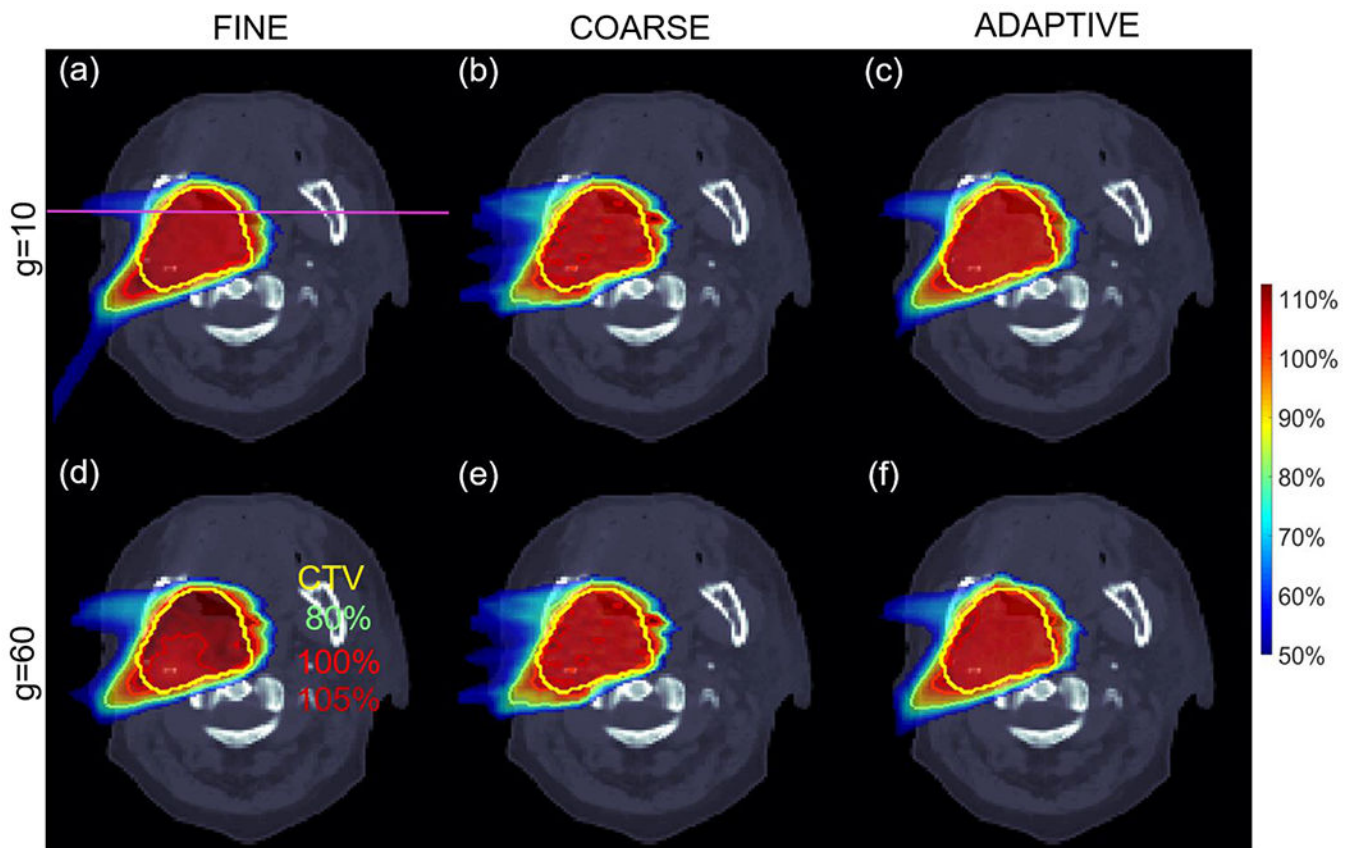
Figure 1.
Spot maps from NS and AS.
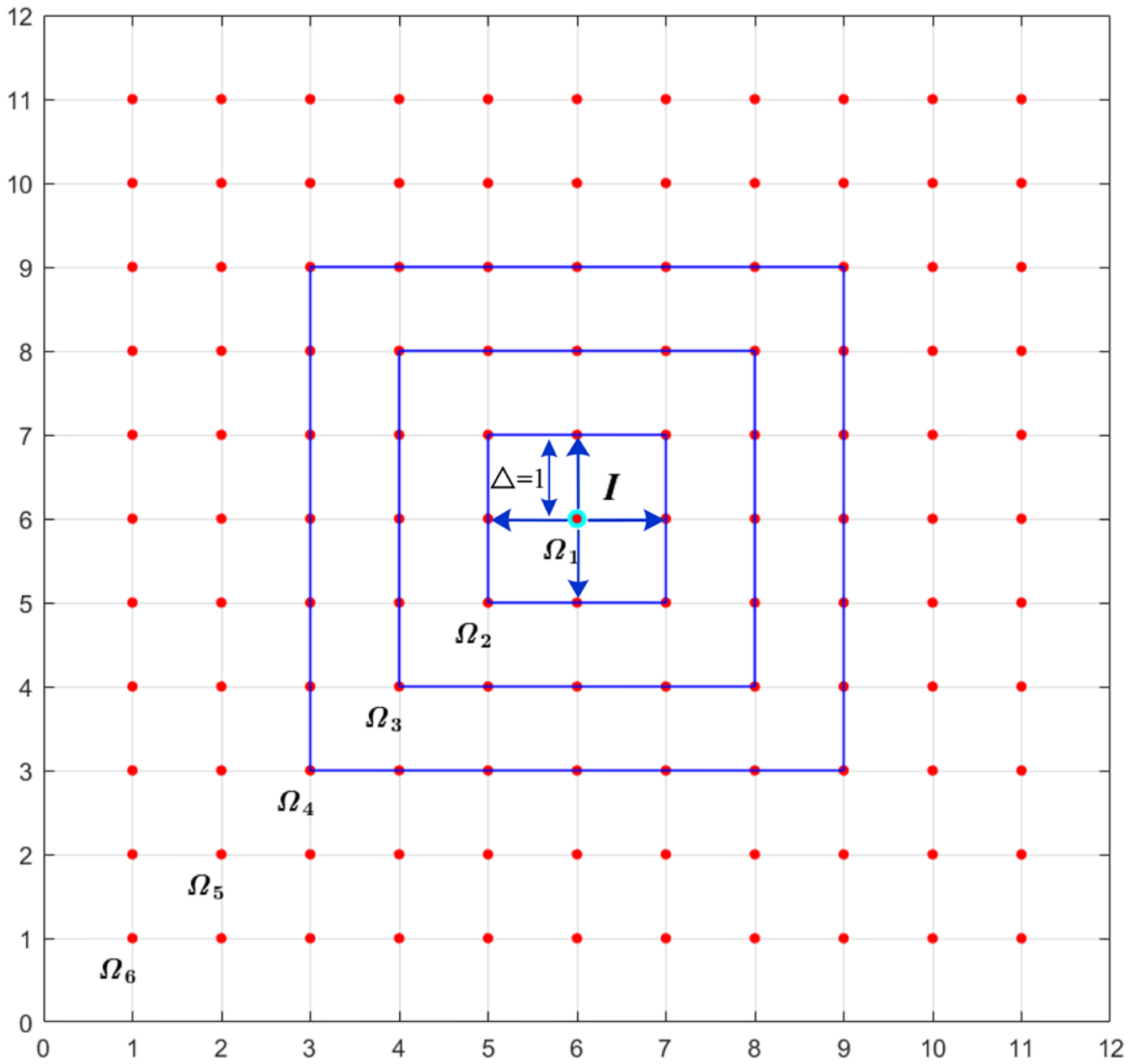
Figure 2.
Spot grouping into concentric squares. Here $X = Y = 11$, $I = (6,6)$, and $C = 6$.
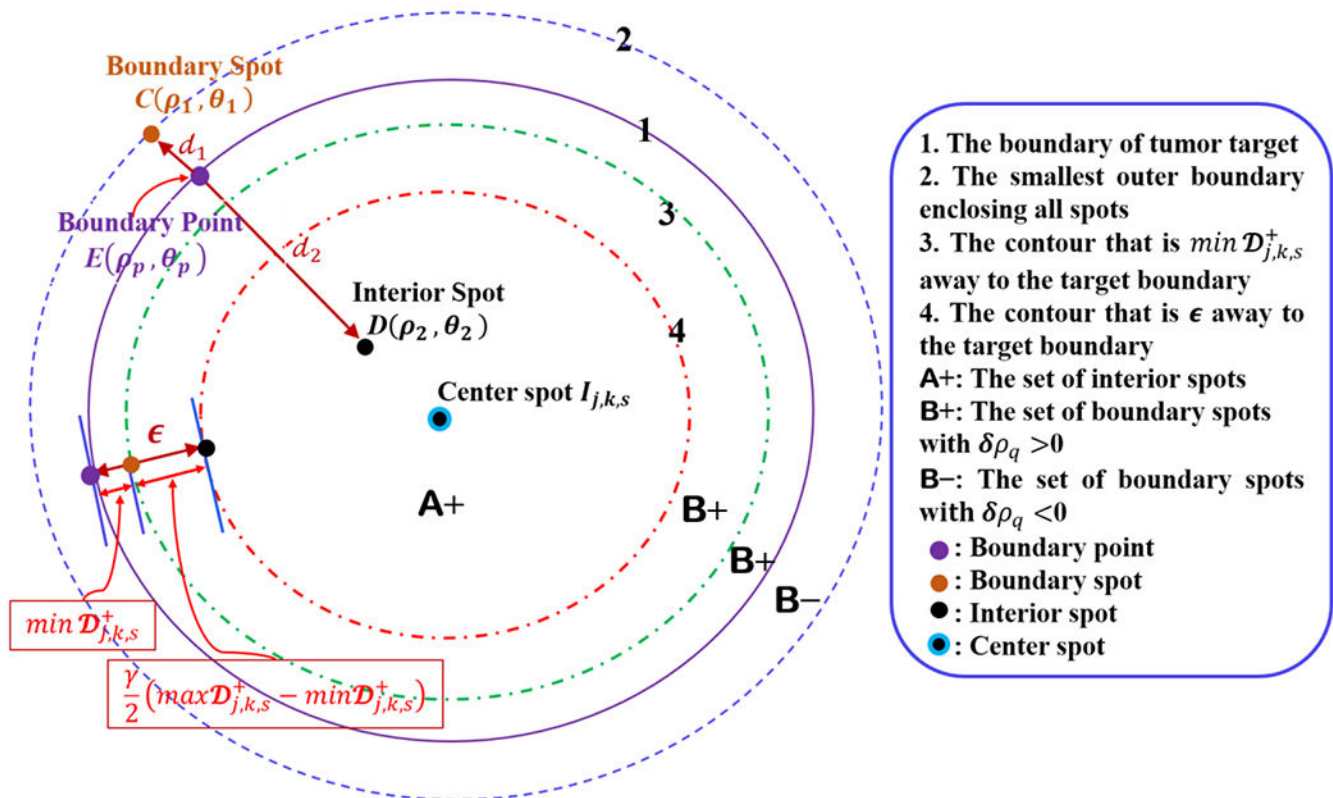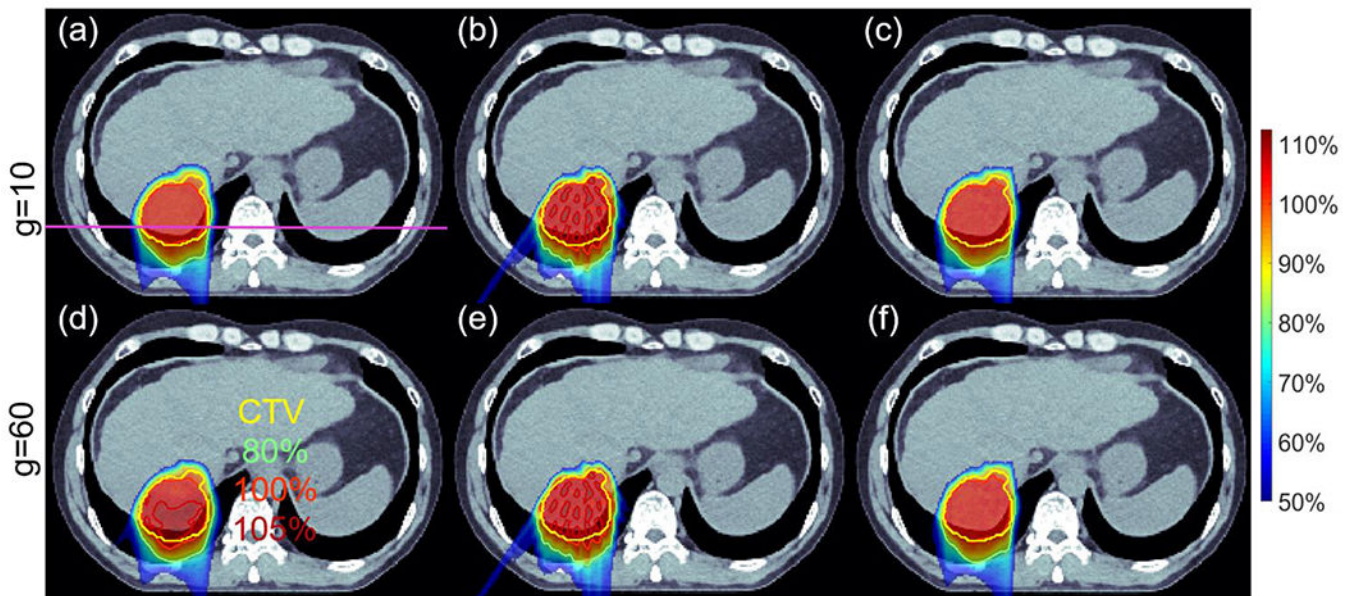
Figure 3.
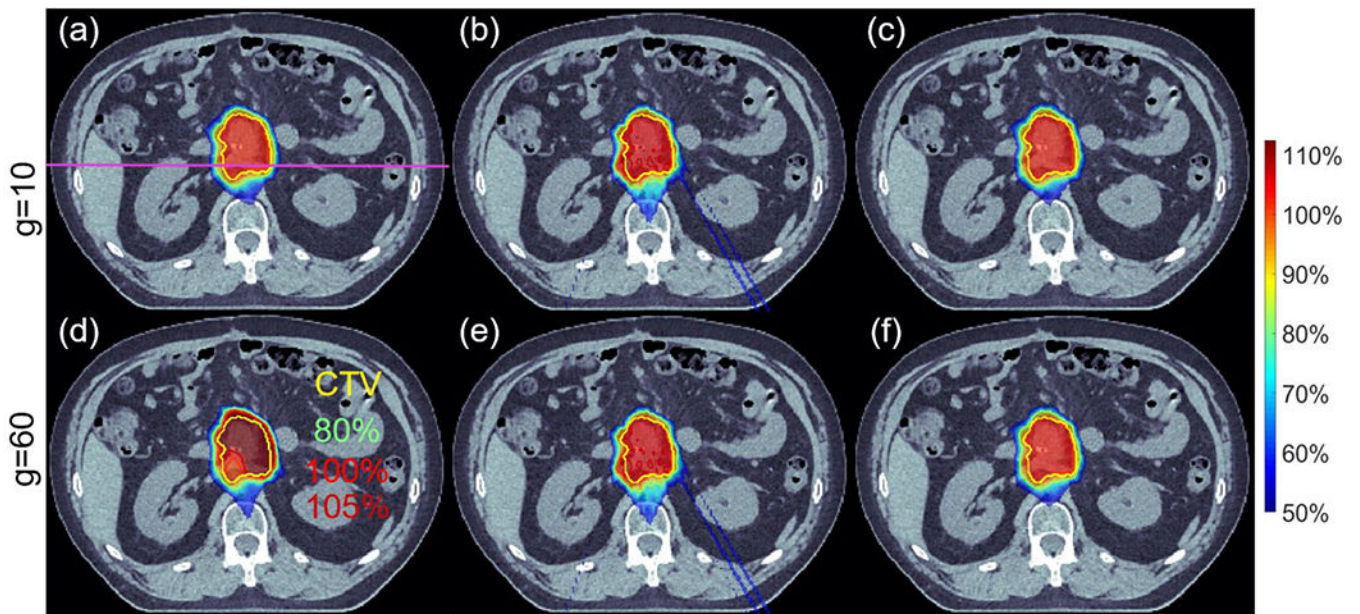Classification into boundary and interior spots.

Figure 4.

HN. (a)-(c) dose plots respectively from FINE, COARSE, ADAPTIVE with g=10; (d)-(f) dose plots respectively from FINE, COARSE, ADAPTIVE with g=60. CTV, 80%, 100% and 105% isodose lines are highlighted in (a)-(f) for comparing target dose conformality. Note that ADAPTIVE and FINE have the same spot spacing, while ADAPTIVE and COARSE have the similar number of spots; the definitions of these three methods are provided in Section 2.5.
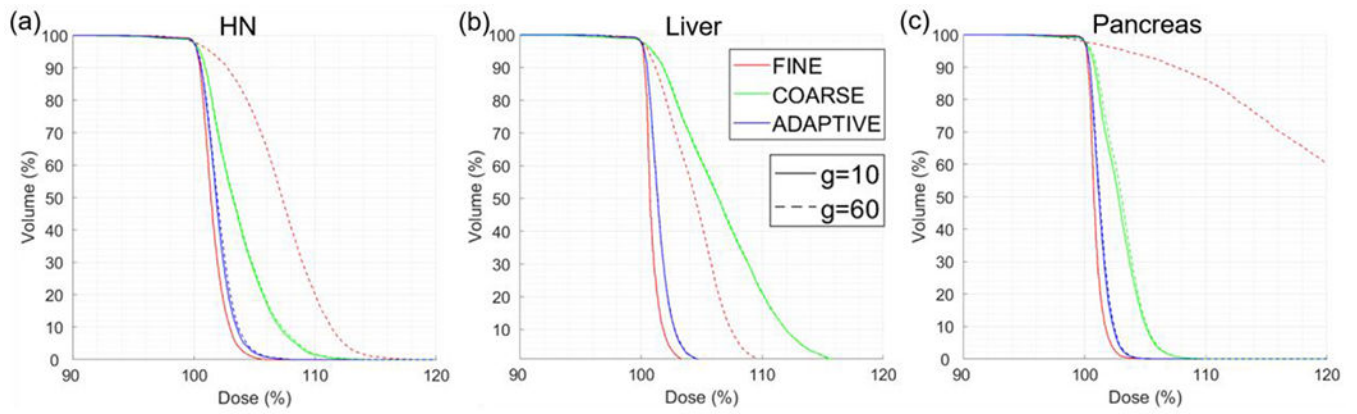
Figure 5.
Liver. (a)-(c) dose plots respectively from FINE, COARSE, ADAPTIVE with g=10; (d)-(f) dose plots respectively from FINE, COARSE, ADAPTIVE with g=60. CTV, 80%, 100% and 105% isodose lines are highlighted in (a)-(f) for comparing target dose conformality. Note that ADAPTIVE and FINE have the same spot spacing, while ADAPTIVE and COARSE have the similar number of spots; the definitions of these three methods are provided in Section 2.5.
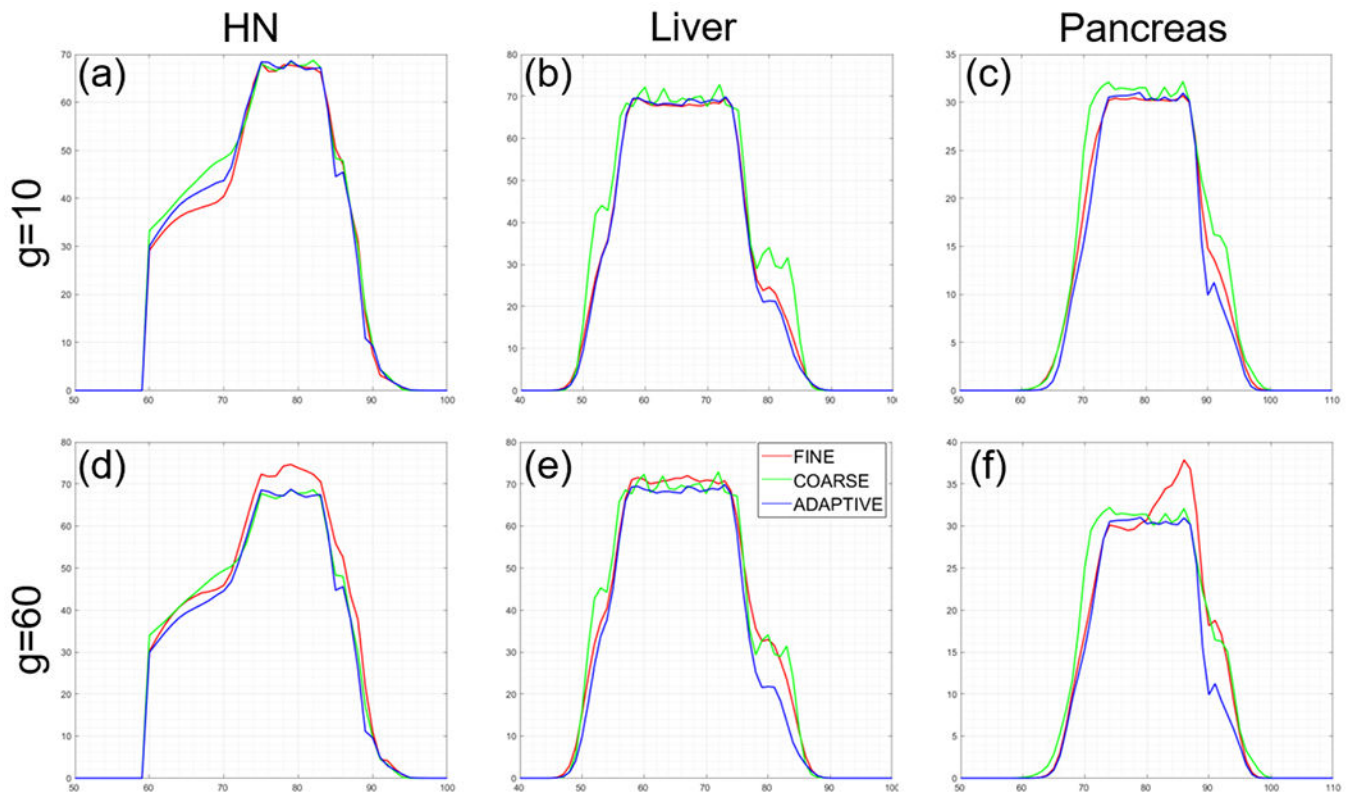
Figure 6.
Pancreas. (a)-(c) dose plots respectively from FINE, COARSE, ADAPTIVE with g=10; (d)-(f) dose plots respectively from FINE, COARSE, ADAPTIVE with g=60. CTV, 80%, 100% and 105% isodose lines are highlighted in (a)-(f) for comparing target dose conformality. Note that ADAPTIVE and FINE have the same spot spacing, while ADAPTIVE and COARSE have the similar number of spots; the definitions of these three methods are provided in Section 2.5.

Figure 7.
Target DVH plots. (a) HN; (b) liver; (c) pancreas. Since all plans have the same normalization, DVH plots are truncated with the plotting window [90%, 120%] of target dose.

Figure 8.
Profiles. (a)-(c) profile plots respectively from HN, liver, pancreas with g=10; (d)-(f) profile plots respectively from HN, liver, pancreas with g=60. The profiles are from the lines in Fig. 4–6(a).

Table 1.

Validation cases.

Case	Fractionation	Uncertainty	Beam Angles
HN	2.0Gy×33	(3mm, 3.5%)	290°, 260°, 210°, 260°[1]
Liver	4.5Gy×15	(5mm, 3.5%)	145°, 215°, 180.1°
Pancreas	6.0Gy×5	(5mm, 3.5%)	150°, 300°, 200°

[1] with couch kick of 30°.

Table 2.

HN.

HN	FINE		COARSE		ADAPTIVE	
N_x	59,400		6,640		6,244	
g	10	60	10	60	10	60
f	9.0	56.6	19.7	19.7	10.0	11.0
D_{max}	106.5	118.4	114.3	115.4	109.9	110.8
D_{5cc}	55.9	72.6	58.7	59.1	56.8	57.6
#x g	38,997	21,361	4,888	4,198	4,741	4,161
$\|x\|_I$	1.48	1.48	1.56	1.58	1.47	1.50

N_X: number of spots available for optimization; g: MMU threshold (unit: 10^6 protons); f: sum of planning objectives after optimization; D_{max}: maximum dose of CTV (unit: percentage with respect to the target dose; D_{5cc}: minimum dose for 5cc of mandible receiving the highest dose (unit: Gy); #x g: number of nonzero spots after optimization; $\|x\|_I$: sum of nonzero spot weights after optimization (unit: 10^{12} protons).

Table 3.

Liver.

Liver	FINE		COARSE		ADAPTIVE	
N_x	32,704		3,641		3,614	
g	10	60	10	60	10	60
f	11.1	23.3	41.7	42.2	12.2	12.4
D_{max}	106.8	111.9	118.4	118.7	107.2	107.2
D_{max_c}	16.8	25.6	47.5	47.3	18.2	18.5
#x g	27,364	20,031	2,655	2,756	2,975	2,866
$\|x\|_1$	1.38	1.44	1.58	1.60	1.38	1.39

N_X: number of spots available for optimization; g: MMU threshold (unit: 10^6 protons); f: sum of planning objectives after optimization; D_{max}: maximum dose of CTV (unit: percentage with respect to the target dose; $D_{max,c}$: maximum dose of cord (unit: Gy); #x g: number of nonzero spots after optimization; $\|x\|_1$: sum of nonzero spot weights after optimization (unit: 10^{12} protons).

Table 4.

Pancreas.

Pancreas	FINE		COARSE		ADAPTIVE	
N_x	38,238		4,253		4,192	
g	10	60	10	60	10	60
f	1.4	102.9	2.5	2.6	1.5	1.5
D_{max}	106.6	154.8	110.5	110.3	106.5	106.2
D_{1cc}	27.8	32.5	28.5	28.6	27.9	28.0
#x g	28,075	10,970	3,373	3,002	3,327	2,983
$\|x\|_1$	0.75	0.81	0.79	0.80	0.74	0.76

N_X: number of spots available for optimization; g: MMU threshold (unit: 10^6 protons); f: sum of planning objectives after optimization; D_{max}: maximum dose of CTV (unit: percentage with respect to the target dose; D_{1cc}: minimum dose for 1cc of small bowel receiving the highest dose (unit: Gy); #x g: number of nonzero spots after optimization; $\|x\|_1$: sum of nonzero spot weights after optimization (unit: 10^{12} protons).

Table 5.

Undersampling ratios from AS. ζ_b, ζ_{in}, and ζ are the undersampling ratios of spots with respect to interior, boundary, and all target regions defined by Eq. (53)–(55) respectively.

	HN	Liver	Pancreas
ζ_{in} (%)	10.1	10.5	10.3
ζ_b (%)	14.2	18.3	16.9
ζ (%)	10.5	11.1	11.0