HHS Public Access

Author manuscript

*Med Phys*. Author manuscript; available in PMC 2024 June 01.

Published in final edited form as:

*Med Phys*. 2023 June ; 50(6): 3258–3273. doi:10.1002/mp.16392.

# Beam angle optimization for proton therapy via group-sparsity based angle generation method

**Haozheng Shen**[1,#], **Gezhi Zhang**[1,#], **Yuting Lin**[2], **Ronny L Rotondo**[2], **Yong Long**[1], **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

## Abstract

**Background:** In treatment planning, beam angle optimization (BAO) refers to the selection of a subset with a given number of beam angles from all available angles that provides the best plan quality. BAO is a NP-hard combinatorial problem. Although exhaustive search (ES) can exactly solve BAO by exploring all possible combinations, ES is very time-consuming and practically infeasible.

**Purpose:** To the best of our knowledge, (1) no optimization method has been demonstrated that can provide the exact solution to BAO, and (2) no study has validated an optimization method for solving BAO by benchmarking with the optimal BAO solution (e.g., via ES), both of which will be addressed by this work.

**Methods:** This work considers BAO for proton therapy, e.g., the selection of 2 to 4 beam angles for IMPT. The optimal BAO solution is obtained via ES and serves as the ground truth. A new BAO algorithm, namely angle generation (AG) method, is proposed, and demonstrated to provide nearly-exact solutions for BAO in reference to the ES solution. AG iteratively optimizes the angular set via group-sparsity (GS) regularization, until the planning objective does not decrease further.

**Results:** Since GS alone can also solve BAO, AG was validated and compared with GS for 2-angle brain, 3-angle lung, and 4-angle brain cases, in reference to the optimal BAO solutions obtained by ES: the AG solution had the rank (1/276, 1/2024, 4/10626), while the GS solution had the rank (42/276, 279/2024, 4328/10626).

**Conclusions:** A new BAO algorithm called AG is proposed and shown to provide substantially improved accuracy for BAO from current methods with nearly-exact solutions to BAO, in reference to the ground truth of optimal BAO solution via ES.

### Keywords

beam angle optimization; group sparsity; IMPT

yong.long@sjtu.edu.cn .

[#]Co-first authors with equal contribution

Conflict of Interest Statement

None.

## 1. Introduction

Radiation therapy (RT) for treating cancer patients is delivered to tumor targets often with multiple beam angles, such as IMRT [1] and IMPT [2]. Beam angle optimization (BAO) refers to the optimization problem of selecting an optimal subset of beam angles from all available beam angles in terms of treatment plan quality [3]. Especially for proton RT, IMPT treatment planning often consists of a few beam angles [2], where the change of any of these angles may have a great impact on treatment plan quality, e.g., the avoidance of a specific organ-at-risk (OAR). Therefore, the selection of appropriate beam angles is important for proton RT. However, the BAO problem is a NP-hard problem [53], where the computational complexity grows exponentially with respect to the number of angles. Although exhaustive search (ES) can solve BAO exactly by exploring all possible combinations, ES is very time-consuming and practically infeasible. This motivates the development of efficient BAO methods, for which this work will propose a new BAO algorithm called the angle generation (AG) method.

Existing BAO methods can be classified into non-iterative methods [4–6], stochastic methods [7–10], and local-search methods [11–16]. All methods are heuristic and AG is a local-search method. In terms of angle update strategy, Ehrgott et al [12] starts with a desired number of beam angles and then iteratively replaces one angle with improved scoring function value. In contrast, one can also start with many angles and subsequently decrease to the desired number [11] or start with an empty set and subsequently add to the desired number [13]. In this sense, AG is similar to [12]. This choice is motivated by that empirical templates are often available for various tumor sites from clinical experiences [2], such as two lateral opposed beams (90°, 270°) for prostate, and X beam arrangement (45°, 135°, 225°, 315°) for head-and-neck, and these existing templates should boost the BAO quality if serving as initial guesses given the combinatorial nature of BAO. However, AG is different from [12] in three aspects: first, the initial guess does not have to be a set of equidistant beams, which was desirable in [12] since the candidate angle for update was searched around existing angles; second, there is no restriction of the search region for a candidate angle, which is global, i.e., to find a complementary angle to the stationary angles in terms of objective function values via a group-sparsity (GS) optimization with respect to the entire angle set; third, the scores are objective function values from IMPT that are consistent with the angle replacement procedure by solving GS-regularized IMPT (GS-IMPT).

On the other hand, GS can be used to directly solve BAO [16–18,60], e.g., the heterogeneity-weighted GS with novel sensitivity regularization for robust optimization [60]. That is, with sufficiently large GS term, the number of beams decreases during iterations and the algorithm terminates when the number of beams reaches a desired number of beams. However, although GS alone works well for selecting a non-small set of angles (e.g., 10 out of 24), but not for selecting a small set of angles (e.g., 3 out of 24) [16]. Another problem with the direct GS approach is that a $M_1$-beam BAO set is always a subset of $M_2$-beam BAO set for $M_1 \leq M_2$ using alternating direction method of multipliers (ADMM) [31]. While the latter is not the case for the proximal method [54], e.g., the so-called Fast Iterative

Shrinkage-Thresholding Algorithm (FISTA) [47], the proximal solution is highly sensitive to the choice of parameters, which can be problematic in choosing the optimal solution. Note that ADMM and proximal methods are two commonly used optimization methods for nonsmooth and constrained problems, such as GS-IMPT.

Although AG is a general method, this work will focus on IMPT for proton RT, since (1) IMPT often consists of 2 to 4 beam angles, each of which can be significant, and (2) IMPT is becoming the dominant method for delivering proton RT [2]. In comparison, IMRT for photon RT often consists of 7–11 beam angles, and many patients are treated with VMAT instead of IMRT [1]. Therefore, there is a greater need of BAO for IMPT, which is the focus of this study.

To the best of our knowledge, (1) no optimization method has been shown that can solve BAO exactly, and (2) no study has validated a BAO algorithm by benchmarking with optimal BAO solution. This work will aim to address these two unsolved problems with the proposed AG method. That is, the AG method will be developed for BAO and shown to provide nearly-exact solutions to BAO, compared to optimal BAO solutions via ES as the ground truth.

## 2. Methods and Materials

The BAO problem is formulated in Section 2.1. The AG method for solving BAO is introduced in Section 2.2. The angle replacement (AR) algorithm for updating beam angles during AG, by solving group-sparsity-regularized IMPT (GS-IMPT) problem, is presented in Section 2.3. ADMM based optimization algorithm for solving GS-IMPT is developed in Section 2.4. The validation plan of AG is outlined in Section 2.5, using the optimal BAO solution via ES as the ground truth. In terms of connections between these algorithms, (1) AG (Algorithm 1 in Section 2.2) is the overall algorithm that solves BAO; (2) AR (Algorithm 2 in Section 2.3) is Step 2 of AG; (3) ADMM (Algorithm 3 in Section 2.4) solves the GS-IMPT problem in Step 1 of AR.

### 2.1. Beam Angle Optimization

Let us consider the IMPT optimization problem for a set of beam angles $S=\{\theta_i|i\leq B\}$, where $B$ is the number of beam angles available for optimization. The optimization is with respect to proton spot weights $x=\{x_i|i\leq B\}$ with $x_i=\{x_{ij}|j\leq N_i\}$, where $N_i$ is the number of proton spots for the $i^{\text{th}}$ angle.

The general form of total planning objective $F$ including dose-volume constraints [19–23] is

$$F_S(x) = \frac{1}{2}\sum_{m=1}^{C} w_m \sum_{k\in\Omega_m}\Big(\sum_{\theta_i\in S}\sum_{j=1}^{N_i} D_{k,ij}x_{ij} - d_m\Big)^2 = \frac{1}{2}\Big\|\sum_{\theta_i\in S} D_i x_i - d\Big\|^2 = \frac{1}{2}\|Dx-d\|^2 \quad (1)$$

In Eq. (1), there are $C$ planning objectives: $w_m$ is the objective weight, $d_m$ is the dose constraint, $\Omega_m$ is the patient volume or active volume for the $m^{\text{th}}$ objective. For dose-volume

constraints, the active volume $\Omega_m$ depends on dose map within the structure corresponding to the $m^{\text{th}}$ objective, which in turn depends on $x$. Therefore, the optimization with dose-volume constraints is nonconvex, for which iterative convex relaxation method [21–24] can handle and iteratively update $\Omega$ based on $x$ during iterations. $D$ is the dose influence matrix, $D_i$ corresponds to the $i^{\text{th}}$ angle and $D_{k,ij}$ is the dose contribution from $j^{\text{th}}$ spot $x_{ij}$ of the $i^{\text{th}}$ beam to the $k^{\text{th}}$ patient voxel. To simplify the notation, $F$ is denoted formally by the last least-square term in Eq. (1).

In order for IMPT to be deliverable, we consider minimum-monitor-unit (MMU) optimization problem [25–29] on $x$, i.e.,

$$\begin{aligned} f_s(x^*) &= \min_x F_s(x) \\ s.t.\ x &\in \{0\} \cup [G_{\min}, +\infty) \end{aligned}. \tag{2}$$

Here $G_{min}$ is the MMU threshold and the weight of deliverable nonzero $x$ has to be at least $G_{min}$. In Eq. (2), $x^*$ and $f$ denote spot weights and total objective respectively after optimization.

Then the BAO problem is to select a subset of $M$ angles from all available angles, with the best plan quality, i.e., the smallest optimization objective.

$$S^* = \arg\min_{S \in \mathbb{S}} f_S. \tag{3}$$

Here $\Theta$ denotes the set of all available angles for optimization, and $\mathbb{S}$ is a set containing all possible selections of $M$ angles from $\Theta$, i.e., $\mathbb{S} = \{S \subset \Theta | \#(S) = M\}$. BAO is a NP-hard problem. ES can solve BAO exactly by exploring all possible combinations. Although ES is very time-consuming and practically infeasible. ES works for BAO problems of small number of angles, which will be considered in this work for the validation purpose. In the following, we will introduce a new method AG for solving BAO. The effectiveness of AG for BAO will be validated against the optimal BAO solution obtained by ES, which solves IMPT Eq. (2) for all possible sets $S$ in $\mathbb{S}$ and thus provides the ground truth for the BAO problem. ADMM based optimization algorithm for solving the IMPT problem is provided in Appendix A.

### 2.2. Angle Generation Method

To solve BAO Eq. (3), i.e., by finding the best $M$ angles with the smallest planning objective, the AG method starts with an initial set $S_0 = \{\theta_i^0 | i \le M\}$ of $M$ angles, and then iteratively updates this set, e.g., $S_n = \{\theta_i^n | i \le M\}$, until no better plan can be found. During each iteration of AG, a current angle called "the pivot angle" in the set $S_n$ is substituted be an angle called "the replacement angle" outside of $S_n$. Specifically, the $n^{th}$ AG iteration consists of the following steps.

- **Step 1:** The angle $a_n = \theta_i^n \in S_n$, $i \equiv n \pmod M$ is selected as "the pivot angle";
- **Step 2:** "The replacement angle" $a_{n+1} \notin S_n$ is identified via the AR algorithm (Section 2.3);

- **Step 3:** $S_n^{'} = (S_n \backslash \{a_n\}) \cup \{a_{n+1}\}$ is generated by replacing $a_n$ with $a_{n+1}$, and $S_{n+1}$ is the smaller of $S_n$ and $S_n^{'}$ in terms of IMPT optimization objective;
- **Step 4:** The stopping criteria is given by either $n \leq n_{max}$ for a given max iteration number $n_{max}$ or $S_n = S_{n-M+1}$ that implies $S_n$ is optimal and no better angle set than $S_n$ will be found.

The AG algorithm is summarized as follows, with corresponding aforementioned steps.

**Algorithm 1:** Angle Generation (AG) algorithm

**while** $S_n \neq S_{n-M+1}$ and $n \leq n_{max}$ **do**
  $a_n = \theta_i^n, i = n \pmod M$ **(Step 1)**
  $a_{n+1} = \mathrm{AR}(a_n)$ **(Step 2)**
  $S_n^{'} = (S_n \backslash \{a_n\}) \cup \{a_{n+1}\}$ **(Step 3)**
  **if** $f_{S_n^{'}} \leq f_{S_n}$ **then (Step 4)**
    $S_{n+1} = S_n^{'}$
  **Else**
    $S_{n+1} = S_n$
  **end if**
  $n = n + 1$
**end while**

## 2.3. Angle Replacement Algorithm

For the purpose of BAO, i.e., to select a small set of $M$ angles from Θ of many angles, mathematically the GS regularization can minimize the number of angles (i.e., to sparsify with respect to the index $i$ in Eq. (1)), but not the number of spots per angle (i.e., not to sparsify with respect to the index $j$ in Eq. (1)). Therefore, the GS regularization has been utilized for solving BAO directly [17, 18, 20].

However, the GS alone is not sufficient, because (1) the optimal GS solution may not be optimal for BAO, owing to the nonconvexity of BAO; (2) for ADMM, although ADMM is insensitive to the values of GS regularization parameter $\lambda$, the optimal set of $M_1$ angles is always a subset of $M_2$ angles for $M_1 \leq M_2$, which is not true from the perspective of treatment planning; (3) while (2) is not for the proximal method (e.g., FISTA), the proximal method is highly sensitive to the values of $\lambda$, which makes it difficult in choosing the optimal solution unless going through the exhaustive searching with respect to $\lambda$.

In contrast, for AG, we will show (1) the optimal set of $M_1$ angles does not have to be a subset of $M_2$ angles for $M_1 \leq M_2$, which is an advantage from FISTA; (2) AG is not sensitive to the values of $\lambda$ which is an advantage from ADMM; (3) AG can outperform GS and achieve nearly-optimal BAO solution, in reference to the optimal BAO solution via ES as the ground truth.

Although the step of replacement angle selection during AG utilizes GS, the difference is that the GS regularization in AG only applies to "the working set" $W_n = \Theta \backslash T_n$ instead of the entire Θ, where $T_n = S_n \backslash \{a_n\}$ is "the fixed set" by removing the pivot angle $a_n$ from the current set $S_n$. That is the fixed set $T_n$ is not penalized by GS and thus remains to contribute to the plan optimized with GS regularization, regardless of what is left or removed in the working set $W_n$ by GS. This is motivated by the purpose of AG, i.e., to find the replacement angle that is better than the pivot angle, in terms of their compatibility to the rest of angles in

$S_n$. That is, AG utilizes GS to find the best candidates for the replacement angle $a_{n+1}$ that are most complementary to the current set $S_n$ excluding $a_n$, in terms of optimal plan quality.

Mathematically, the GS-IMPT problem is the following:

$$x^* = \arg\min_x F_\Theta(x) + \sum_{i \in W} \lambda_i \|x_i\|_2^p, \tag{4}$$

where $\|x_i\|_2^p = (\sum_j x_{ij}^2)^{p/2}$ is the L2, $p$-norm GS regularization.

In Eq. (4), the total plan objective $F_\Theta$ is from all angles in $\Theta$, while the GS regularization term with L2, $p$-norm ($0<p\leq 1$) is only with respect to the working set $W$, e.g., $W_n$ during the $n^{th}$ AG iteration as explained earlier. We have compared and found negligible difference between $p=1/2$ and $p=1$, and therefore used $p=1$ that is simpler in this work. On the other hand, the GS parameter $\lambda_i$ is to balance the contribution from each beam per spot [18], i.e.,

$$\lambda_i = \lambda \left( \frac{\|D_{PTV,i}\mathbf{1}\|_2}{N_i} \right)^{\frac{p}{2}}, \tag{5}$$

where $D_{PTV,i}$ is the dose influence matrix from the $i^{th}$ beam angle to PTV for beam angle $b$, and $\lambda$ is a tuning parameter that is constant for all angles.

It is a concern that when larger $\lambda$ is needed to sparsify the angles, the contribution from planning objective $F$ in Eq. (4) is less and therefore the iterative process for solving Eq. (4) may not sufficiently account of plan quality. Therefore, we consider another formulation of GS-IMPT [30], where the planning objective $F$ is constantly controlled under a small value $\varepsilon$, i.e.,

$$\begin{aligned} x^* = &\arg\min_x \sum_{i \in W} \lambda_i \|x_i\|_2^p \\ &s.t.\ F_\Theta(x) \leq \varepsilon \end{aligned}. \tag{6}$$

However, we have compared and found negligible difference between Eq. (4) and Eq. (6), and therefore used the original formulation Eq. (4) in this work.

The AR step of AG, i.e., Step 2 in Algorithm 1, consists of the following steps:

- **Step 1:** $x^*$ is derived by solving the GS-IMPT problem Eq. (4) for the working set $W_n$, which is determined by $a_n$, i.e., $W_n = \Theta \backslash T_n$ with $T_n = S_n \backslash \{a_n\}$;
- **Step 2:** A candidate angle set $C = \{\theta_i^n \in W_n | i \leq N_c\}$ is selected that consists of the largest $N_c$ angles in the L2, $p$-norm value ($N_c = 3$ in this work);
- **Step 3:** The IMPT problems corresponding to each angle set $Q_i = T_n \cup \{\theta_i^n\}$ for $\theta_i^n \in C$ are solved, for which optimized objective value is denoted by $f_{Qi}$;
- **Step 4:** $\theta_i^n$ with the minimal $f_{Qi}$ is selected as the replacement angle $a_{n+1}$.

The AR algorithm is summarized as follows, with corresponding aforementioned steps.

**Algorithm 2:** Angle Replacement (AR) algorithm for $a_{n+1} = \mathrm{AR}(a_n)$

$x^* = \text{GS-IMPT}(a_n)$ **(Step 1)**

$C = \{\theta_i^* \in W_n \mid \max \left\|x_i^*\right\|_2^p, \#C = N_s\}$ **(Step 2)**

**for** $\theta_j^* \in C$ **(Step 3)**

$Q_t = T_n \cup \{\theta_t^*\}$

$f_{Q_t} = \mathrm{IMPT}(Q_t)$

**end for**

$a_{n+1} = \{\theta_t^* \in C \mid \max f_{Q_t}, \#C = 1\}$ **(Step 4)**

## 2.4. Solving GS-IMPT via ADMM

This section presents the solution algorithm for GS-IMPT problem Eq. (4) using ADMM [31–33].

ADMM is a general optimization method for solving a variety of problems, and is particularly suitable for dealing with nonconvex constraints and non-differentiable regularization (e.g., GS) through the variable splitting. For example, we have developed ADMM algorithms for solving convex problems with linear data fidelity, such as image reconstruction in 4D CT [34], spectral CT [35], MRI [36], 4D cone-beam CT [37], mega-voltage CT [38], CT [39], and breast CT [40, 41], and nonconvex problems with nonlinear data fidelity, such as image reconstruction in quantitative photoacoustic tomography [42–44], and cine cone-beam CT [45, 46].

Note that IMPT and GS-IMPT can also be efficiently solved by the state-of-the-art proximal method such as FISTA [18, 47]. However, we have compared and found negligible difference between ADMM and FISTA for the purpose of AG. ADMM based solution algorithm is provided here, while FISTA based solution algorithm is provided in [18].

Eq. (4) is recapped here for the convenience of the presentation

$$\begin{aligned} &\min_x \|Dx - d\|^2 + \sum_{i \in W} \lambda_i \|x_i\|_2^p, \\ &s.t.\ x \in \{0\} \cup [G_{\min}, +\infty) \end{aligned} \tag{7}$$

where the GS term is only applied to the angles in the work set $W$.

Similar to Appendix A, a dummy variable $z_1$ is introduced, i.e., $z_1 = x$, to decouple the MMU constraint. Another dummy variable $z_2$ is introduced, i.e., $z_2 = x$, to decouple the GS term. As a result, the constrained problem Eq. (7) can be decoupled into an unconstrained problem of $x$ with planning objectives, and two constrained problems of $z_1$ and $z_2$ respectively that have no planning objectives and thus can be solved analytically.

We start with the augmented Lagrangian function of Eq. (7)

$$\begin{aligned} L(x, z_1, u_1, z_2, u_2) = \|Dx - d\|^2 + \sum_{i \in W} \lambda_i \|z_{2i}\|_2^p + \mu_1 \|x - z_1 + u_1\|_2^2 + \mu_2 \|x - z_2 + u_2\|_2^2 \\ s.t.\ z \in \{0\} \cup [G_{\min}, +\infty) \end{aligned} \tag{8}$$

.

Here $u_1$ is the dual variable of $z_1$, and $\mu_1$ is the relaxation parameter for the constraint $z_1=x$; $u_2$ is the dual variable of $z_2$, and $\mu_2$ is the relaxation parameter for the constraint $z_2=x$.

From Eq. (8), the ADMM solution to Eq. (7) consists of the following iterations

$$\begin{cases} x^{n+1} = \underset{x}{\operatorname{argmin}} L(x, z_1^n, z_2^n, u_1^n, u_2^n) \\ z_1^{n+1} = \underset{z_1 \in \{0\} \cup [G_{\min}, +\infty)}{\operatorname{argmin}} L\left(x^{n+1}, z_1, z_2^n, u_1^n, u_2^n\right) \\ z_2^{n+1} = \underset{z_2}{\operatorname{argmin}} L\left(x^{n+1}, z_1^{n+1}, z_2, u_1^n, u_2^n\right) \\ u_1^{n+1} = u_1^n + x^{n+1} - z_1^{n+1} \\ u_2^{n+1} = u_2^n + x^{n+1} - z_2^{n+1} \end{cases}. \tag{9}$$

In Eq. (9), the $x$-problem and $z_1$-problem can be handled similarly to Eq. (A4) and (A5) respectively; the $z_2$-problem is separable and has analytic solution $z_2^{n+1}=S_t(x^{n+1}+u_2^n)$ with $t=\lambda/2\mu_2$ and $S_t$ given by the followings for $p=1$ and $p=1/2$ [48] respectively.

$$S_{t,1}(x) = x - x \cdot \min\left(\frac{t}{\|x\|_2}, 1\right) \tag{10}$$

$$S_{t,1/2}(x) = \begin{cases} 0, & t\|x\|_2^{-1.5} > \frac{2\sqrt{6}}{9} \\ x\sqrt{\frac{2}{\sqrt{3}}\sin\left(\frac{1}{3}\left(\arccos\left(\frac{3\sqrt{3}}{4}t\|x\|_2^{-1.5}\right) + \frac{\pi}{2}\right)\right)}, & \text{otherwise} \end{cases} \tag{11}$$

Note that the GS updates Eq. (10) and (11) are group-wise instead of element-wise: that is the spots per beam angle are jointly updated together.

The ADMM algorithm for solving GS-IMPT is summarized as follows.

**Algorithm 3:** $x^* = \text{GS-IMPT}(a_n)$

**while** $n \le n_{\max,GS-IMPT}$ **do**

$x^{n+1} = \left(D^T D + \mu_1 + \mu_2\right)^{-1}\left(D^T d + \mu_1\left(z_1^n - u_1^n\right) + \mu_2\left(z_2^n - u_2^n\right)\right)$ (x-problem in Eq. (9) solved by CG)

$z_1^{n+1} = S(x^{n+1} + u_1^n)$ (Eq. (A5))

$z_2^{n+1} = S_{1/2,\mu}(x^{n+1} + u_2^n)$ (Eq. (10) or (11))

$u_1^{n+1} = u_1^n + x^{n+1} - z_1^{n+1}$

$u_2^{n+1} = u_2^n + x^{n+1} - z_2^{n+1}$

$n = n + 1$

**end while**

## 2.5. Comparative Planning Study

We considered BAO for a brain case with $M=2$ ("2-angle brain") and $M=4$ ("4-angle brain") respectively, and a lung case with $M=3$ ("3-angle lung"). First we consider the scenario with the full angular set $\Theta$ of 24 evenly-distributed coplanar beam angles. The BAO problem was to select the best $M$ angles from $\Theta$. The choice of a relatively small number $B=24$ is because it allows us to solve all possible IMPT solutions via ES to find the optimal BAO solution. On the other hand, as one reduces angular spacing to increase number of beam angles, the

angles get closer to each other and the incremental gain of plan quality with increasing number angles starts to diminish.

For the validation purpose, the optimal BAO solution was attained using ES by solving all possible combinations of angles, which serves as the group truth for BAO. ES was equivalent to solve 24×23/2=276 IMPT problems for 2-angle brain case, 24×23×22/(3×2)=2024 IMPT problems for 3-angle lung case, 24×23×22×21/(4×3×2)=10626 IMPT problems for 4-angle brain case, for which the best solution was selected as the ES solution and also the ground truth.

In benchmark to the ES solution, the proposed angle generation method ("AG") was validated in comparison with the GS method ("GS") that directly solved BAO [16–18]. In terms of computational time, AG (equivalent to solving a few IMPT problems) or GS (equivalent to solving 1 IMPT problem) was negligible compared to ES. For optimizing plan quality, after GS or AG was solved for the best $M$ angle, the IMPT problem for selected M angles was solved as the final solution. Thus, the solutions from GS, AG and ES were consistent with each other as they were from the same IMPT solution algorithm, with the only difference in the choice of beam angles. Moreover, all plans were normalized after optimization with D95=100% at PTV.

Next we also consider BAO with a relatively large number $B=72$. In this case, ES was not performed due to the size of the problem, and AG was compared with GS. To demonstrate the generality of AG, the solution algorithms for solving IMPT and GS-IMPT in this case of $B=72$ were based on FISTA [18, 47], which was primarily based on ADMM for $B=24$ [31–33].

MatRad [49] was used to generate dose influence matrices using 5 mm lateral spacing, and 3 mm longitudinal spacing (energy), on 3 mm-resolution dose grid. All the experiments were performed with i9-10900K CPU, for which AG took 7–25 minutes.

Robust optimization was also considered for the lung case with 5mm for setup uncertainty and 3.5% for range uncertainty via probabilistic formulation [52,21]. The plan was normalized with respect to the nominal case. The dose and DVH plots in Fig. 12 and 13 were based on average dose of all uncertainty scenarios.

In the tables, the conformity index (CI) is defined as $V_{100}^2/(V\times V'_{100})$ ($V_{100}$: PTV volume receiving ≥100% of prescription dose; V: PTV volume; $V'_{100}$: total volume receiving ≥100% of prescription dose). The value of CI is between 0 and 1, with optimal being 1. $S^*$ is the set of selected angles; $f$ is the optimized plan objective value; rank is the ranking of $f$ in all ES solutions.

## 3. Results

### 3.1. AG v.s. ES: nearly-exact BAO solution via AG

ES was used to solve IMPT problems Eq. (2) for all possible combinations of angles, i.e., in $\mathbb{S} = \{S \subset \Theta | \ \#(S) = M\}$. IMPT planning objectives of all plans were sorted in the ascending order in Fig.1, for which the smaller rank corresponds to smaller objective values and thus

better plan quality. Since ES solved BAO Eq. (3) exactly, the optimal ES solution was the optimal BAO solution, which had the smallest rank (i.e., 1). As shown in Fig. 1 and Table 1, the AG solution was ranked 1 for both 2-angle brain and 3-angle lung, and 4 for 4-angle lung. Therefore, AG provided the nearly-exact solution to BAO. In contrast, the GS solution was ranked 42/276 (i.e., 42 of 276 total plans) for 2-angle brain, 279/2024 for 3-angle lung, and 4328/10626 for 4-angle brain respectively.

### 3.2. AG v.s. GS: AG improved plan quality from GS

Compared to GS, AG improved rank and plan objective values in Fig. 1 and Table 1, which implies AG improved plan quality from GS. The improved plan quality via AG from GS was also confirmed by comparison of dosimetric parameters in Table 1, comparison of DVH plots in Fig. 2 and comparison of dose plots in Fig. 3–5. For example, as shown in Table 1, compared to GS, AG improved target dose conformity quantified by CI, from 0.6 to 0.62 for 2-angle brain, from 0.9 to 0.92 for 3-angle lung, and from 0.65 to 0.68 for 4-angle brain respectively.

### 3.3. Solution convergence of AG

By the design of AG algorithm, the solution converges with non-increasing plan objectives during AG iterations, which was experimentally confirmed and presented in Fig. 6. Based on Algorithm 1, AG found the optimal solution and terminated the optimization in 3, 5 and 14 AG iterations respectively for 2-angle brain, 3-angle lung, and 4-angle lung respectively, as shown in Fig. 6.

### 3.4. AG was insensitive to the initial guess

To test the sensitivity of AG to the initial guess ($B=24$), for 2-angle brain, we ran AG with 8 randomly-chosen initial angular set $S_0$, i.e., (1,2), (3,9), (5,18), (2,24), (15,19), (11,13), (17,22) and (7,9), which all converged to (5,20) that ranked 1/276; for 3-angle lung, we ran AG with 8 randomly-chosen $S_0$, i.e., (1,2,3), (1,3,9), (1,10,18), (2,17,24), (3,15,18), (7,11,13), (12,17,22) and (2,7,9), which all converged to (14,19,22) that ranked 1/2024; for 4-angle brain, we ran AG with 8 randomly-chosen $S_0$, i.e., (1,2,3,4), (1,15,19,24), (1,5,7,20), (1,9,10,13), (3,7,12,18), (5,8,11,14), (6,12,17,21) and (6,7,19,20), are all (5,7,12,20), which all converged to (5,7,12,20) that ranked 4/10626.

To test the sensitivity of AG to the initial guess (B=72), for 2-angle brain, we ran AG with 16 randomly-chosen initial angular set S0, i.e., (1,2), (18,55), (55,56), (9,42), (4,7), (2,15), (25,56), (16,30), (44,55), (38,39), (70,71), (9,44), (5,18), (33,71), (14,28) and (6,66), and these all converged to (13,58), i.e., (60°,285°), which is the same as the result with B=24; for 3-angle lung, we ran AG with 16 randomly-chosen S0, i.e., (1,25,49), (2,31,41), (3,21,51), (7,17,66), (7,19,29) (7,32,72), (9,36,69), (10,56,67), (10,59,65), (12,13,40), (12,39,60), (12,70,71), (13,23,39), (13,29,48), (13,48,49) and (14,55,57) , and these converged to (37,46,60), (42,51,60), (42,51,60), (42,51,60), (42,51,60), (42,51,60), (42,51,60), (42,51,60), (42,55,65), (42,51,60), (42,51,60), (42,51,60), (42,51,60), (37,46,60), (42,51,60) and (42,51,64) respectively, which are very close to the corresponding B=24 result (14,19,22), i.e., (195°,270°,315°); for 4-angle brain, we ran AG with 16 randomly-chosen S0, i.e., (5,14,18,31), (18,29,30,56), (14,22,40,54), (23,32,33,36), (16,22,43,45),

(15,17,34,61), (11,13,17,32), (19,20,30,43), (22,24,31,37), (36,42,53,67), (17,36,38,45), (31,57,66,69), (28,32,55,57), (19,37,51,54), (13,23,39,57) and (13,19,38,56), and these converged to (13,19,31,56), (13,19,30,57), (14,22,40,54), (13,19,36,56), (13,19,45,58), (13,19,34,58), (13,19,32,56), (13,19,30,57), (13,21,37,58), (13,19,36,56), (13,19,38,56), (13,19,31,56), (13,19,28,57), (13,19,37,56), (13,19,39,58) and (13,19,38,56) respectively, which are very close to the corresponding B=24 result (5,7,12,20), i.e., (60°,90°,165°,285°).

Therefore, the AG solution was insensitive to the initial guess S0, and the results with B=72 were similar to those with B=24.

### 3.5. ADMM v.s. FISTA for GS

GS can be solved by either ADMM or FISTA. The comparison of ADMM and FISTA are presented in Table 2 and Fig. 7. While ADMM was insensitive to $\lambda$ except for large values of $\lambda$, FISTA was sensitive to $\lambda$ and the case-by-case tuning or ES seems to be needed to obtain good FISTA solutions. The sensitivity of FISTA to $\lambda$ can be problematic for choosing the optimal solution unless using ES with respect to $\lambda$, which is however time-consuming. In Table 2, the FISTA results were from the best scenario via ES with respect to $\lambda$: FISTA was worse than ADMM for 2-angle brain, but better than ADMM for 3-angle lung and 4-angle brain.

### 3.6. ADMM v.s. FISTA for AG

To investigate the algorithm dependence of AG, we ran AG with both ADMM and FISTA. The results summarized in Table 3 suggest that there was nearly no difference in solving AG by ADMM or FISTA, despite that (1) FISTA was sensitive to $\lambda$ and (2) a $M_1$-beam BAO set is always a subset of $M_2$-beam BAO set for $M_1 \leq M_2$ when solving GS via ADMM. Note that although (14,19.22) was ranked 1/2024 for ADMM based ES and (15,19.22) was ranked 1/2024 for FISTA based ES, both ADMM based AG and FISTA based AG found the optimal BAO solution. Moreover, the difference in plan quality between (14,19.22) and (15,19.22) was negligible, with both being 0.12 in two-decimal rounding.

### 3.7. Comparison of two GS-IMPT formulations for AG

The comparison of two GS-IMPT formulations Eq. (4) and Eq. (6) for AG was presented in Table 4, which suggest that Eq. (6) was not as expected to be better than, but turned out to be slightly worse than Eq. (4) for AG.

### 3.8. L2,1/2-norm v.s. L2,1-norm for AG

We compared the use of L2,1/2-norm ($p=1/2$) and L2,1-norm ($p=1$) in GS Eq. (4) for AG. The angular sets during iterations are summarized in Table 5, which shows that $p=1/2$ and $p=1$ found the same angular set, while $p=1/2$ had 1 fewer iteration than $p=1$.

### 3.9. AG v.s. GS with B=72

For $B=72$, despite of missing of ES as the ground truth, AG again improved plan quality from GS. For example, AG had smaller optimized plan objective values, better CI values, smaller OAR dosimetric parameters than GS as shown in Table 6. This improvement in plan quality was evident by comparing DVH plots in Fig. 8, and dose plots in Fig. 9–11.

### 3.10. AG v.s. GS under robust optimization

The comparison of dosimetric parameters, DVH, and dose plots are presented in Table 7, Fig. 12, and 13 respectively, using robust optimization for the lung case, for which ES was solved as the reference for the ground truth. As shown in Table 7, the AG solution was ranked 2, and thus AG still provided the nearly-exact solution to BAO under robust optimization, in reference to the ES solution. On the other hand, the GS solution was ranked 175 out of 2024. In comparison to GS, AG substantially improved the solution ranking, which was reflected in plan quality as presented in Table 7, Fig. 12, and 13, e.g., improved dose conformity by increasing CI from 0.63 to 0.66.

## 4. Discussion

While GS can directly solve BAO [16–18], the direct GS method via either ADMM or FISTA is imperfect considering physics perspective of treatment planning. With ADMM, for GS, the optimal set of $M_1$ angles is always a subset of $M_2$ angles for $M_1 \leq M_2$, while in practice the optimal set of $M_1$ angles does not have to be a subset of $M_2$ angles since BAO is a nonconvex problem. With FISTA, the optimal solution is highly sensitive to the GS regularization parameter $\lambda$ and therefore is practically difficult to achieve without going through exhaustive searching in $\lambda$, which is time-consuming. In comparison, the AG method makes more sense than the direct GS method, since (1) AG is insensitive to the value of $\lambda$, and (2) $M_1$-optimal angular set does not have to be a subset of $M_2$-optimal angular set for AG. For example, the optimal angular set via the direct GS method was (7,19), (7,19,20), (6,7,19,20) for 2-angle, 3-angle, 4-angle brain respectively, while the optimal angular set via AG was (5,20), (6,13,20), (5,7,12,20) for 2-angle, 3-angle, 4-angle brain respectively; on the other hand, the rank of GS solutions was 42/276, 521/2024, 4328/10626 respectively, while the rank of AG solutions was 1/276, 3/2024, 4/10626 respectively.

Although AG provided nearly-exact solutions to the BAO problem, we do not claim that AG exactly solves BAO, because BAO is a nonconvex NP-hard problem. The experiments in this work were limited to BAO to select 2–4 beam angles, i.e., M=2, 3, 4, and the number of all angles available for optimization was limited to 24, i.e., $B=24$, in order to run ES as the ground truth for BAO. However, it is expected that AG will still outperform the direct GS method for larger $M$ and larger $B$, e.g., the results with $B=72$ as shown in Section 3.9. For example, the angular set from the direct GS method can be used as initial guess for AG, and the plan objective will be non-increasing throughout the AG iterations as demonstrated in Fig. 6. Because of the nonconvexity of BAO, a thorough investigation of degree of improvement via AG from GS is to be carried out in a future work, e.g., with non-coplanar beams, for which however the exact BAO solution may not be available as the ground truth since the number of beams can be too big for ES to solve.

The AG method can be regarded as a general framework for BAO. Although in this work the update to the angular set is based on GS, other methods such as column generation method [50] or deep learning method [51] can be potentially used to update the angular set for AG. The integration of these methods to AG will be of interest as future works.

In current AG method, the update to the angular set is limited to one angle per AG iteration. Due to the nonconvex nature of BAO, it is possible that the update of multiple angles to the angular set per AG iteration may further improve the optimality of AG solutions. For example, for BAO to select $M$ angles, the strategy to update up to two angles to the angular set per AG iteration can include $M$ updates of one angle at a time and $M(M-1)/2$ updates of two angles at a time, which however should be computationally expensive.

As presented in Table 2–4 respectively, it was found that the performance of AG was not sensitive to the choice of the optimization algorithm (ADMM or FISTA), the GS-IMPT formulation (Eq. (4) or (6)), and the $p$ value (1/2 or 1). It was also found that the performance of AG was not sensitive to whether it was the MMU constraint ($G_{min}>0$) or the nonnegative constraint ($G_{min}=0$). Because the MMU constraint is nonconvex and the nonnegative constraint is convex, the presented results were with the MMU constraint instead of the nonnegative constraint to demonstrate the effectiveness of AG, even in the presence of nonconvex MMU constraint.

Spot-scanning arc therapy is an emerging proton modality [55] and rapidly developing [56–59] as a viable alternative to fixed-beam IMPT that can potentially offer a combination of advantages in plan quality and delivery efficiency, compared to fixed beam IMPT. But currently fixed-beam IMPT remains to be the mainstream for proton treatments, for which BAO is still relevant.

## 5. Conclusion

We have developed a new BAO algorithm called the AG method, which was shown to provide substantially improved accuracy for BAO from current methods (e.g., GS or FISTA) with nearly-exact solutions to BAO, in reference to the ground truth of optimal BAO solution via ES. It was found that (1) AG is relatively insensitive to the initial guess, compared to GS or FISTA, (2) AG is stable with respect to the choice of optimization engine (e.g., GS or FISTA) for solving the GS-IMPT problem.

## Acknowledgment

The authors are very thankful to the valuable comments from anonymous reviewers. This research is partially supported by NIH grants R37CA250921, R01CA261964, and a KUCC physicist-scientist recruiting grant.

## Appendix A: Solving IMPT via ADMM

This section presents ADMM based solution algorithm for the IMPT problem Eq. (2). For the convenience of presentation, the IMPT problem is recapped here

$$\begin{aligned} &\min_{x} \|Dx-d\|^2 \\ &s.t.\ x\in\{0\}\cup[G_{\min},+\infty) \end{aligned}. \tag{A1}$$

With ADMM, a standard trick to handle the constrained optimization problem such as Eq. (A1) is to introduce a dummy variable $z$ for the constraint, i.e., $z=x$, so that the constrained

problem Eq. (A1) can be decoupled into an unconstrained problem of $x$ with planning objectives and a constrained problem of $z$ that has no planning objectives and thus can be solved by an analytic formula, which will become clear shortly.

To derive ADMM solutions, we start with the augmented Lagrangian function of Eq. (A1)

$$\begin{aligned} L(x,z,u) &= \|Dx-d\|^2+\mu\|x-z+u\|_2^2, \\ s.t.\ z &\in \{0\}\cup[G_{\min},+\infty) \end{aligned} \tag{A2}$$

where $u$ is the dual variable of $z$, and $\mu$ is the relaxation parameter for the constraint $z=x$.

From Eq. (A2), the ADMM solution to Eq. (A1) consists of the following iterations

$$\begin{cases} x^{n+1} = \underset{x}{\operatorname{argmin}} L(x,z^n,u^n) \\ z^{n+1} = \underset{z\in\{0\}\cup[G_{\min},+\infty)}{\operatorname{argmin}} L\left(x^{n+1},z,u^n\right). \\ u^{n+1} = u^n + x^{n+1} - z^{n+1} \end{cases} \tag{A3}$$

As explained in [31,32], the purpose of $u$-update is to relax the need of increasing $\mu$ during iterations to enforce $z=x$: a fixed value of $\mu$ instead of an increasing sequence is sufficient for Eq. (A3) to converge, which avoids numerical instability owing to large $\mu$.

The $x$-problem of Eq. (A3) is a standard least-square problem, which has the optimal condition

$$\left(D^TD+\mu\right)x^{n+1} = D^Td+\mu(z^n-u^n), \tag{A4}$$

which is solved here by conjugate gradient (CG) method without explicit matrix inversion.

The $z$-problem of Eq. (A3) is separable and has point-wise analytic solution $z^{n+1}=S(x^{n+1}+u^n)$ with $S$ given by

$$S(x) = \begin{cases} 0, & x \le G_{\min}/2 \\ \max(G_{\min},x), & x > G_{\min}/2 \end{cases} \tag{A5}$$

The ADMM algorithm for solving IMPT is summarized as follows.

**Algorithm A1:** $f_x = \mathrm{IMPT}(S)$

**while** $n \le n_{\max\_IMPT}$ **do**

$x^{n+1} = \left(D^TD+\mu\right)^{-1}\left(D^Td+\mu\left(z^n-u^n\right)\right)$ (Eq. (A4) solved by CG)

$z^{n+1} = S(x^{n+1}+u^n)$ (Eq. (A5))

$u^{n+1} = u^n + x^{n+1} - z^{n+1}$

$n = n+1$

**end while**

## References

[1]. Bortfeld T, 2006. IMRT: a review and preview. Phys Med Biol. 51, R363. [PubMed: 16790913]
[2]. Paganetti H ed., 2018. Proton therapy physics. CRC press.

[3]. Ehrgott M, Güler Ç, Hamacher HW, et al. , 2010. Mathematical optimization in intensity modulated radiation therapy. Ann Oper Res. 175, 309–365.

[4]. Soderstrom S and Brahme A, 1992. Selection of suitable beam orientations in radiation therapy using entropy and Fourier transform measures. Phys Med Biol. 37, 911.

[5]. Pugachev A and Xing L, 2001. Pseudo beam's-eye–view as applied to beam orientation selection in intensity-modulated radiation therapy. Int J Radiat Oncol. 51, 1361–1370.

[6]. D D'Souza W, Meyer RR and Shi L, 2004. Selection of beam orientations in intensity-modulated radiation therapy using single-beam indices and integer programming. Phys Med Biol. 49, 3465. [PubMed: 15379026]

[7]. Stein J, Mohan R, Wang XH et al. , 1997. Number and orientations of beams in intensity-modulated radiation treatments. Med Phys. 24, 149–160. [PubMed: 9048355]

[8]. Haas OCL, Burnham KJ and Mills JA, 1998. Optimization of beam orientation in radiotherapy using planar geometry. Phys Med Biol. 43, 2179. [PubMed: 9725597]

[9]. Djajaputra D, Wu Q, Wu Y et al. , 2003. Algorithm and performance of a clinical IMRT beam-angle optimization system. Phys Med Biol. 48, 3191. [PubMed: 14579860]

[10]. Hou Q, Wang J, Chen Y, et al. , 2003. Beam orientation optimization for IMRT by a hybrid method of the genetic algorithm and the simulated dynamics. Med Phys. 30, 2360–2367. [PubMed: 14528958]

[11]. Das S, Cullip T, Tracton G, et al. , 2003. Beam orientation selection for intensity-modulated radiation therapy based on target equivalent uniform dose maximization. Int J Radiat Oncol. 55, 215–224.

[12]. Ehrgott M and Johnston R, 2003. Optimisation of beam directions in intensity modulated radiation therapy planning. OR Spectrum. 25, 251–264.

[13]. Meedt G, Alber M and Nüsslin F, 2003. Non-coplanar beam direction optimization for intensity-modulated radiotherapy. Phys Med Biol. 48, 2999. [PubMed: 14529207]

[14]. Engel K and Tabbert E, 2005. Fast simultaneous angle, wedge, and beam intensity optimization in inverse radiotherapy planning. Optim Eng. 6, 393–419.

[15]. Craft D, 2007. Local beam angle optimization with linear programming and gradient search. Phys Med Biol. 52, N127. [PubMed: 17374906]

[16]. Jia X, Men C, Lou Y et al. , 2011. Beam orientation optimization for intensity modulated radiation therapy using adaptive l2, 1-minimization. Phys Med Biol. 56, 6205–6222. [PubMed: 21891848]

[17]. O'Connor D, Yu V, Nguyen D et al. , 2018. Fraction-variant beam orientation optimization for non-coplanar IMRT. Phys Med Biol. 63, 045015. [PubMed: 29351088]

[18]. 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. [PubMed: 29394454]

[19]. Bortfeld T, Stein J, and Preiser K, 1997. Clinically relevant intensity modulation optimization using physical criteria. 12th Int. Conf. on the Use of Computers in Radiation Therapy, 1–4.

[20]. Wu Q, and Mohan R, 2000. Algorithms and functionality of an intensity modulated radiotherapy optimization system. Med. Phys, 27, 701–711. [PubMed: 10798692]

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

[22]. 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]

[23]. 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. [PubMed: 34800301]

[24]. 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.

[25]. 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]

[26]. 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]

[27]. 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]

[28]. 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]

[29]. Cai JF, Chen RC, Fan J et al. , 2022. Minimum-monitor-unit optimization via a stochastic coordinate descent method. Phys Med Biol. 67, 015009.

[30]. Pan L and Chen X, 2021. Group sparse optimization for images recovery using capped folded concave functions. SIAM J Imaging Sci. 14, 1–25.

[31]. Boyd S, Parikh N, Chu E et al. , 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn. 3, 1–122.

[32]. Goldstein T and Osher S, 2009. The split Bregman method for L1-regularized problems. SIAM J Imaging Sci. 2, 323–343.

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

[34]. Gao H, Cai JF, Shen Z et al. , 2011. Robust principal component analysis-based four-dimensional computed tomography. Phys Med Biol. 56, 3181. [PubMed: 21540490]

[35]. Gao H, Yu H, Osher et al. , 2011. Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM). Inverse Probl. 27, 115012. [PubMed: 22223929]

[36]. Gao H, Lin H, Ahn CB et al. , 2011. PRISM: A divide-and-conquer low-rank and sparse decomposition model for dynamic MRI. UCLA CAM Report, 11–26.

[37]. Gao H, Li R, Lin Y et al. , 2012. 4D cone beam CT via spatiotemporal tensor framelet. Med Phys. 39, 6943–6946. [PubMed: 23127087]

[38]. Gao H, Qi XS, Gao Y et al. , 2013. Megavoltage CT imaging quality improvement on TomoTherapy via tensor framelet. Med Phys. 40, 081919. [PubMed: 23927333]

[39]. Zhou W, Cai JF and Gao H, 2013. Adaptive tight frame based medical image reconstruction: a proof-of-concept study for computed tomography. Inverse Probl. 29, 125006.

[40]. Zhao B, Gao H, Ding H et al. , 2013. Tight-frame based iterative image reconstruction for spectral breast CT. Med Phys. 40, 031905. [PubMed: 23464320]

[41]. Ding H, Gao H, Zhao B et al. , 2014. A high-resolution photon-counting breast CT system with tensor-framelet based iterative image reconstruction for radiation dose reduction. Phys Med Biol. 59, 6005. [PubMed: 25230204]

[42]. Gao H, Zhao H and Osher S, 2010. Bregman methods in quantitative photoacoustic tomography. Cam Report. 30, 3043–3054.

[43]. Gao H, Osher S and Zhao H, 2012. Quantitative Photoacoustic Tomography, Mathematical Modeling in Biomedical Imaging II, Lecture Notes in Mathematics. Springer Berlin Heidelberg. 131–158.

[44]. Gao H, Feng J and Song L, 2015. Limited-view multi-source quantitative photoacoustic tomography. Inverse Probl. 31, 065004.

[45]. Cai JF, Jia X, Gao H et al. , 2014. Cine cone beam CT reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study. IEEE T Med Imaging. 33, 1581–1591.

[46]. Liu J, Zhang X, Zhang X et al. , 2015. 5D respiratory motion model based image reconstruction algorithm for 4D cone-beam computed tomography. Inverse Problems, 31, 115007.

[47]. Beck A and Teboulle M, 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci. 2, 183–202.

[48]. Möllenhoff T, Strekalovskiy E, Möller M et al., 2015, January. Low rank priors for color image regularization. In International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (pp. 126–140). Springer, Cham.

[49]. 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]

[50]. Romeijn HE, Ahuja RK, Dempsey JF et al. , 2005. A column generation approach to radiation therapy treatment planning using aperture modulation. SIAM J Optimiz. 15, 838–862.

[51]. Sadeghnejad Barkousaraie A, Ogunmolu O, Jiang S et al. , 2020. A fast deep learning approach for beam orientation optimization for prostate cancer treated with intensity-modulated radiation therapy. Med Phys. 47, 880–897. [PubMed: 31868927]

[52]. Unkelbach J, Alber M, Bangert M, et al. , 2018. Robust radiotherapy planning. Phys Med Biol. 63, 22TR02.

[53]. Bovet DP, Crescenzi P and Bovet D, 1994. Introduction to the Theory of Complexity. London: Prentice Hall.

[54]. Parikh N and Boyd S, 2014. Proximal algorithms. Foundations and trends® in Optimization, 1, 127–239.

[55]. 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. [PubMed: 27869083]

[56]. Gu W, Ruan D, Lyu Q, et al. , 2020. A novel energy layer optimization framework for spot-scanning proton arc therapy. Med Phys. 47, 2072–2084. [PubMed: 32040214]

[57]. Zhang G, Shen H, Lin Y, et al. , 2022. Energy layer optimization via energy matrix regularization for proton spot-scanning arc therapy. Med Phys. In Press.

[58]. Wuyckens S, Saint-Guillain M, Janssens G, et al. , 2022. Treatment planning in arc proton therapy: Comparison of several optimization problem statements and their corresponding solvers. Comput Biol Med, 105609. [PubMed: 35803749]

[59]. Engwall E, Battinelli C, Wase V, et al. , 2022. Fast robust optimization of proton PBS arc therapy plans using early energy layer selection and spot assignment. Phys Med Biol. 67, 065010.

[60]. Gu W, Neph R, Ruan D, et al. , 2019. Robust beam orientation optimization for intensity-modulated proton therapy. Med Phys. 46, 3356–3370. [PubMed: 31169917]

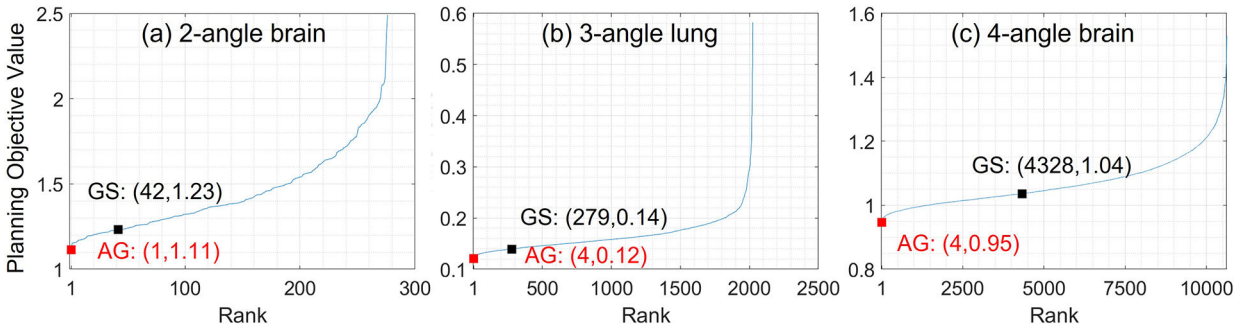


**Figure 1.**

Comparison of planning objective values optimized with AG and GS via ADMM for *B=24*. In each plot, planning objective values for all possible combinations are solved via ES and plotted in the ascending order. AG is marked in red and GS is marked in black, with their (rank, planning objective value) in the bracket.

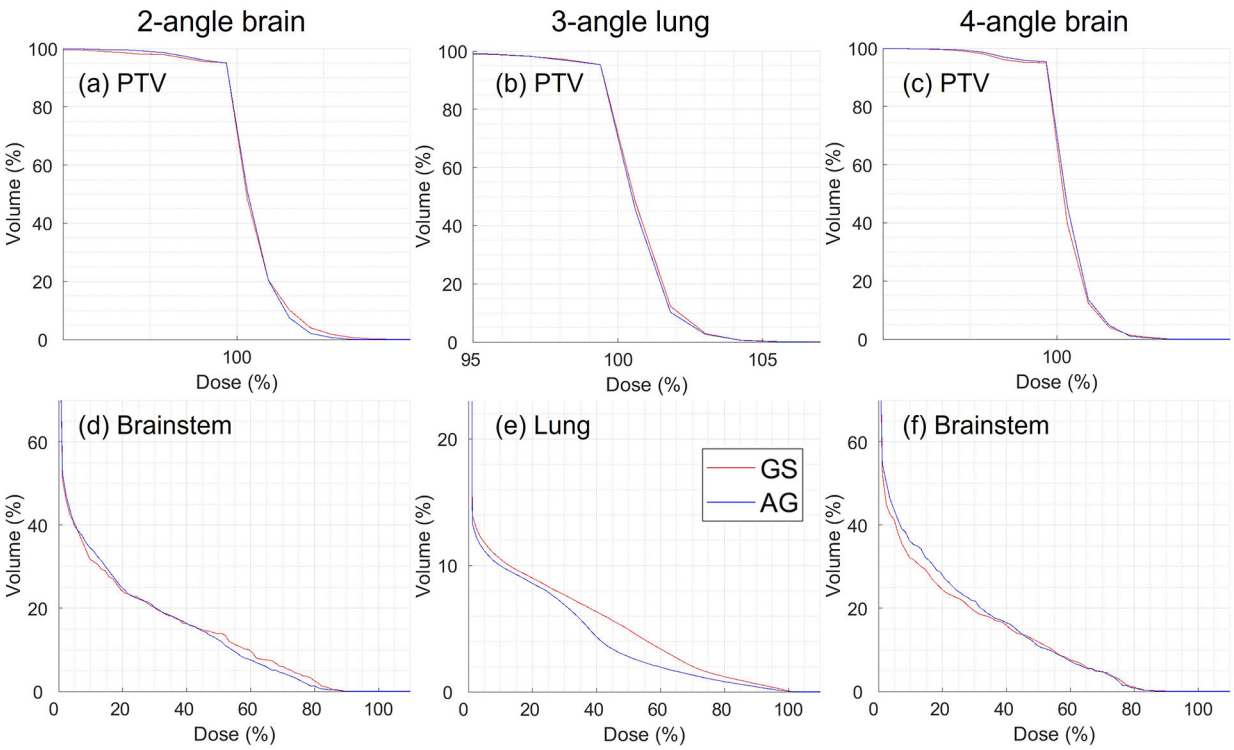


**Figure 2.**
Comparison of DVH plots between GS (red curves) and AG (blue curves) via ADMM with $B=24$.

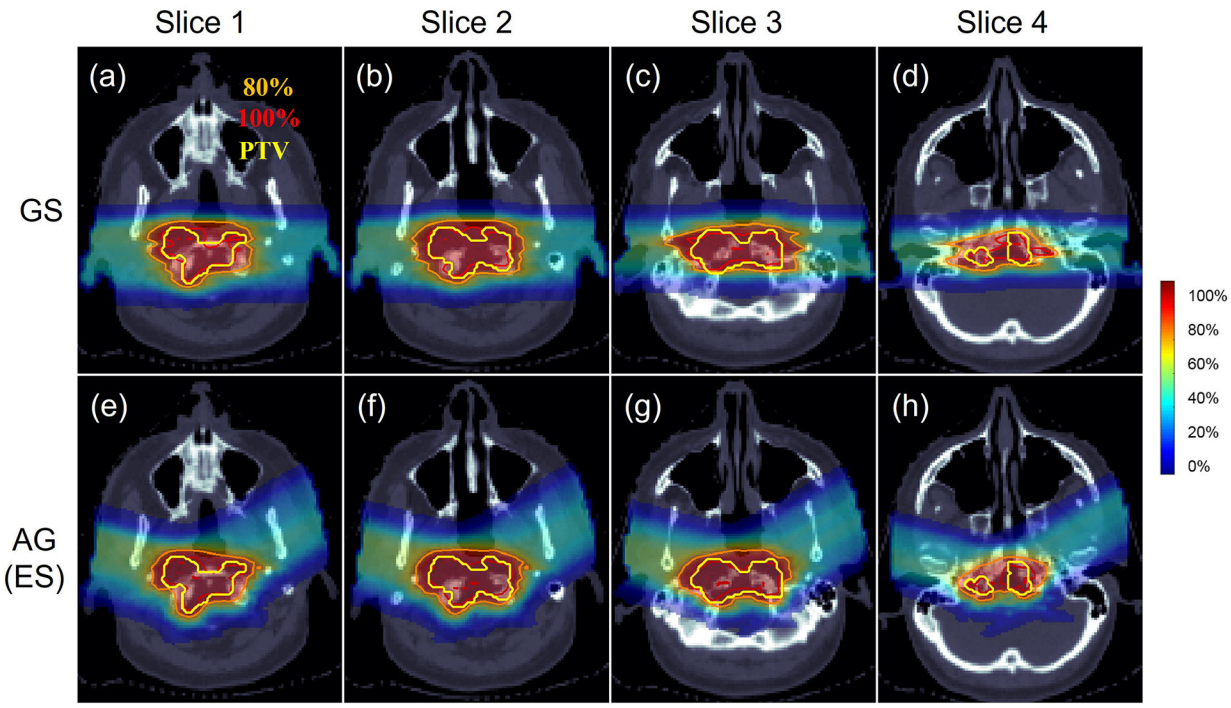


**Figure 3.**

Comparison of dose plots for 2-angle brain via ADMM with *B=24*, for which the optimal ES solution is available as the ground truth. ES is the same as AG, as AG exactly solved BAO. The dose plot window is [0%, 110%]. 100% isodose line, 80% isodose line and CTV are highlighted in dose plots.

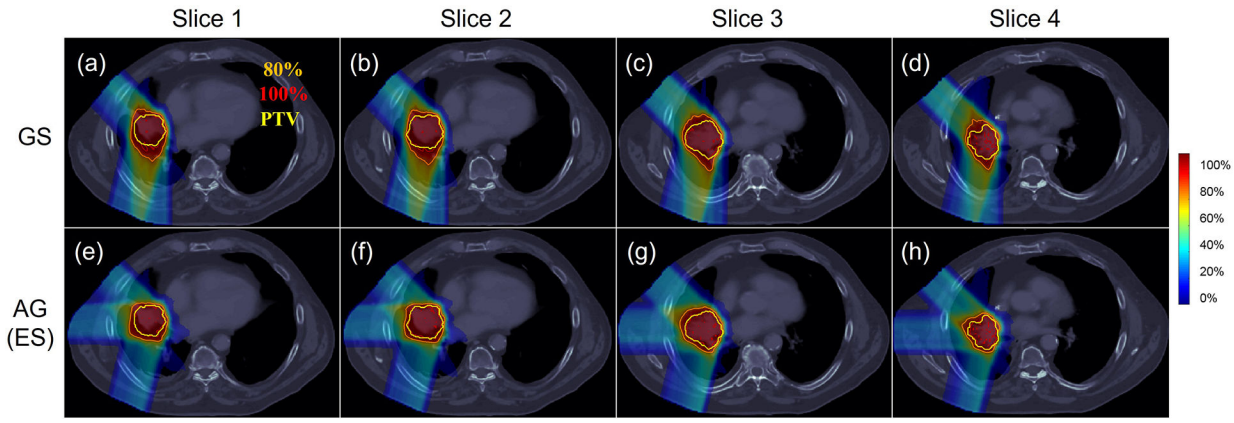


**Figure 4.**

Comparison of dose plots for 3-angle lung via ADMM with *B=24*, for which the optimal ES solution is available as the ground truth. ES is the same as AG, as AG exactly solved BAO. The dose plot window is [0%, 110%]. 100% isodose line, 80% isodose line and CTV are highlighted in dose plots.

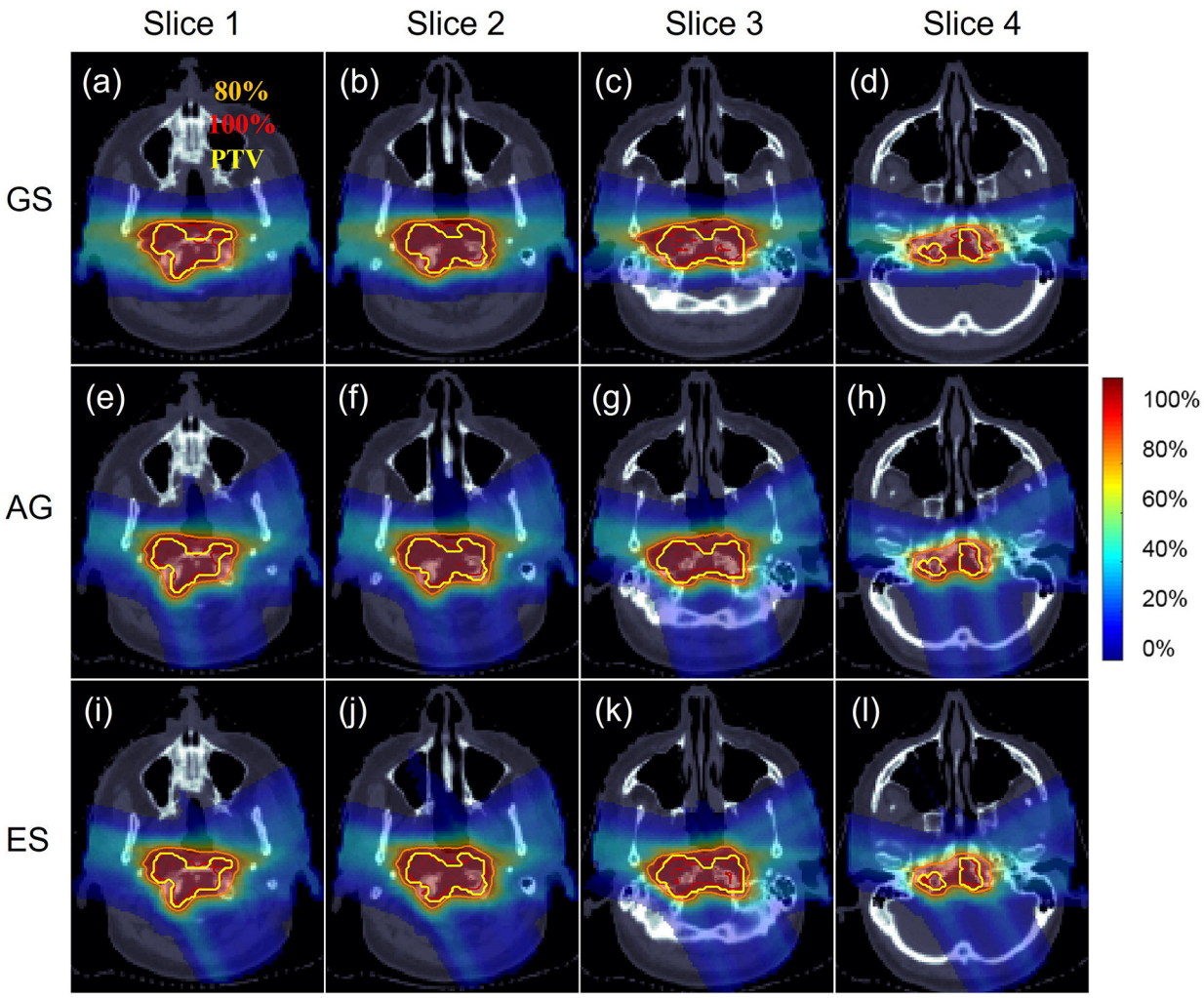


**Figure 5.**

Comparison of dose plots for 4-angle brain via ADMM with *B=24*, for which the optimal ES solution is available as the ground truth. The dose plot window is [0%, 110%]. 100% isodose line, 80% isodose line and CTV are highlighted in dose plots.

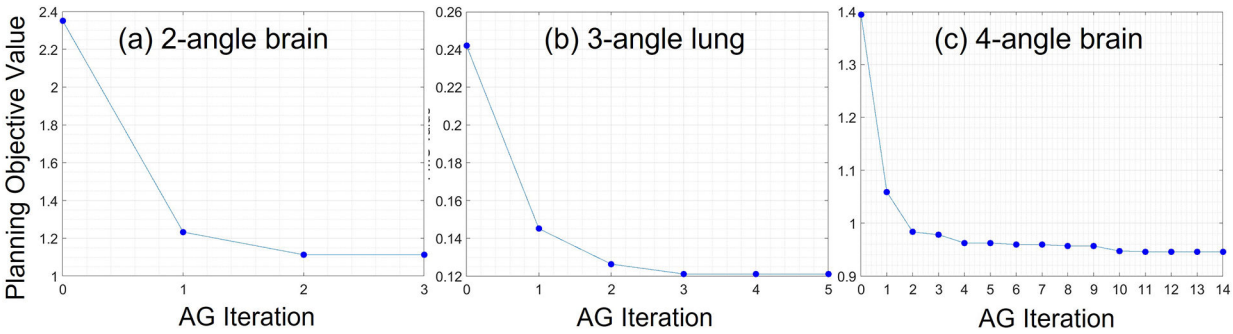


**Figure 6.**
The planning objective values are monotonically non-increasing during AG iterations.

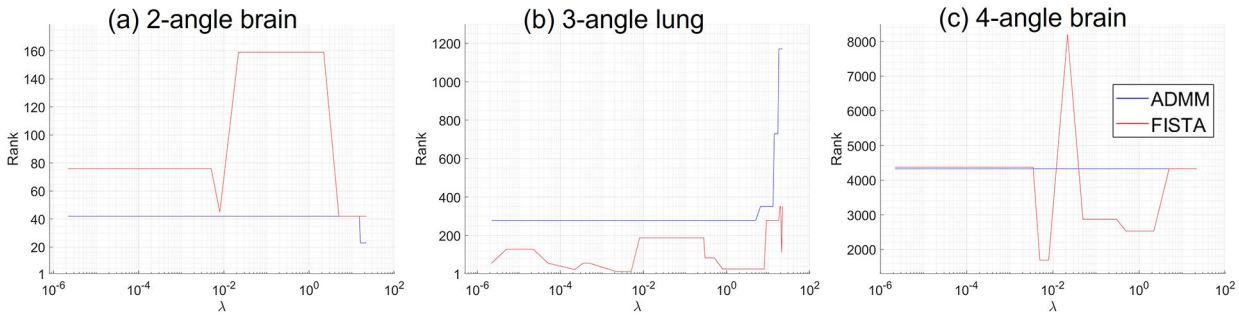


**Figure 7.**

Comparison of ADMM and FISTA in dependence of ranks of optimized plan quality with respect to the values of $\lambda$ in Eq. (5). While ADMM is insensitive to $\lambda$ except for large values of $\lambda$, FISTA is sensitive to $\lambda$, which can be problematic for choosing the optimal solution.

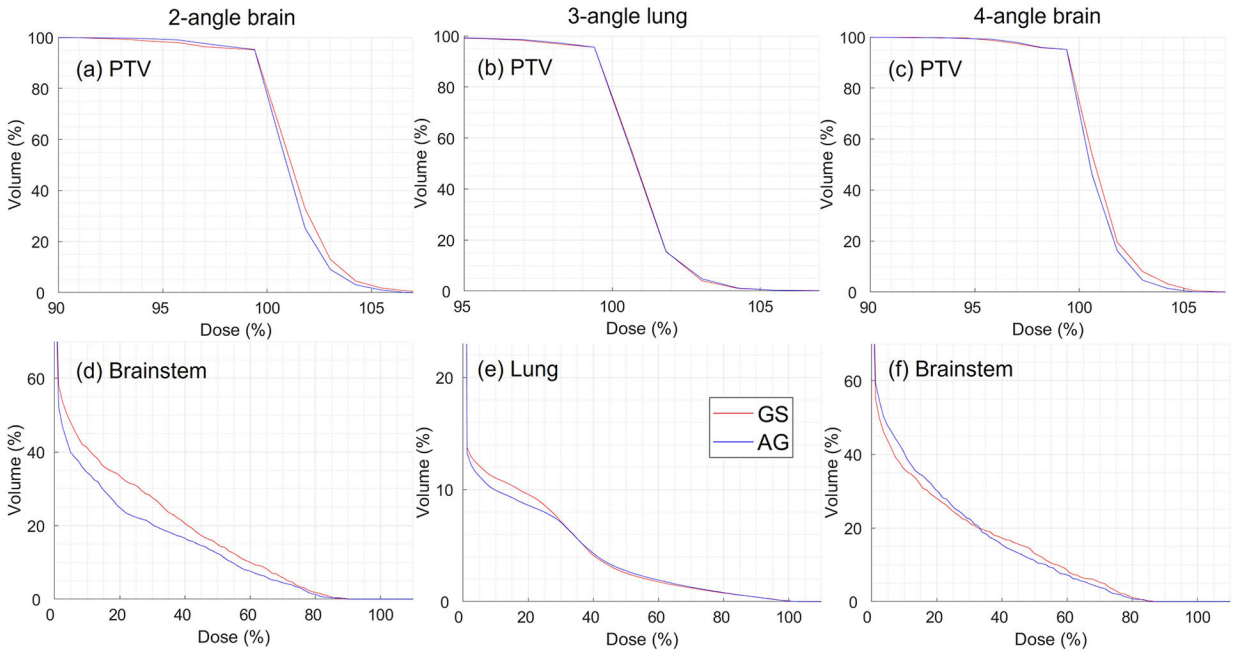


**Figure 8.**
Comparison of DVH plots between GS and AG via FISTA with *B=72*.

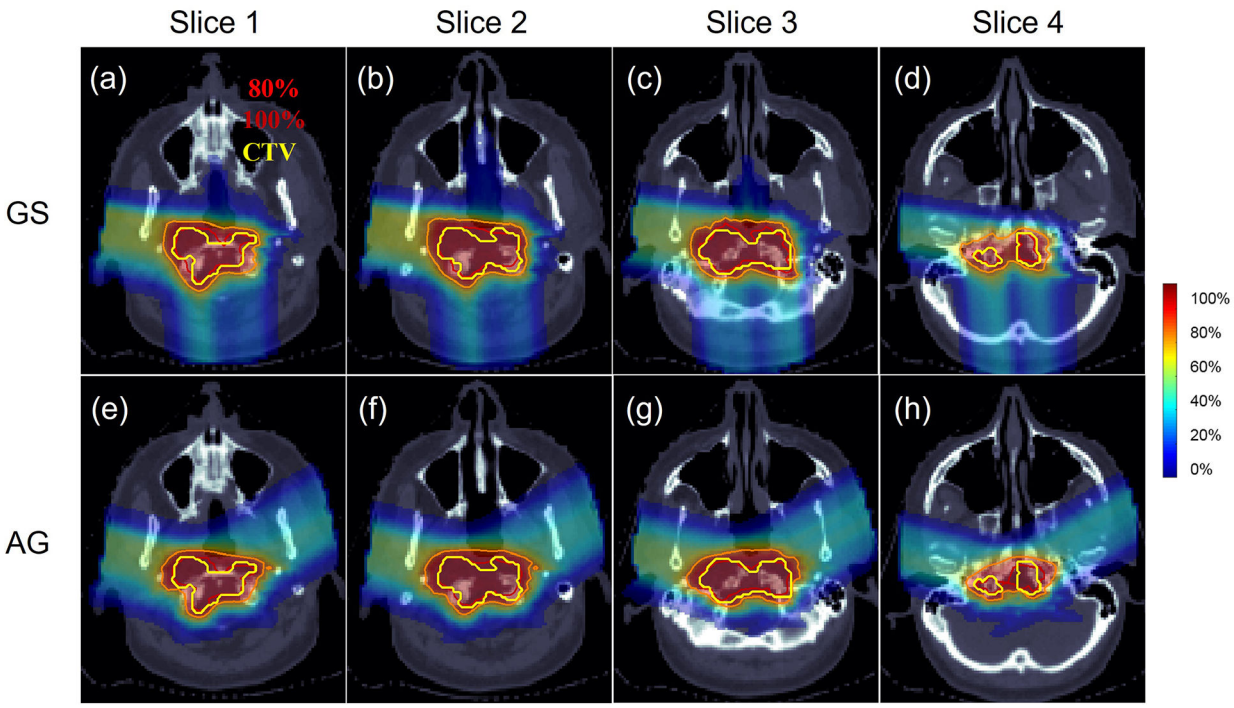


**Figure 9.**

Comparison of dose plots for 2-angle brain via FISTA with $B=72$, for which no ES solution is available as the ground truth. The dose plot window is [0%, 110%]. 100% isodose line, 80% isodose line and CTV are highlighted in dose plots.

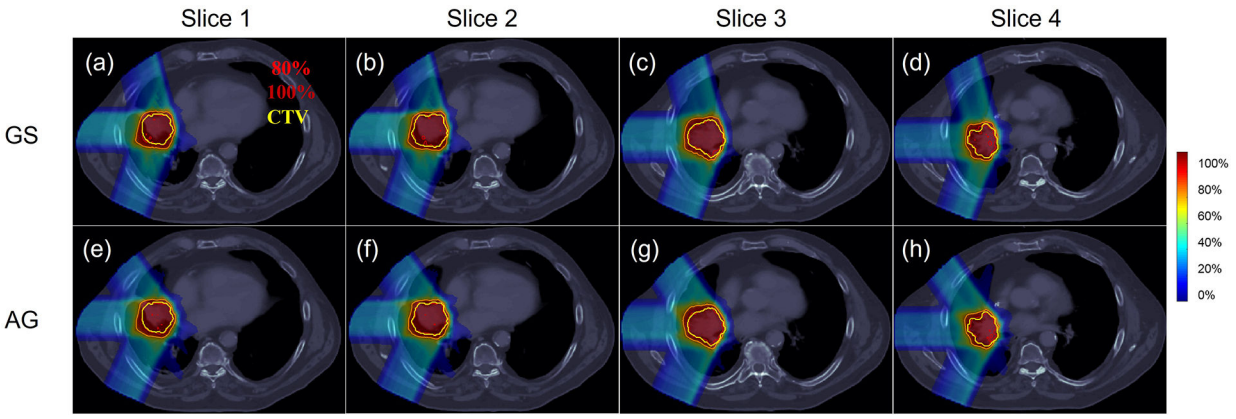


**Figure 10.**

Comparison of dose plots for 3-angle lung via FISTA with *B=72*, for which no ES solution is available as the ground truth. The dose plot window is [0%, 110%]. 100% isodose line, 80% isodose line and CTV are highlighted in dose plots.

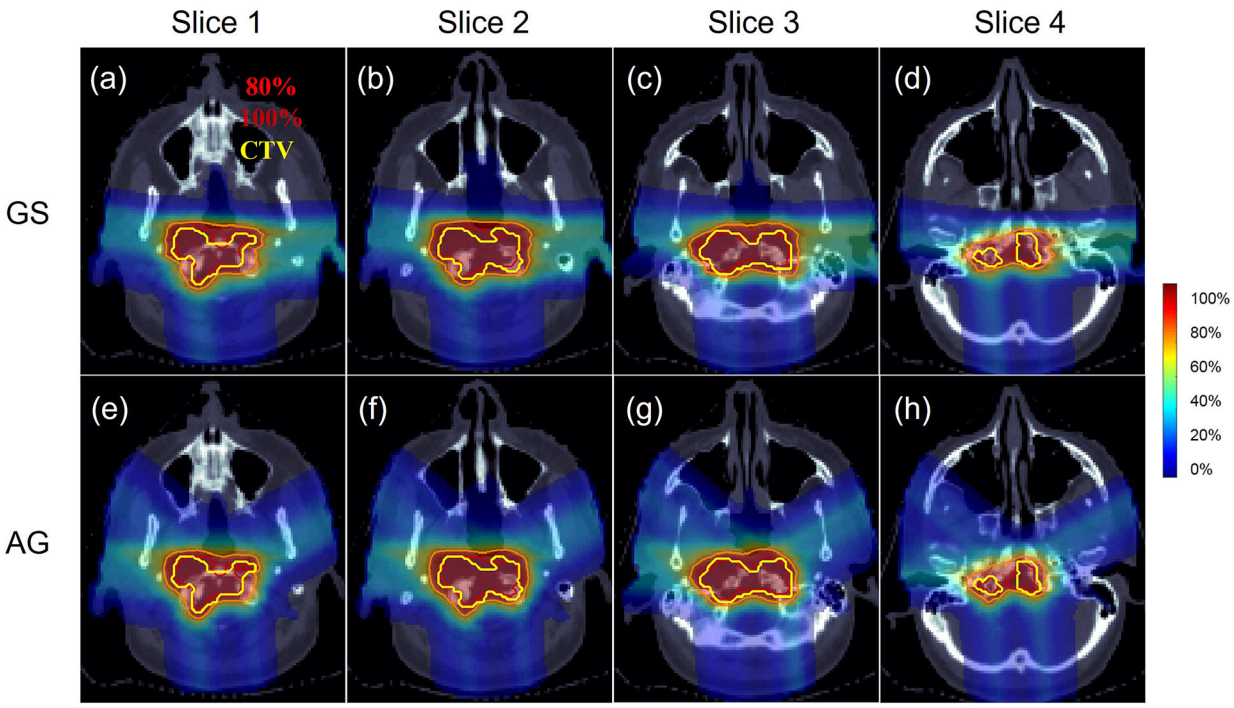


**Figure 11.**

Comparison of dose plots for 4-angle brain via FISTA with $B=72$, for which no ES solution is available as the ground truth. The dose plot window is [0%, 110%]. 100% isodose line, 80% isodose line and CTV are highlighted in dose plots.

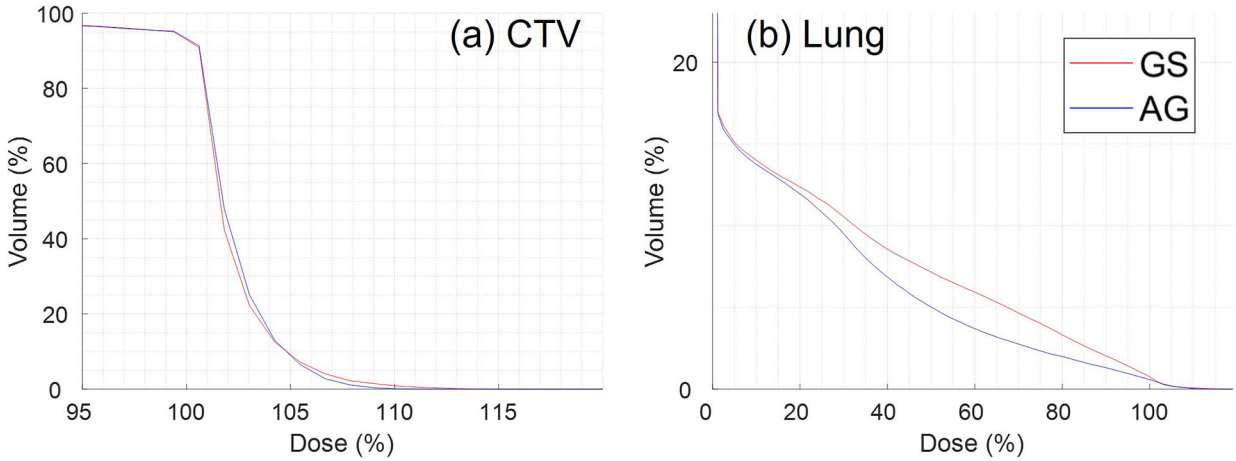


**Figure 12.**

Comparison of DVH plots for 3-angle lung between GS and AG via ADMM with *B=24* from robust optimization.

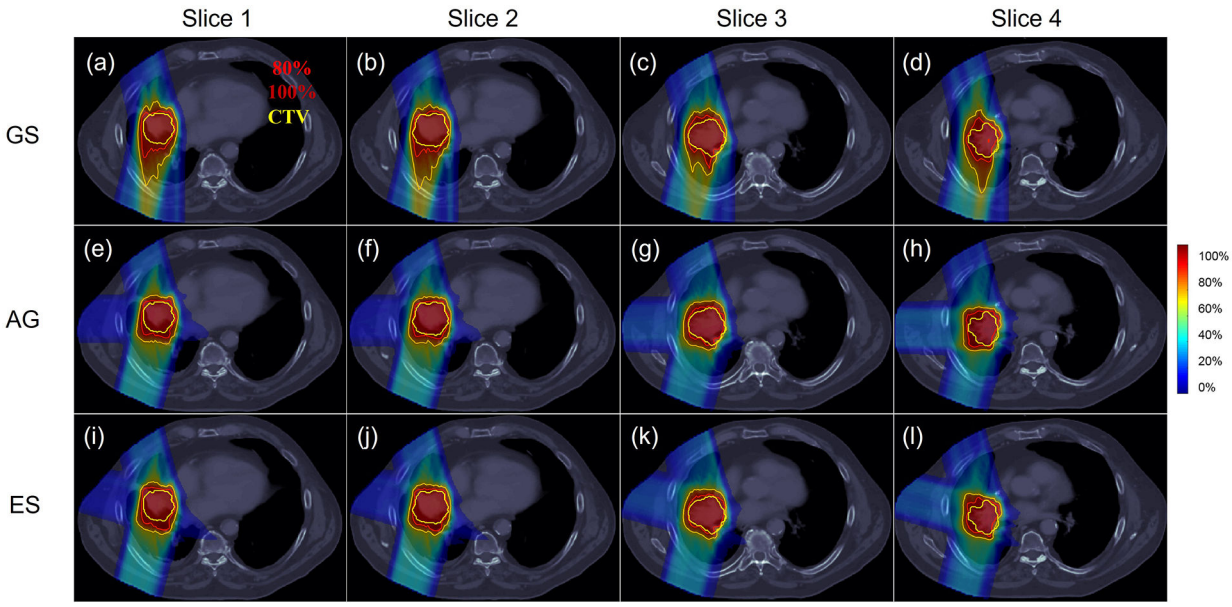


**Figure 13.**

Comparison of dose plots for 3-angle lung from robust optimization via ADMM with *B=24*, for which the optimal ES solution is available as the ground truth. The dose plot window is [0%, 110%]. 100% isodose line, 80% isodose line and CTV are highlighted in dose plots.

**Table 1.**

Comparison of GS and AG in benchmark to ES via ADMM with $B=24$.

| Case | Parameters | GS | AG | ES |
|---|---|---|---|---|
| 2-angle brain | CI | 0.60 | 0.62 | 0.62 |
| | $D_{max,\ brainstem}$ (Gy) | 17.82 | 17.75 | 17.75 |
| | $V_{10Gy,\ brainstem}$ (cc) | 3.11 | 2.75 | 2.75 |
| | $S^*$ | (7,19) | (5,20) | (5,20) |
| | $F$ | 1.23 | 1.11 | 1.11 |
| | Rank | 42 | 1 | 1 |
| 3-angle lung | CI | 0.90 | 0.92 | 0.92 |
| | $D_{mean,\ lung}$ (Gy) | 1.89 | 1.57 | 1.57 |
| | $S^*$ | (13,14,22) | (14,19,22) | (14,19,22) |
| | $F$ | 0.14 | 0.12 | 0.12 |
| | Rank | 279 | 1 | 1 |
| 4-angle brain | CI | 0.65 | 0.68 | 0.64 |
| | $D_{max,\ brainstem}$ (Gy) | 18.05 | 17.29 | 17.13 |
| | $V_{10Gy,\ brainstem}$ (cc) | 2.62 | 2.40 | 2.70 |
| | $S^*$ | (6,7,19,20) | (5,7,12,20) | (5,7,11,20) |
| | $f$ | 1.04 | 0.95 | 0.94 |
| | rank | 4328 | 4 | 1 |

**Table 2.**

Comparison of ADMM and FISTA for GS. The FISTA results here are from the best scenario by exhaustively searching the values of $\lambda$.

| Case | Parameters | ADMM | FISTA |
|---|---|---|---|
| 2-angle brain | CI | 0.60 | 0.60 |
| | $D_{max,\ brainstem}$ (Gy) | 17.82 | 17.91 |
| | $V_{10Gy,\ brainstem}$ (cc) | 3.11 | 2.73 |
| | $S^*$ | (7,19) | (19,21) |
| | $F$ | 1.23 | 1.24 |
| | rank | 42 | 51 |
| 3-angle lung | CI | 0.90 | 0.90 |
| | $D_{mean,\ lung}$ (Gy) | 0.52 | 0.56 |
| | $S^*$ | (13,14,22) | (14,19,24) |
| | $F$ | 0.14 | 0.13 |
| | rank | 279 | 12 |
| 4-angle brain | CI | 0.65 | 0.66 |
| | $D_{max,\ brainstem}$ (Gy) | 18.05 | 18.11 |
| | $V_{10Gy,\ brainstem}$ (cc) | 2.62 | 2.40 |
| | $S^*$ | (6,7,19,20) | (4,6,12,22) |
| | $F$ | 1.04 | 1.00 |
| | rank | 4328 | 1693 |

**Table 3.**

Comparison of ADMM and FISTA for AG using 3-angle lung.

| Parameters | ADMM | FISTA |
|---|---|---|
| CI | 0.92 | 0.92 |
| $S^*$ | (14,19,22) | (15,19,22) |
| $f$ | 0.12 | 0.12 |
| rank | 1 | 1 |

**Table 4.**

Comparison of two GS-IMPT formulations Eq. (4) and (6) for AG using 3-angle lung.

| Parameters | Eq. (4) | Eq. (6) |
|---|---|---|
| CI | 0.92 | 0.90 |
| $S^*$ | (14,19,22) | (14,16,19) |
| $f$ | 0.12 | 0.12 |
| rank | 1 | 36 |

**Table 5.**

Comparison of *p=1/2* and *p=1* in GS for AG using 3-angle lung.

| Iteration | *p=1/2* | | | *p=1* | | |
|---|---|---|---|---|---|---|
| | ***S*** | ***f*** | **rank** | ***S*** | ***f*** | **rank** |
| Initialization | (1,2,3) | 0.24 | 1963 | (1,2,3) | 0.24 | 1963 |
| 1 | (1,19,3) | 0.15 | 467 | (1,13,3) | 0.16 | 940 |
| 2 | (1,19,14) | 0.13 | 14 | (1,13,21) | 0.14 | 229 |
| 3 | (22,19,14) | 0.12 | 1 | (19,13,21) | 0.13 | 74 |
| 4 | (22,19,14) | 0.12 | 1 | (19,14,22) | 0.12 | 1 |
| 5 | (22,19,14) | 0.12 | 1 | (19,14,22) | 0.12 | 1 |
| 6 | - | - | - | (19,14,22) | 0.12 | 1 |

**Table 6.**

Comparison of GS and AG via FISTA with $B=72$.

| Case | Parameters | GS | AG |
|---|---|---|---|
| 2-angle brain | CI | 0.82 | 0.84 |
| | $D_{max,\ brainstem}$ (Gy) | 18.15 | 17.80 |
| | $V_{10Gy,\ brainstem}$ (cc) | 3.35 | 2.78 |
| | $S^*$ | (36,57) | (13,58) |
| | $F$ | 1.12 | 1.11 |
| 3-angle lung | CI | 0.92 | 0.92 |
| | $D_{mean,\ lung}$ (Gy) | 0.55 | 0.52 |
| | $S^*$ | (41,55,70) | (42,55,66) |
| | $F$ | 0.13 | 0.12 |
| 4-angle brain | CI | 0.86 | 0.87 |
| | $D_{max,\ brainstem}$ (Gy) | 17.25 | 16.88 |
| | $V_{10Gy,\ brainstem}$ (cc) | 2.92 | 2.48 |
| | $S^*$ | (18,19,36,57) | (13,36,56,63) |
| | $F$ | 1.01 | 0.94 |

**Table 7.**

Comparison of GS and AG in benchmark to ES from robust optimization via ADMM with $B=24$ using 3-angle lung.

| Parameters | GS | AG | ES |
|---|---|---|---|
| CI | 0.63 | 0.66 | 0.64 |
| $D_{mean,\ lung}$ (Gy) | 0.93 | 0.79 | 0.80 |
| $S^*$ | (13,14,24) | (14,19,24) | (14,20,24) |
| $F$ | 8.13 | 5.94 | 5.83 |
| Rank | 175 | 2 | 1 |