

The first direct method of spot sparsity optimization for proton arc therapy

Lewei Zhao^{a*} , Juntao You^{b*}, Gang Liu^{a,c}, Sophie Wuyckens^d, Xiliang Lu^{e†} and Xuanfeng Ding^{a†}

^aProton Therapy Center, Department of Radiation Oncology, Corewell Health William Beaumont University Hospital, Royal Oak, MI, USA;

^bDepartment of Mathematics, The Hong Kong University of Science and Technology, Hong Kong SAR, PR China; ^cCancer Center, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, PR China; ^dMolecular Imaging, Radiotherapy and Oncology (MIRO), UCLouvain, Belgium; ^eSchool of Mathematics and Statistics, and Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan, PR China

ARTICLE HISTORY Received 8 November 2022; Accepted 21 January 2023

Introduction

Spot-scanning proton arc (SPArc) therapy was introduced in 2016 offering superior target conformity and simplifying the proton therapy treatment workflow [1,2]. It is considered an advanced intensity modulated proton therapy (IMPT) technique through arc trajectory optimization. This new approach allows the proton system to deliver the proton beam following a sequence of control points while the gantry continuously rotates around the patient.

One of the challenges in the clinical implementation of proton arc therapy is to minimize its beam delivery time (BDT), as the plan might contain numerous energy layers and spots. In the beginning stage of the pencil beam scanning (PBS) clinical implementation, energy layer switching time (ELST) is the bottleneck due to the technical limitation that prolongs the treatment delivery time. Thus, several studies focused on reducing energy layer switching time through energy sequence optimization [3–7]. However, a recent study [8] found that the BDT is approximately proportional to the spot number for IBA's ProteusONE PBS proton therapy system in Beaumont where about half the treatment delivery time is spent on the spot switching [9]. It becomes a new bottleneck of SPARC therapy, which normally contains thousands of spots making it very challenging to deliver in a timely manner. Thus, it is critical to reducing the spot number while maintaining the optimal treatment plan quality.

Searching for the minimum spot number solution can be considered as a sparsity optimization problem [10,11], where the sparsity is the number of zero-valued elements divided by the total number of elements. The most natural measure of sparsity is the counting function $\|\cdot\|_0$, called usually the l_0 -norm [12]. The l_0 -norm counts the total number of non-zero elements of a vector (spots). Minimizing the number of nonzeros of the solution (its l_0 -norm) is a difficult nonconvex and non-deterministic polynomial-time hardness (NP hard) optimization problem [13,14]. One widely used

replacement method is using l_1 -norm [13,15] as convex approximation to the l_0 -norm (See [supplemental material](#) for their comparison). Meanwhile, a novel primal dual active set (PDAS) algorithm was developed by Bergounioux et al. [16] and Ito et al. [11] to solve the nonsmooth optimization [17] (not necessarily differentiable optimization) problem. For l_0 -regularized minimization problem, Jiao et al. [18] and Huang et al. [19] propose a PDASC algorithm, which couples the PDAS with a continuation strategy on the regularization parameter in outer iteration. Each inner iteration first identifies the active set from both primal and dual variables, then updates the primal variable by solving a (typically small) least-squares problem defined on the active set, from which the dual variable can be updated explicitly. It is observed that the PDASC algorithm yields optimal solutions that are comparable with other methods, e.g., OMP, CoSaMP, and AIHT [20], but usually with less computing time.

In this paper, we introduced the first optimization framework for SPARC by searching the optimal spot sparsity solution using PDASC or we called it SPARC-PDASC. The proposed study aims to solve the two main challenges (a) simultaneous optimization of treatment delivery efficiency and dosimetric plan quality; (b) optimization speed, the proposed approach can be hundreds of times more efficient. The successful integration of the PDASC algorithm into SPARC could serve as one option to provide more delivery-efficient plans.

Material and Methods

Problem formulation

Based on the previously studied beam delivery sequence model of the new generation of the single room system, Beaumont's IBA ProteusONE, proton treatment delivery time is actually dominated by spot switching time (SSWT) [9]. SSWT is approximately linearly dependent on spot numbers [8]. Let $\mathbf{w} = \{w_i\}_{i=1}^n$ be the vector of spot intensity (MU),

CONTACT Xuanfeng Ding  xuanfeng.ding@beaumont.org  Proton Therapy Center, Department of Radiation Oncology, Corewell Health William Beaumont University Hospital, 3571 W 13 Mile Rd, Royal Oak, MI 48073, USA

*Lewei Zhao and Juntao You contribute equally to the study.

†Xiliang Lu and Xuanfeng Ding share the senior authorship.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/0284186X.2023.2172689>.

where n is the full (initial) spot number. Spot index set $\mathbf{S} = \{1, 2, \dots, n\}$. $w_i = 0$ means this spot has been removed. The true spots $w_i > 0$ of nonzero MU form a primal dual active set denoted by A .

Let $\Phi \in \mathbb{R}^{m \times n}$ be kernel fluence matrix (also called the dose deposition matrix [21]) simulated using MatRad dose calculation engine [22], where m is the total number of voxels. Denote $\mathbf{p}_0 \in \mathbb{R}^m$ as prescribed dose vector. It is a common approach to minimize the least-squares fitting, i.e., $\frac{1}{2} \|\Phi \mathbf{w} - \mathbf{p}_0\|_2^2$ [13,23]. Further, by utilizing the sparsity in $\mathbf{w} \in \mathbb{R}^n$, it is possible to introduce efficient solvers for the problem. In the work, to enforce a sparse solution to the objective function, we consider solving the following l_0 -regularized optimization problem:

$$\underset{\mathbf{w} \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2} \|\Phi \mathbf{w} - \mathbf{p}_0\|_2^2 + \lambda \|\mathbf{w}\|_0 \quad (1a)$$

$$\text{subject to } \mathbf{w} \geq 0, \quad (1b)$$

where $\lambda > 0$ is a regularization parameter, controlling the sparsity level of the regularized solution, and (1b) follows from the fact that \mathbf{w} is non-negative.

The necessary and sufficient condition for a coordinate-wise minimizer \mathbf{w} to a problem (1a) is given by [18]

$$\begin{cases} |w_i + d_i| < \sqrt{2\lambda} \Rightarrow w_i = 0 \\ |w_i + d_i| = \sqrt{2\lambda} \Rightarrow w_i = 0 \text{ or } d_i = 0 \\ |w_i + d_i| > \sqrt{2\lambda} \Rightarrow d_i = 0 \end{cases} \quad (2)$$

where $\mathbf{d} = \{d_i\}_{i=1}^n = \Phi^T(\mathbf{p}_0 - \Phi \mathbf{w})$ denotes the dual variable.

For (1b), we can simply project the estimation of (1a) onto the space \mathbb{R}_+^n .

Algorithm procedure

Our algorithm includes an outer iteration and inner loop. The initial spot intensity is chosen as zero vector $\mathbf{w}^0 = \mathbf{0}$. Regularization parameter is chosen among intervals $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$. In k th outer iteration ($k = 1, 2, \dots, K_{\max}$), λ is given by

$$\lambda^k = \frac{1}{2} \|\Phi^T \mathbf{p}_0\|_{l_\infty} \exp \left[\ln \lambda_{\max} - \frac{k-1}{K_{\max}-1} \ln \frac{\lambda_{\max}}{\lambda_{\min}} \right] \quad (3)$$

At j th inner loop ($j = 1, 2, \dots, J_{\max}$), it identifies the active set from primal variable \mathbf{w}^{j-1} and dual variable \mathbf{d}^{j-1} by condition (2):

$$A^j = \{i : |w_i^{j-1} + \eta d_i^{j-1}| > \sqrt{2\lambda_k}\}, \quad (4)$$

where $\eta > 0$ is a step size parameter to be adjusted when condition number of Φ is large, \mathbf{w}^j and \mathbf{d}^j are updated by solving a (typically small) least-squares problem defined on the dual active set:

$$\begin{cases} \mathbf{w}_{A^j}^j = \mathbf{0} \end{cases} \quad (5a)$$

$$\begin{cases} \Phi_{A^j}^T \Phi_{A^j} \mathbf{w}_{A^j}^j = \Phi_{A^j}^T \mathbf{p}_0 \end{cases} \quad (5b)$$

$$\begin{cases} \mathbf{d}^j = \Phi^T(\mathbf{p}_0 - \Phi \mathbf{w}^j) \end{cases} \quad (5b)$$

The complete procedure is summarized below:

Algorithm 1 SPARC-PDASC algorithm

- 1: Set parameter $\eta, \lambda_{\min}, \lambda_{\max}, K_{\max}$ and J_{\max} .
- 2: **for** $k = 1, 2, \dots, K_{\max}$ **do**
- 3: Set λ^k by formula (3)
- 4: **for** $j = 1, 2, \dots, J_{\max}$ **do**
- 5: Compute the A^j by (4)
- 6: Check the inner stopping criterion $A^{j-1} = A^j$
- 7: Update the \mathbf{w}^j and \mathbf{d}^j by (5)
- 8: **end for**
- 9: Check sparsity for outer stopping criterion
- 10: **end for**
- 11: **return** $\mathbf{w} = \max\{\mathbf{w}, \mathbf{0}\}$, the best solution found

We can adjust the outer stopping criterion in the loop to obtain the different sparsity level solutions.

Evaluation

Two clinical cases: One intracranial cancer case and one lung cancer case from a previous study [8] were used for the evaluation. For each patient, only the primary clinical target volume (CTV) was considered in the optimization. The prescription dose for the intracranial tumor is 54 Gy in 30 fractions (1.8 Gy per fraction) and for the lung tumor is 48 Gy in 4 fractions (12 Gy per fraction) using stereotactic body radiotherapy (SBRT) [24]. Their CT resolution is in 2.5 mm. Single arc was used in the planning. Other specific information is listed in [supplemental material table 1](#).

The cumulative dose volume histogram (DVH) was plotted for evaluating the plan quality [25]. In addition, the optimization time is compared with the original SPARC algorithm (SPARC-original, [1]) which was implemented in a commercial treatment planning system (TPS) RayStation (RaySearch Laboratories AB, Stockholm, Sweden, ver. 6) through in-house scripts. The optimizer of RayStation RayOptimizer [26] utilizes the sequential quadratic programming (SQP) software package NPSOL [27] in which the SQP uses Broyden–Fletcher–Goldfarb–Shanno (BFGS) updates [21]. The sparsity optimization is implemented by gradually increasing the parameter of the minimal spot weight. It is one of the limitations when filtering out low-weight spots in RayStation which prohibits the user to reactive the filtered spots [28]. SPARC-original is an iterative approach, which will take much longer time to optimize. The objective function in RayStation is set as a uniform dose function, which penalizes the deviation from a certain dose level [29], to compare with l_2 -norm objective term in PDASC.

Results

We obtained different sparsity-level solutions by manually adjusting the outer stopping criterion. [Figure 1](#) showed the comparison between 96.6% high spot sparsity and 11.2% low sparsity for lung SBRT cases. Their optimized spot intensities were compared in [Figure 1\(A\)](#), which indicates that less spot number plan has greater intensity. The DVH is

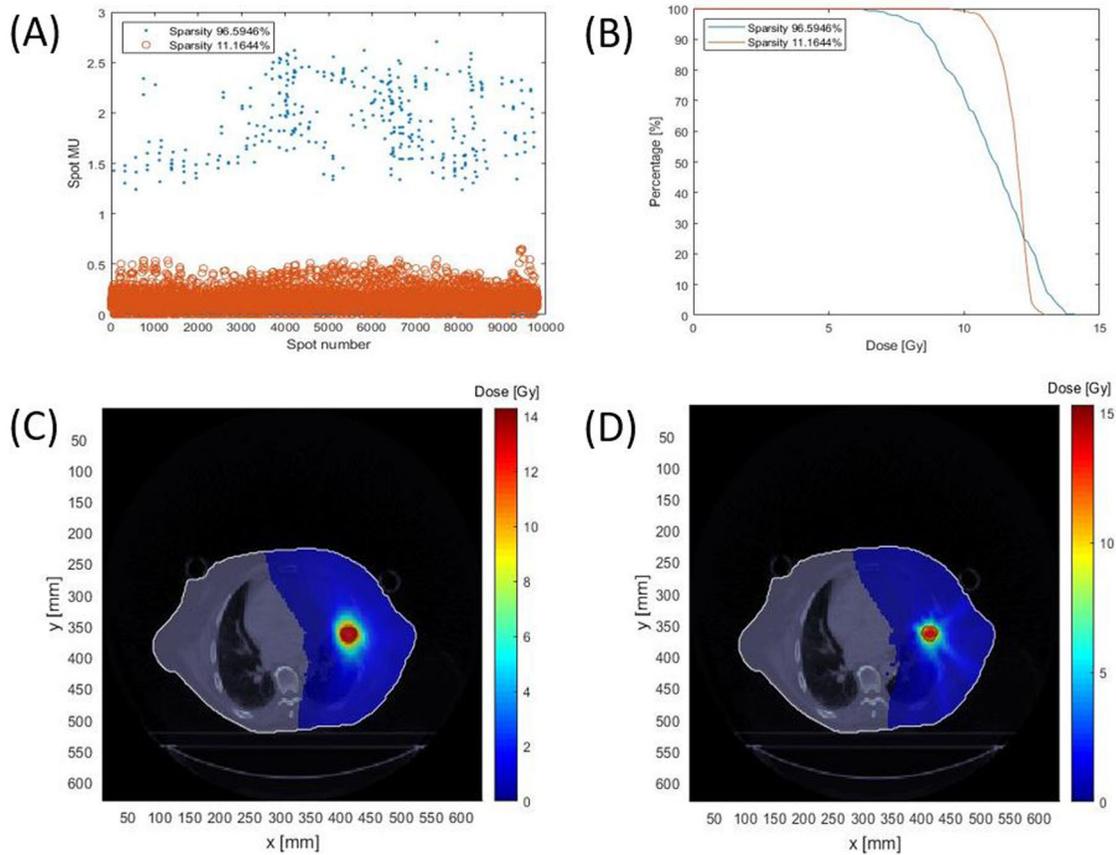


Figure 1. An example of comparison between high sparsity and low sparsity in a lung SBRT case. The prescription dose is 48 Gy and 12 Gy per fraction. (A) High sparsity resulted in a higher averaged spot MU. (B) Dose-volume histogram (DVH) per fraction shows low sparsity has better plan quality. (C) Dose distribution for 11.2% sparsity. (D) Dose distribution for 96.6% sparsity.

compared in Figure 1(B). High spot sparsity reduced the plan quality because of losing too much degree of freedom in terms of spot number. Their plan dose was demonstrated in Figure 1(C,D).

Supplemental material Figure s1 showed the comparison between 93.4% high spot sparsity plan and 11.7% low sparsity for the intracranial case. Figure s1A compared their spot MU. Since high sparsity plan has much fewer spot number, their spot MU (blue dot, around 0.2 MU) are greater than the low sparsity plan (red circle, around 0–0.05 MU). Figure s1B compared their DVH. High spot sparsity (blue line) loses the plan quality due to the lack of optimization freedom from the limited spot number. Figures s1C and s1D show their plan dose. Spot MU in Figure 1(A) (about 1.5–2.5 MU for high sparsity plan and 0–0.5 MU for low sparsity plan) is greater than in Figure s1A because SBRT has a lower number of fractions.

Figure 2 plotted the plan quality and optimization time with respect to different sparsity level for both intracranial and lung SBRT case, which are compared with SPARC-original implemented in RayStation. Figure 2(A,C) showed the plan quality degraded as the number of spots is reduced. The plan quality is measured by relative objective value, which is normalized by the lowest sparsity plan's clinic objective value at each case for different methods separately. Compared to SPARC-original, SPARC-PDASC was able to find a better plan quality with the same spot sparsity, which could be critical

to the SPARC therapy. In the optimization speed evaluation (Figure 2(B,D)), the study found that the PDASC could significantly reduce the optimization time: this new planning framework could effectively improve the optimization speed by a factor of about three hundred on average (8.8 times to 536.5 times from 20%–80% sparsity) compared to the SPARC-original implemented in RayStation. Opposite to RayStation's sparsity optimization through filtering spot from low sparsity to high sparsity, SPARC-PDASC searches the global sparsity solution from its zero initial solution. The higher the sparsity, the less optimization time PDASC spends. This can be theoretically explained by Jiao et al. [18]: there is an important monotonicity relationship on the active set during the PDASC iteration. The greater the active set size is, the greater the computation complexity will be (See supplemental material for complexity analysis).

Discussion

Shortening the beam delivery time (BDT) is important to radiation therapy which can be used to mitigate the intra-fractional motion [30] and improve the treatment throughput. PDASC combines the fast local convergence of the active set technique and the globalizing property of the continuation technique for sparsity optimization, which demonstrates its value in proton arc optimization and plan generation. The study demonstrated that the BDT could be shortened

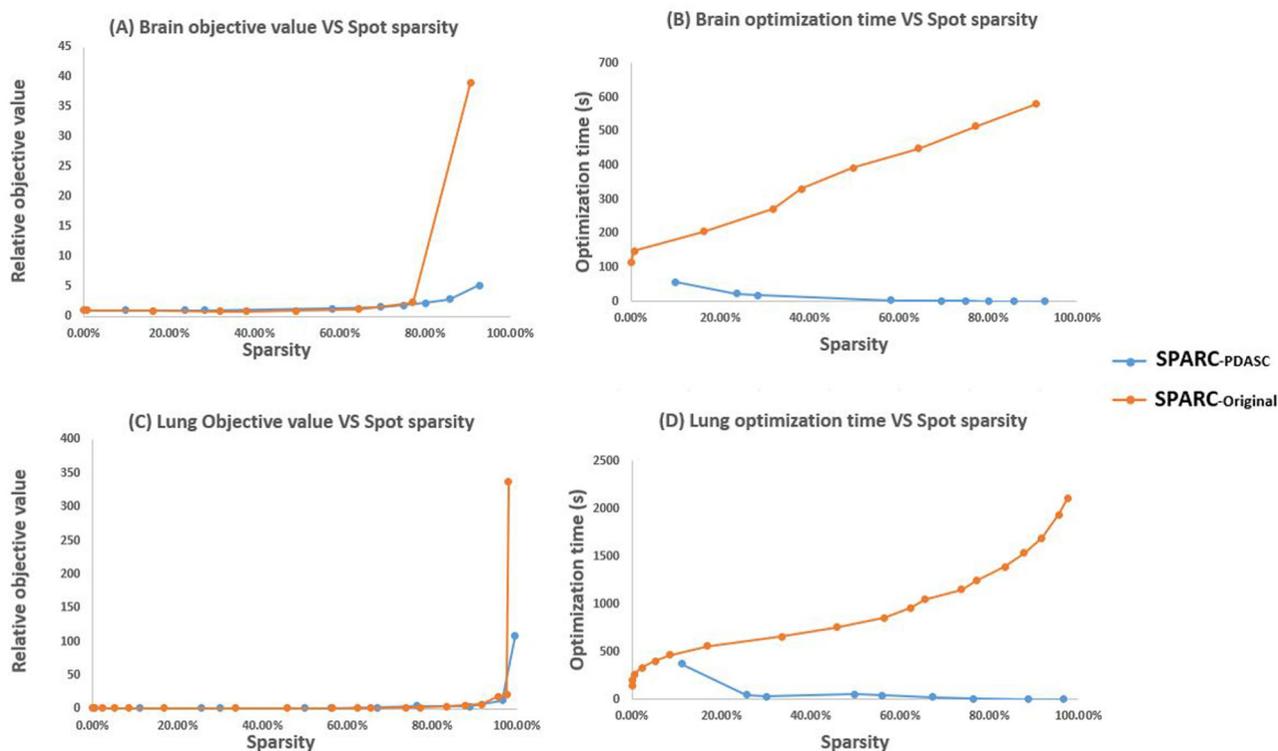


Figure 2. Compared with SPARC-original implemented in RayStation (Stockholm, Sweden). (A) Brain case objective value vs. sparsity curve using SPARC-PDASC. (B) Brain optimization time vs. sparsity curve. (C) Lung SBRT case objective value vs. sparsity curve using SPARC-PDASC. (D) Lung SBRT optimization time vs. sparsity curve.

through PDASC with a faster optimization speed compared to the SPARC-original. However, too many constraints on the BDT might affect the plan quality as the treatment plan starts losing degrees of freedom in terms of the number of spots. An optimal proton arc therapy plan is to find a trade-off between the plan quality and BDT. In our institution's proton delivery system, the BDT is reduced by decreasing the spot number through sparsity optimization. In reality, a tradeoff might consist of a series of solutions, which could be given by Pareto surface through multi-criterion optimization (MCO) [31]. PDASC could provide a proper optimizer for MCO search, i.e., the individual point on a Pareto surface.

In the clinic, only spot with MU being greater than a machine threshold can be delivered. PDASC has separation from zero property [18]:

$$\min_{i \in A} w_i \geq \sqrt{2\lambda} \quad (6)$$

But λ could be finally chosen as a very small value in PDASC iteration, which might not satisfy the machine threshold. How to choose the proper stopping criterion to meet the machine threshold might be a potential improvement to this algorithm.

One limitation of this work is that robustness optimization is not considered in this method. The difficulty is how the mathematically-defined robustness in this PDASC optimization formulation remains unclear, which will be the focus of the next development (See [supplemental material](#) for more discussion). Another limitation of this paper is only investigating spot numbers without considering the spot scanning paths. Different scanning paths lead to different

dose distributions and BDT due to the contribution of the unintended transit dose and spot swi time between spots [32]. A simulated annealing algorithm has been applied to scanning path optimization of particle therapy beams [33]. It is a heuristic algorithm to optimize the scanning path for quasi-discrete scanned beams. PDASC used in this paper might provide a methodology for a direct method for proton arc spot scanning path optimization.

Conclusion

The study developed the first fast-planning framework for proton arc therapy spot-sparsity optimization. Such advancement in spot-sparsity optimization is critical to the SPARC therapy for an efficient treatment delivery with a balanced plan quality. This work also paves the roadmap for clinical implementation in the TPS platform efficiently.

Disclosure statement

The corresponding author X.D report honorarium from the Ion Beam Applications (IBA) and ELEKTA speaker Bureau.

Funding

G.L is supported by the National Natural Science Foundation of China (No.12005072). L.Z is supported by Beaumont Seed Grant Award. This work was supported by William Beaumont Hospital.

ORCID

Lewei Zhao  <http://orcid.org/0000-0002-1050-3184>

Data availability statement

The data that support the findings of this study are available from the corresponding author, X.D, upon reasonable request.

References

- [1] Ding X, Li X, Zhang JM, et al. Spot-scanning proton arc (sparc) therapy: the first robust and delivery-efficient spot-scanning proton arc therapy. *Int J Radiat Oncol Biol Phys.* 2016;96(5):1107–1116.
- [2] Liu G, Li X, Qin A, et al. Is proton beam therapy ready for single fraction spine sbirs?—a feasibility study to use spot-scanning proton arc (sparc) therapy to improve the robustness and dosimetric plan quality. *Acta Oncol.* 2021;60(5):653–657.
- [3] Liu G, Li X, Zhao L, et al. A novel energy sequence optimization algorithm for efficient spot-scanning proton arc (sparc) treatment delivery. *Acta Oncol.* 2020;59(10):1178–1185.
- [4] Gu W, Ruan D, Lyu Q, et al. A novel energy layer optimization framework for spot-scanning proton arc therapy. *Med Phys.* 2020; 47(5):2072–2084.
- [5] Engwall E, Battinelli C, Wase V, et al. Fast robust optimization of proton pbs arc therapy plans using early energy layer selection and spot assignment. *Phys. Med. Biol.* 2022;67(6):065010.
- [6] Wuyckens S, Saint-Guillain M, Janssens G, et al. Treatment planning in arc proton therapy: comparison of several optimization problem statements and their corresponding solvers. *Comput Biol Med.* 2022;148:105609.
- [7] Zhang G, Shen H, Lin Y, et al. Energy layer optimization via energy matrix regularization for proton spot-scanning arc therapy. *Med Phys.* 2022;49(9):5752–5762.
- [8] Zhao L, Liu G, Li X, et al. An evolutionary optimization algorithm for proton arc therapy. *Phys. Med. Biol.* 2022;67(16):16NT01.
- [9] Zhao L, Liu G, Chen S, et al. Developing an accurate model of spot-scanning treatment delivery time and sequence for a compact superconducting synchrocyclotron proton therapy system. *Radiat Oncol.* 2022;17(1):1–19.
- [10] Peleg D, Meir R. A bilinear formulation for vector sparsity optimization. *Signal Process.* 2008;88(2):375–389.
- [11] Ito K, Kunisch K. A variational approach to sparsity optimization based on lagrange multiplier theory. *Inverse Prob.* 2014;30(1): 015001.
- [12] Nikolova M. Relationship between the optimal solutions of least squares regularized with 0-norm and constrained by k-sparsity. *Appl Comput Harmon Anal.* 2016;41(1):237–265.
- [13] Donoho DL, Tsaig Y, Drori I, et al. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Trans Inf Theory.* 2012;58(2):1094–1121.
- [14] Feng M, Mitchell JE, Pang JS, et al. Complementarity formulations of l0-norm optimization problems. *Industrial Engineering and management sciences technical report northwestern university, Evanston, IL, USA.* 2013. 5.
- [15] Kumar S, Mamun K, Sharma A. Csp-tnm: optimizing the performance of riemannian tangent space mapping using common spatial pattern for mi-bci. *Comput Biol Med.* 2017;91:231–242.
- [16] Bergounioux M, Ito K, Kunisch K. Primal-dual strategy for constrained optimal control problems. *SIAM J Control Optim.* 1999; 37(4):1176–1194.
- [17] Bagirov A, Karmitsa N, Mäkelä MM. *Introduction to nonsmooth optimization: theory, practice and software.* Vol. 12. Switzerland: Springer International Publishing; 2014.
- [18] Jiao Y, Jin B, Lu X. A primal dual active set with continuation algorithm for the 0-regularized optimization problem. *Appl Comput Harmon Anal.* 2015;39(3):400–426.
- [19] Huang J, Jiao Y, Jin B, et al. A unified primal dual active set algorithm for nonconvex sparse recovery. *Statistical Science.* 2021;36(2):215–238.
- [20] Blumensath T. Accelerated iterative hard thresholding. *Signal Process.* 2012;92(3):752–756.
- [21] Sundström J. Scenario dose calculation for robust optimization in proton therapy treatment planning [master's thesis]. Sweden: Lund University; 2021.
- [22] Wieser HP, Cisternas E, Wahl N, et al. Development of the open-source dose calculation and optimization toolkit matrad. *Med Phys.* 2017;44(6):2556–2568.
- [23] Needell D, Vershynin R. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of Computational Mathematics.* 2009;9(3):317–334.
- [24] Onishi H, Shirato H, Nagata Y, et al. Stereotactic body radiotherapy (sbirt) for operable stage i non-small-cell lung cancer: can sbirt be comparable to surgery? *Int J Radiat Oncol Biol Phys.* 2011;81(5):1352–1358.
- [25] Khan FM, Gibbons JP. *Khan's the physics of radiation therapy.* Philadelphia, PA: Lippincott Williams & Wilkins; 2014.
- [26] RaySearch. *RayStation product configurations overview brochure.* Sweden; 2017.
- [27] Gill PE, Murray W, Saunders MA, et al. *User's guide for npsol (version 4.0): a fortran package for nonlinear programming.* Stanford Univ CA Systems Optimization Lab; 1986.
- [28] Laboratories R. *RayStation 6 user manual.* Sweden; 2016. RSL-D-RS-6.0-USM-EN-1.0-2016-12-22.
- [29] Laboratories R. *A guide to optimization in RayStation.* Sweden; 2016. RSL-D-RS-6.0-OPT-EN-1.0-2016-12-22.
- [30] Kim S, Li J, Liu C, et al. Evaluation of intra-fraction patient movement for the head & neck imrt. *Int J Radiat Oncol, Biol, Phys.* 2002;54(2):319.
- [31] Craft DL, Halabi TF, Shih HA, et al. Approximating convex pareto surfaces in multiobjective radiotherapy planning. *Med Phys.* 2006; 33(9):3399–3407.
- [32] Kang JH, Wilkens JJ, Oelfke U. Demonstration of scan path optimization in proton therapy. *Med Phys.* 2007;34(9):3457–3464.
- [33] Pardo J, Donetti M, Bourhaleb F, et al. Heuristic optimization of the scanning path of particle therapy beams. *Med Phys.* 2009; 36(6):2043–2051.