Radiotherapy Dosimetry: A Review on Open-Source Optimizer

Paul Dubois

May 26, 2023

Abstract

Radiotherapy dosimetry plays a crucial role in optimizing treatment plans for cancer patients. In this study, we investigate the performance of a dozen standard state-of-the-art open-source optimizers for radiotherapy dosimetry. Our evaluation includes the use of TGG119 benchmark cases as well as one real case obtained from the Institute du Cancer de Montpellier (ICM). Among the tested optimizers, Newton CG demonstrates the fastest convergence in terms of the number of iterations. However, when considering the computation time per iteration, LBFGS emerges as the most efficient optimizer. These findings shed light on the performance of open-source optimizers for radiotherapy dosimetry, aiding practitioners in selecting suitable optimization tools for efficient treatment planning.

1 Introduction

Radiotherapy, a widely utilized intervention for cancer treatment, employs ionizing radiation to eliminate malignant cells. Intensity-modulated radiation therapy (IMRT) has emerged as a notable technique within radiotherapy, aiming to deliver high radiation doses to tumors while minimizing exposure to healthy surrounding tissues [4]. Traditional IMRT strategies typically employ a set number of beams, often 5, 7, or 9, originating from various angles around the patient, commonly distributed evenly [3]. Each beam's intensity is modulated to optimize the delivery of radiation doses to the tumor while reducing exposure to healthy tissues. This approach surpasses the effectiveness of the 3D-conformal radiotherapy (3D-CRT) technique [10] [16] [20]. To facilitate precise and efficient radiation delivery, a computer-controlled device called the multi-leaf collimator (MLC) is utilized to shape the radiation beam according to the contours of the tumor.

The effectiveness of a radiotherapy treatment plan relies on the optimization procedure, which involves a series of steps aimed at ensuring the optimal delivery of radiation in accordance with the prescribed guidelines of medical practitioners. Typically, computer software is employed to facilitate the optimization process, taking into account various factors such as patient anatomy, the size and location of the tumor and organs, and the radiation objectives defined by medical professionals.

While there has been comparison between commercial software[18] [13], the aim of this paper is to investigate the best optimizer for this task among the open source optimizers.

Pre-dose-optimization The initial stage of the optimization process entails the creation of a virtual representation of the patient's anatomical structure using advanced medical imaging modalities, such as computed tomography (CT) or magnetic resonance imaging (MRI) scans. This model is subsequently utilized to accurately determine the size and location of the tumor, as well as to delineate the surrounding healthy tissues that necessitate protection from radiation exposure. Following this, the radiation dose required for effective treatment is established, typically based on dose-volume objectives defined by physicians (e.g., ensuring that 95% of the planning target volume receives a minimum dose of 75 Gy). Determining the appropriate dose takes into consideration factors such as tumor characteristics, location, size, as well as the patient's medical history and overall health status. These essential steps in the optimization process are carried out by medical professionals with expertise in radiotherapy treatment planning.

Radiotherapy doses The subsequent step entails the computation of the radiation dose distribution within the patient's volumetric anatomy. This is achieved by simulating a particular configuration of the multi-leaf collimator (MLC) on the patient's body, utilizing the available medical imaging data. The resulting computed dose represents a mapping from the three-dimensional volume of the patient's anatomy to a scalar value measured in Grays (Gy), which denotes the absorbed radiation energy. In practical implementation, a discrete representation of the dose distribution is utilized, wherein the dose is calculated for each individual voxel comprising the patient's anatomical structure.

Dose-Volume Histograms Medical professionals have meticulously identified and delineated the pertinent anatomical structures within the patient's anatomy. This allows the computation of dose-volume histograms (DVHs) for each structure, predicated on a specified dose distribution. The dose-volume objectives are subsequently represented as specific points on the DVH curve, which correspond to the desired minimum or maximum dose constraints. These objectives delineate the desired thresholds that should be upheld, with points on the DVH curve either located above (for minimum dose constraints) or below (for maximum dose constraints) the prescribed thresholds.

Dose Evaluation Physicians employ multiple criteria to assess the quality of a radiation dose administered during treatment. Initially, they scrutinize the three-dimensional distribution of the dose across the patient's anatomy, focusing on the spatial allocation among different anatomical structures, as well as identifying the presence, number, and locations of regions with excessive radiation (referred to as "hot spots"). Subsequently, physicians conduct a thorough analysis of the dose-volume histograms (DVHs) to evaluate the degree of compliance with predefined DVH objectives. This crucial evaluation step aims to safeguard the adjacent healthy tissues from unnecessary radiation exposure. By optimizing the treatment plan and meticulously assessing the quality of the dose distribution, physicians strive to ensure the attainment of the most favorable outcome for the patient.

2   Methods

To ensure the precision and effectiveness of radiation therapy, a robust dose optimization process are essential.

2.1   Radiotherapy Dosimetry

Radiotherapy dose optimization can be conceptualized as an inverse problem, whereby the objective is to determine the most suitable radiation dose distribution that aligns with the desired treatment outcome [21]. In other terms, the challenge lies in identifying the radiation intensity or fluence maps that deliver the prescribed dose to the tumor while minimizing exposure to healthy tissues.

Mathematical optimization algorithms are employed to address the inverse problem of radiotherapy dose optimization. These algorithms aim to identify the optimal solution by minimizing a predefined objective function that encompasses treatment goals and constraints. Typically, the objective function includes terms that penalize both underdosing and overdosing of the tumor and overdosing of surrounding healthy tissues. It may also incorporate terms that account for the complexity or deliverability of the treatment plan.

Efficiently solving this optimization problem often involves designing the objective function to be convex, thereby providing a well-defined target for the optimization process. Gradient-based methods, Newtonian algorithms, or quasi-Newtonian algorithms are commonly employed for this purpose. We aim at benchmarking state-of-the-art open-source optimization algorithms for the specific task of radiotherapy dosimetry.

2.2   Data

In this research endeavor, our focus was to evaluate the various open-source optimizers. We used the widely recognized TG-119 [14] cases as a benchmark for evaluating radiation therapy plans optimization. The TG-119 dataset provides specific dose goals, which we incorporated into our proposed cost function.

We also used one real case of prostate cancer treatment from ICM. For this case, doctors had provided specific dose goals, that we again incorporated into our proposed cost function.

The simulation of the beams was done using TheraPanacea dose engine, which uses collapse cone convolution techniques, and is conformal to other simulator available on the market.

2.3   Objective function

The cost function is formulated as a weighted sum of multiple objectives, with each objective corresponding to a specific dose goal. The formulation is as follows:

$$f(\mathbf{d}) = \sum_{o \in \mathcal{O}} w_o f_o(\mathbf{d})$$

where:

- $\mathbf{d}$ represents the dose distribution at the voxel level, and $\mathbf{d}[s]$ denotes the dose on voxels within the structure s

- $\mathcal{O}$ denotes the set of objectives corresponding to dose volume goals

- w_o signifies the weight assigned to the objective $o \in \mathcal{O}$

- o_s, o_d, and o_v refer to the structure, dose, and volume goals of the objective $o \in \mathcal{O}$

The objective function $f_o(\mathbf{d})$ is computed based on the specific type of dose volume constraint: If o represents a maximum dose volume constraint[1], $f_o(\mathbf{d})$ is calculated as $\sum_{d \in \mathbf{d}[o_s]}(d - o_d)^2_+$; if o represents a minimum dose volume constraint[2], then $f_o(\mathbf{d})$ is calculated as $\sum_{d \in \mathbf{d}[o_s]}(o_d - d)^2_+$. The formulation involves a squared over/under-dose penalty function.

In addition to the above, we introduced a regularization term that penalizes variations in bixel values between neighboring regions, also employing a squared penalty.

The optimization process involves finding the optimal bixel values ($\mathbf{b}$) by solving $\mathbf{d} = \mathbf{Lb}$, where $\mathbf{L}$ is a precomputed dose-influence matrix mapping bixels to voxels. Notably, since negative energy rays are physically infeasible, we ensured that each bixel value is non-negative ($b \geq 0 \quad \forall b \in \mathbf{b}$). To achieve this, we computed $\mathbf{d} = \mathbf{L}|\mathbf{b}|$, where $|\mathbf{b}|$ denotes the element-wise absolute value of $\mathbf{b}$.

By construction, the objective function is convex. Consequently, minimizing the objective function with a given set of weights should invariably converge to the same radiotherapy plan. To generate different treatment doses for the same patient case, dosimetrists can play withe the weights of each sub-objective function. This is outside the scope of this small review article, so we decided to just set the weights of all constraints equal to one.

2.4 Open-source Optimizers

We tried to have a comprehensive test of available open-source optimizers, here is a short description of the ones tested:

(Stochastic) Gradient Descent Is an optimization algorithm that iteratively updates the model parameters in the direction of the negative gradient of the objective function. In our case, it is not stochastic, since it calculates the gradient using the current solution[3] [8].

Conjugate Gradient Is an iterative optimization algorithm commonly used to solve systems of linear equations or quadratic optimization problems. It iteratively computes conjugate directions and updates the solution along these directions, aiming to minimize the objective function [6]. Conjugate Gradient is often applied in scenarios where the Hessian matrix is unavailable or expensive to compute.

---

[1]e.g.: top 20% of the volume should receive at most 30 Gy
[2]e.g.: 95% of the volume should receive at least 70 Gy
[3]Our objective function has all its inputs as parameters, so there is no notion of stochasticity.

Newton Newton's method is an iterative optimization algorithm that uses the second-order derivative (Hessian matrix) to find the minimum of a function. It updates the current estimate by taking into account both the first-order derivative (gradient) and the second-order derivative [15].

SLSQP (Sequential Least Squares Programming) is a sequential quadratic programming algorithm used for constrained optimization. It iteratively solves a sequence of quadratic programming subproblems to find the optimal solution subject to constraints [2].

RMSprop (Root Mean Square Propagation) is an optimization algorithm that addresses the problem of diminishing learning rates in traditional gradient descent methods. It divides the learning rate by the root mean square of the past gradients, which helps to stabilize and speed up convergence [7].

BFGS-based

Pure BFGS (Broyden-Fletcher-Goldfarb-Shanno) is a quasi-Newton method that approximates the Hessian matrix using updates based on gradient information. It performs a line search to determine the step size that minimizes the objective function along the search direction [5].

L-BFGS (Limited-memory BFGS) is a variation of BFGS that uses a limited-memory approach to approximate the Hessian matrix. It stores a limited number of past gradient and parameter values to compute an approximate inverse Hessian matrix efficiently [11].

Adam-based

Pure Adam (Adaptive Moment Estimation) is an optimization algorithm that combines ideas from both adaptive learning rates and momentum methods. It computes adaptive learning rates for each parameter based on estimates of the first and second moments of the gradients [9].

RAdam (Rectified Adam) is a variant of the Adam optimizer that introduces a rectification term to stabilize the adaptive learning rate. It aims to address some convergence issues that can occur in Adam by dynamically adjusting the variance of the adaptive learning rate [12].

NAdam (Nesterov Adam) combines the Nesterov accelerated gradient method with the Adam optimizer. It incorporates Nesterov momentum into the Adam update rule to improve convergence and provide better generalization [19].

AdamDelta Is another variant of the Adam optimizer that replaces the second moment estimates (variance) with a delta parameter. It eliminates the need for storing and updating the moving average of the squared gradients, which can be beneficial in memory-constrained settings [22].

Adamax   Is an extension of the Adam optimizer that uses the infinity norm (max norm) of the gradients instead of the L2 norm. It is designed to handle sparse gradients more effectively and can be particularly useful in deep learning models [1].

Rprop   (Resilient Backpropagation) is an optimization algorithm specifically designed for neural networks. It updates the weights based on the sign of the gradient, adjusting the step size adaptively. Rprop performs weight updates independently for each weight parameter [17].
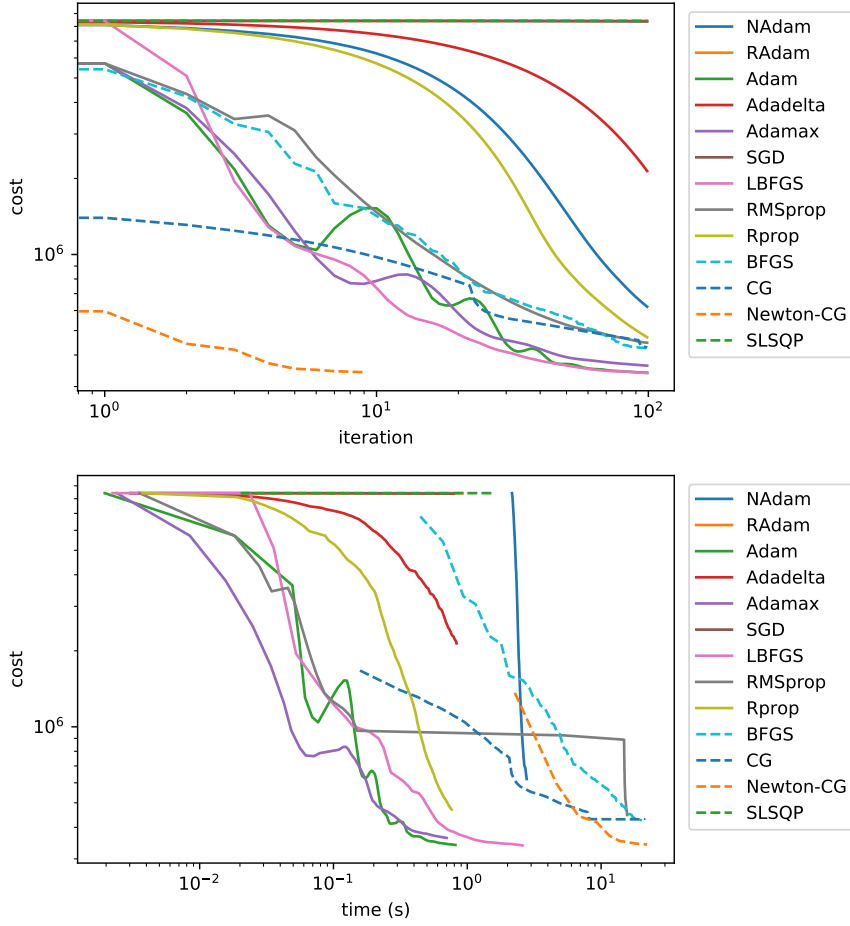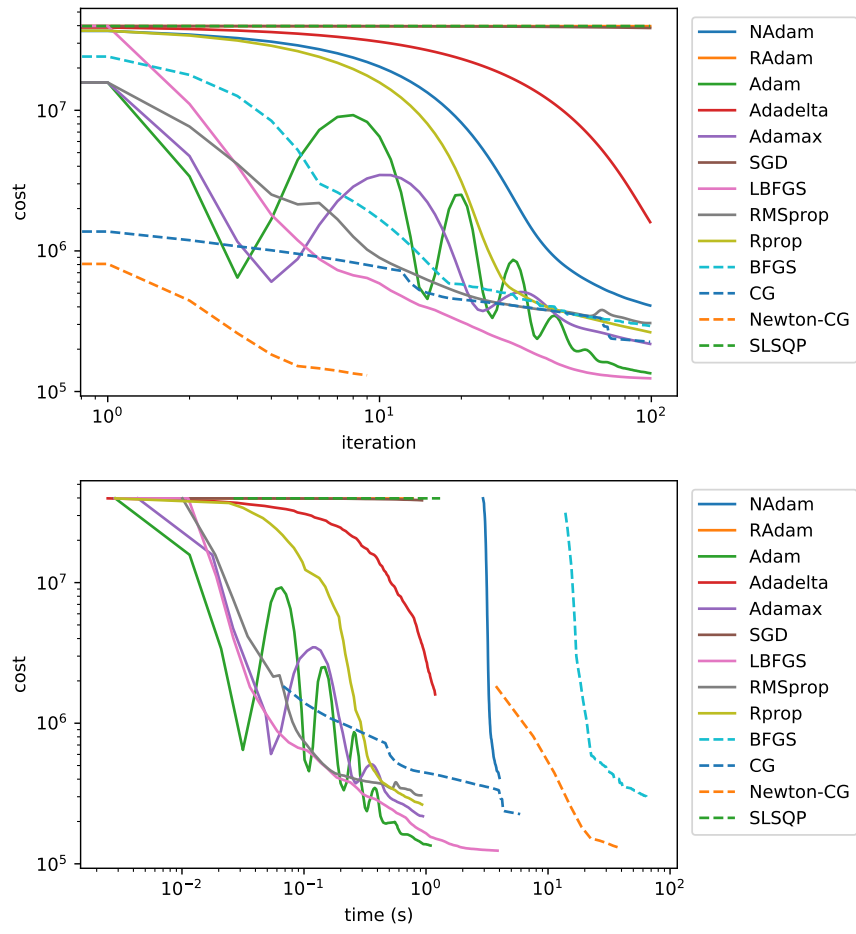
3   Results

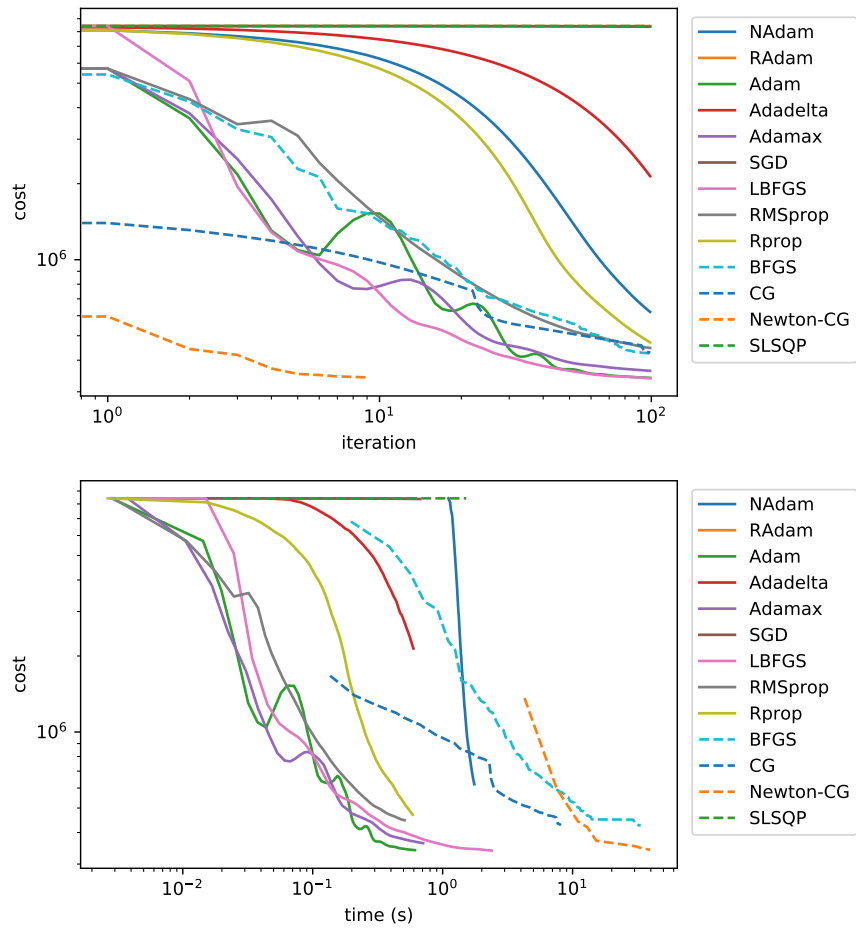Figure 1: TGG 119: Multiple PTVs

Figure 2: TGG 119: Head and Neck
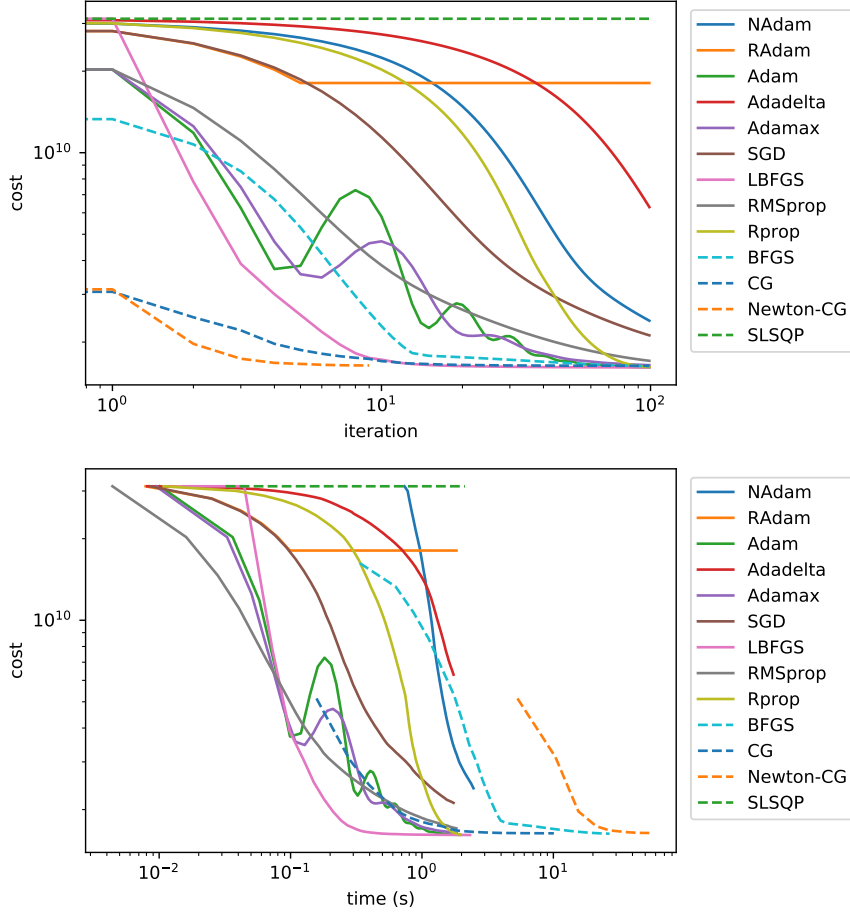
Figure 3: TGG 119: Prostate

Figure 4: ICM: (Typical) Prostate

Newton's method   Looking only at iterations-wise graph, Newton's method appears to be the very best, reaching within 10 steps a stable converged state in all 4 cases studied. However, Newton's method steps are very expensive to compute, since it uses second order derivative (the Hessian) that is expensive to compute (in terms of calculation time).

Best Algorithms   Other than Newton's method, three algorithms have similar performances: Adam, RMSprop and LBFGS. Adam (and RMSprop, to a lesser extent) appear to have more "wavy" cost curves, while LBFGS cost decreases in a more stable way. These observations are true both iteration and time-wise.

TGG 119 Multiple PTVs is the smallest problem, and th real ICM prostate case is the largest problem (in terms of patient/organs/structure volume size). It appears that as the problem size gets bigger, LBFGS performs better than RMSprop and Adam.

LBFGS vs BFGS It would be expected that BFGS performs better than LBFGS in terms of iterations, but not in terms of time (since LBFGS is a fast approximation of the BFGS technique). However, we observe that LBFGS outperforms BFGS even on the iterations-wise graph. This suggests that the limited memory approximation made are biased towards suitable directions in these type of problems.

4   Discussion

References

[1] Somenath Bera and Vimal K Shrivastava. Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification. *International Journal of Remote Sensing*, 41(7):2664–2683, 2020.

[2] JF Bonnans, Jean Charles Gilbert, C Lemaréchal, and Claudia Sagastizabal. *Numerical Optimization – Theoretical and Practical Aspects.* 01 2006.

[3] Thomas Bortfeld. Imrt: a review and preview. *Physics in Medicine and Biology*, 51(13):R363, jun 2006.

[4] Gary A. Ezzell, James M. Galvin, Daniel Low, Jatinder R. Palta, Isaac Rosen, Michael B. Sharpe, Ping Xia, Ying Xiao, Lei Xing, and Cedric X. Yu. Guidance document on delivery, treatment planning, and clinical implementation of imrt: Report of the imrt subcommittee of the aapm radiation therapy committee. *Medical Physics*, 30(8):2089–2115, 2003.

[5] Roger Fletcher and Chengian Xu. Hybrid methods for nonlinear least squares. *IMA Journal of Numerical Analysis*, 7(3):371–389, 1987.

[6] Magnus R. Hestenes and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*, 49:409–435, 1952.

[7] Geoffrey Hinton, Nitish Srivastava, and Kevin Swersky. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. *Cited on*, 14(8):2, 2012.

[8] J.B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis.* Grundlehren Text Editions. Springer Berlin Heidelberg, 2012.

[9] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[10] Thomas P. Kole, Osarhieme Aghayere, Jason Kwah, Ellen D. Yorke, and Karyn A. Goodman. Comparison of heart and coronary artery doses associated with intensity-modulated radiotherapy versus three-dimensional conformal radiotherapy for distal esophageal cancer. *International Journal of Radiation Oncology*Biology*Physics*, 83(5):1580–1586, 2012.

[11] Dong C. Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45(1):503–528, Aug 1989.

[12] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond, 2021.

[13] Erick Mora-Ramirez, Lore Santoro, Emmanuelle Cassol, Juan C Ocampo-Ramos, Naomi Clayton, Gunjan Kayal, Soufiane Chouaf, Dorian Trauchessec, Jean-Pierre Pouget, Pierre-Olivier Kotzki, Emmanuel Deshayes, and Manuel Bardiès. Comparison of commercial dosimetric software platforms in patients treated with ^{177} Lu-DOTATATE for peptide receptor radionuclide therapy. *Med. Phys.*, 47(9):4602–4615, September 2020.

[14] Dinesh Kumar Mynampati, Ravindra Yaparpalvi, Linda Hong, Hsiang-Chi Kuo, and Dennis Mah. Application of aapm tg 119 to volumetric arc therapy (vmat). *Journal of Applied Clinical Medical Physics*, 13(5):108–116, 2012.

[15] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer series in operations research and financial engineering. Springer, New York, NY, 2nd ed. edition, 2006. MAB0014.001: M 11.0268.

[16] David Palma, Emily Vollans, Kerry James, Sandy Nakano, Vitali Moiseenko, Richard Shaffer, Michael McKenzie, James Morris, and Karl Otto. Volumetric modulated arc therapy for delivery of prostate radiotherapy: Comparison with intensity-modulated radiotherapy and three-dimensional conformal radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, 72(4):996–1001, 2008.

[17] Martin A. Riedmiller and Heinrich Braun. Rprop - a fast adaptive learning algorithm. 1992.

[18] Bin Tang, Xiangyu Liu, Xianliang Wang, Shengwei Kang, Pei Wang, Jie Li, and Lucia Clara Orlandini. Dosimetric comparison of graphical optimization and inverse planning simulated annealing for brachytherapy of cervical cancer. *J Contemp Brachytherapy*, 11(4):379–383, August 2019.

[19] Ange Tato and Roger Nkambou. Improving adam optimizer. 2018.

[20] Lynn J. Verhey. Comparison of three-dimensional conformal radiation therapy and intensity-modulated radiation therapy systems. *Seminars in Radiation Oncology*, 9(1):78–98, 1999. Radiation Therapy Treatment Optimization.

[21] S Webb. The physical basis of imrt and inverse planning. *The British Journal of Radiology*, 76(910):678–689, 2003. PMID: 14512327.

[22] Matthew D. Zeiler. Adadelta: An adaptive learning rate method, 2012.