Swarm and Evolutionary Computation 26 (2016) 191–201

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

journal homepage: www.elsevier.com/locate/swevo

Regular Paper

# Investigations of a GPU-based levy-firefly algorithm for constrained optimization of radiation therapy treatment planning

Georgios Kalantzis [a], Charles Shang [a,c], Yu Lei [b], Theodora Leventouri [a]

[a] Florida Atlantic University, Department of Physics, Boca Raton, FL 33431, USA
[b] University of Nebraska Medical Center, Department of Radiation Oncology, Omaha, NE 68198, USA
[c] Lynn Cancer Institute, Boca Raton Regional Hospital, Boca Raton, FL 33486, USA

ARTICLE INFO

*Article history:*
Received 19 December 2014
Received in revised form
16 August 2015
Accepted 22 September 2015
Available online 26 October 2015

*Keywords:*
FireFly algorithm
Optimization
Radiation therapy
IMRT
GPU

ABSTRACT

Intensity modulated radiation therapy (IMRT) affords the potential to decrease radiation therapy associated toxicity by creating highly conformal dose distribution to tumor. Inverse optimization of IMRT treatment plans is often a time intensive task due to the large scale solution space, and the indubitably complexity of the task. Furthermore, the incorporation of conflicting dose constraints in the treatment plan, usually introduces an additional degree of intricacy. Metaheuristic algorithms have been proposed in the past for global optimization in IMRT treatment planning. However one disadvantage of the aforementioned methods is their extensive computational cost. One way to ameliorate their performance deficiency is to parallelize the application. In the current study we propose a GPU-based levy-firefly algorithm (LFA) for constrained optimization of IMRT treatment planning. The evaluation of our method was realized for two treatment cases: a prostate and a head and neck (H&N) cancer IMRT plans. The studies indicated an ascendable increase of the speedup factor as a function of the number of pencil beams with a maximum of $\sim$11, whereas the performance of the algorithm was decreasing as a function of the population of the swarm particles. In addition, from our simulation results we concluded that 200 fireflies were sufficient for the algorithm to converge in less than 80 iterations. Finally, we demonstrated the effect of penalizing factors on constraining the maximum dose at the organs at risk (OAR) by impeding the dose coverage of the tumor target. The impetus behind our study was to elucidate the performance and generic attributes of the proposed algorithm, as well as the potential of its applicability for IMRT optimization problems.



## 1. Introduction

Intensity Modulated Radiation therapy (IMRT) is an advanced means of treatment modality delivering highly modulated external megavoltage radiation beams utilizing linear accelerators. For many types of cancer, such as prostate and nasopharyngeal cancer the use of IMRT allows a highly intensive treatment of the tumor volume while limiting the radiation dose to adjacent healthy tissues [1,2]. In contrast to the traditionally 3D conformal radiotherapy (3DCRT) where the treatment is delivered with large uniform beams, in IMRT the intensity of each beam varies within the treatment field. This is achieved by dividing the radiation field into a collection of pencil beams (beamlets) [3,4]. When planning a radiation therapy case, the dose constraints are assigned to both the target and surrounding normal structures. These dose constraints need not to be given in terms of the dose assigned to each point in the body, but rather are usually phrased in terms of aggregate functions such as maximum or minimum dose, and dose limit for a given volume (dose-volume constraints). Then, according to the prescribed dose objectives, numerical inverse optimization is performed in order to determine the individual intensities of the beamlets [5,6]. The optimized intensity maps are then decomposed into a series of deliverable multi-leaf collimator (MLC) patterns in the sequencing step. It is worth note that alternative optimization methods, such as the direct aperture optimization (DAO), have been proposed in the past where the aperture shapes and beamlet intensities are optimized simultaneously. In that way all of the MLC delivery constraints of the leaf sequencing algorithm are included in the optimization [7–9]. Traditionally, the optimization model contains a single objective function subject to a set of hard constraints on the treatment plan.

A number of increasingly sophisticated mathematical programming models have been proposed for the inverse treatment planning and deconvolution process (see, e.g. Ref. [10] for a recent overview of this topic). For instance, the well-known Newton–Raphson algorithm is gradient-based and it works well for smooth unimodal problems. Gradient-based algorithms have been proposed in the past for optimizing single-objective problems in radiation therapy treatment planning under volume-dose constraints [11–14]. However for optimization problems with

http://dx.doi.org/10.1016/j.swevo.2015.09.006
2210-6502/

discontinuity, a derivative-free non-gradient algorithm (i.e. Nelder–Mead downhill simplex) is preferred. Zhu and Xing recently proposed a total-variation based compressed sensing technique to better balance the tradeoff between fluence modulation complexity and deliverability [15]. Kalantzis et al. [16] have introduced an accelerated reduced order prioritized optimization method where the IMRT optimization is performed in a pre-sampled eigen mode space of the beamlets intensities. Romeijn et al. [17] have followed a linear programming approach for fluence-based IMRT optimization by approximating any convex objective function by a piecewise linear convex function [17]. Wang et al. [18] have employed mixed integer linear programming (MILP) to optimize beam orientations and beam weights, whereas Xing et al. [19] have used a nonlinear programming model, the weighted least squares for IMRT optimization.

An alternative approach to the aforementioned deterministic methods is the Metaheuristic Algorithms (MAs) which form an important part of contemporary global optimization algorithms. MAs are often nature-inspired and they are now among the most widely used algorithms for optimization problems [20]. These algorithms have been developed by mimicking the most successful processes in nature, including biological systems as well as physical and chemical processes. Convergence analysis of a few algorithms such as the particle swarm optimization (PSO) show some insight, but in general mathematical analysis of MAs remains unsolved and still an ongoing active research topic [21]. The main components of any metaheuristic algorithm are: intensification and diversification, or exploitation and exploration. Diversification aims to generate diverse solutions so as to explore the search space on the global scale, while intensification focus on the search in a local region by exploiting the information that a current solution is found in this region [22]. MAs have been proposed in the past for dosimetric optimization of LINAC-based IMRT treatment [23–25] and robotic radiosurgery [26]. Simulated annealing [27], PSO [28] and a hybridized genetic algorithm with an ant colony [29] have been also applied to beam angle optimization (BAO). Finally, other metaheuristics, such as the Bat Algorithm [30] and Memetics [31,32] have also found applications in optimization of IMRT and Gamma Knife treatment planning.

A global optimization stochastic algorithm which has attracted interest from researchers is the Firefly algorithm (FA). As a novel literature, the FA is a metaheuristic, nature inspired optimization algorithm developed by X. Yang [33], it is based on the idealized behavior of the emitted light from the fireflies, in the summer sky in the tropical temperature regions. Although the FA has many similarities with other swarm intelligence algorithms, such as Artificial Bee Colony (ABC), Bacterial Foraging (BFA) and Particle Swarm Optimization (PSO), it is indeed much simpler both in concept and implementation [34–36]. Additionally, recent developments have demonstrated the superiority of the FA in performance compared to other metaheuristic algorithms for solving various optimization tasks [37–40]. One of the key advantages of the FA is the global communication among the swarming particles (i.e. fireflies), which can provide a quick convergence by switching from exploration to exploitation. However, if we allow the algorithm to switch to exploitation stage too quickly, it may lead to stagnation soon after the initial stage.

Due to the high dimensionality of the search space, IMRT optimization is a computationally demanding task. One way to ameliorate that issue is the parallelization of the optimization algorithm. Previous studies have signified promising results towards that direction for various computational platforms. Ziegenhein et al. [41] have demonstrated a parallelized quasi-Newton method on a multi-core CPU with the usage of pre-calculated dose influence data sets. Na et al. [42] suggested a web-based radiation therapy planning system at the Amazon Elastic Compute Cloud (EC2). Men et al. [43] proposed a gradient projection method for GPU-based quadratic optimization model for IMRT treatment planning. Although all metaheuristic algorithms are simple in terms of complexity and easy to implement, they require extensive computational resources. That is due to the extensive iterative calculations and random number generators required for their execution. That issue becomes more apparent for large scale optimization problems with thousands variables, such as the IMRT optimization. Efforts have been made towards the parallelization of MAs for radiation planning optimization. Nazareth et al. [44], describes the use of a genetic algorithm that is run on a distributed computing platform for BAO. Fiege et al. [25], describe the application of a parallel Matlab-based multiobjective genetic algorithm (Ferret) for IMRT optimization.

This paper suggests, for the first time to our best knowledge, an immerging GPU-based Levy Firefly algorithm for constraint optimization in radiation therapy treatment planning. The applicability of the proposed method is demonstrated through two cases of IMRT treatment planning: an early stage prostate cancer and a head and neck (H&N) cancer case. Performance tests were conducted for both cases in comparison to a sequential version of the algorithm executed on a CPU.

## 2. Methods

### 2.1. Treatment planning preparation

The treatment planning system we have used in this investigation is the computational environment for radiotherapy research (CERR) [45], a MATLAB®-based implementation of a treatment planning suite for radiation therapy (Fig. 1).

The main advantages of CERR is the user-friendly graphical user interface (GUI), access to dose deposition matrices, implementation of programming modules from the user and availability of the toolboxes of MATLAB®. Based on the defined target and structures of the planning CT set, the number and orientation of 6 MV photon beams to be used in the treatment are selected by the planner. In conformal therapy using uniform or flat radiation fields, the arrangement of the beams is carefully selected to spare the beam projectiles without collisions and minimize overlap with the normal organs at risk (OARs). For IMRT plans each beam is further subdivided into a rectilinear grid of beamlets with typical size of $0.5\times0.5\ \mathrm{cm}^2$, but may be as small as $0.2\times0.2\ \mathrm{cm}^2$. An optimization algorithm is then utilized to find the beamlet intensities to maximize the dose to the planning tumor volume (PTV) while minimizing the dose at the OARs. Fig. 1 illustrates an example of prostate cancer treatment plan with a single anterior beam. The intensities of its beamlets are depicted schematically on the top part of the beam.

In an IMRT treatment planning process, a two dimensional photon fluence map (a set of beamlet intensities that can be controlled individually) must be specified for each beam Fig. 2). Then the calculated dose to each unit of volume (voxel) for all the structures provides the feedback of optimization [46,47]. In the context of IMRT optimization the requirements are in the form of the prescription dose to the PTV, subject to maximum dose constraints at the organs at risk (OARs), and can be described as follows:

$$f(x)=\arg\min\sum_{i\in T}\frac{(D_i-D_{PTV})^2}{N_T}\quad i\in PTV$$
$$s.t.\quad \max D_j\le D_{\max,OARj}\quad \forall j\in OAR_j \tag{1}$$

where $N_T$ is the number of voxels in the target structure or else PTV, $D_{PTV}$ is the prescribed dose of the target structure, $D_j$ and