Implementation of Gradient Projection Algorithm to Radiation Therapy Treatment Planning

Laleh Kardar, Gino J. Lim, Jiming Peng, Wenhua Cao, Guven Kaya
Department of Industrial Engineering
University of Houston, Houston, TX, USA

Abstract

The fluence map optimization (FMO) is a key element in radiation therapy treatment planning. In this study, we implement two types of gradient projection algorithm to a dose-based objective function formulated as a bound-constrained quadratic programming (BCQP) problem. The purpose of this implementation is to assess the validity and convergence properties of these algorithms. In addition, we investigate the possibility of being trapped in a local minimum when using different initial intensity distributions. We use two cancer cases to illustrate the performance of these algorithms, including one prostate and one head-and-neck case. Our preliminary results indicate that the algorithms are producing clinically acceptable plans in a reasonable amount of time, and essentially converge to the same plans regardless of the starting point used in the algorithms.

Keywords
Radiation therapy, optimization, gradient projection algorithm, fluence map optimization.

1. Introduction

The fluence map optimization problem is one of the most widely studied problems of radiation therapy treatment planning that determines the fluence of each beamlet in each beam. The FMO problem has received considerable attention recently, and a substantial amount of discussions regarding model formulations as well as algorithmic design have appeared in the literature. Different types of objective functions have been used to optimize the fluence of beamlets including dose-volume-based, dose-based, and biology-based. Deasy [1] pointed out that several model formulations for the FMO model are in fact unimodal (a model with a single local minimum). However, some FMO models, particularly those that employ dose-volume constraints, are nonconvex which may lead to the existence of multiple local minima. A combination of a unimodal model formulation and an efficient algorithm would be a more robust approach to solving the FMO problem to optimality.

Many optimization algorithms, including stochastic algorithms [2,3] and gradient based algorithms [4,5], have been utilized to solve the FMO problem. Stochastic algorithms such as simulated annealing (SA) are used when one expects multiple local minima in the objective function over the feasible region. Stochastic algorithms have the advantage of escaping local entrapment. However, they can be slow in reaching the global optimum. Gradient based algorithms are commonly used for FMO because of their convergence. But there is a risk of being trapped in local minima [1]. As a successful example, Wu and Mohan [6] optimized several clinically relevant plans multiple times with randomly assigned initial intensities (i.e., starting points), and found that resulting plans converged to essentially the same dose distribution. They argued that the presence of the multiple local minima are not considered impediments in finding satisfactory solutions in the routine optimization of intensity modulated radiation therapy (IMRT) plans using gradient methods. Zhang et al. [5] compared the relative convergence speed properties of different gradient algorithms for dose-volume and biological objective functions. Their study has demonstrated that it is reasonable to use Newton’s method without being concerned about being trapped at local minima as long as uniform or random initial intensity distributions are used.

Unlike unconstrained optimization algorithms discussed above, linear programming (LP) based algorithms have been proposed to impose clinically acceptable constraints such as upper and lower dose bound constraints [7,8]. However, LP has not garnered much attention in the medical physics community due to its limitations in formulating suitable objective functions and constraints. On the other hand, quadratic programming (QP) has received a great deal of attention in radiation therapy, with many algorithms proposed in this type of formulation. It is well known that, even
for a unimodal model, many algorithms are not guaranteed to solve it to optimality without needing unlimited computational time. Also, many algorithms have no means to recognize the fact that a globally optimal solution has been found. In fact, it has been pointed out that such algorithms occasionally get trapped in local minima, resulting in inferior treatment plans.

Therefore, we focus our study on the performance of a gradient projection (GP) algorithm to solve FMO problem in intensity modulated proton therapy (IMPT) in this paper. A dose-based objective function is used to optimize the fluence of beamlets. The convergence properties of the GP algorithm are examined by applying uniform or random initial intensity distributions to two types of cancer cases: one prostate and one head-and-neck case.

The remainder of the paper is organized as follows. In Section 2, we describe the optimization model and the implementation of the GP algorithm to solve the model. We also describe our method of choosing initial intensity distributions in this section. In Section 3, we present our results for the convergence properties of gradient projection algorithm with discussions. Finally, conclusions are presented in Section 5.

2. Method and Materials

As mentioned previously, we define our objective function in terms of dose limits. We use two types of gradient projection methods for optimizing the objective function. The following subsections describe the optimization model and the solution methods in detail.

2.1 Optimization model

Our model considers different types of structures such as the planning target volume (PTV) which includes the cancerous area; the organs-at-risk (OARs) or critical structures that are located close to the PTV. Without loss of generality, our model assumes that the structures are irradiated using a set of beams, each corresponding to a particular beam angle. For each beam angle, the aperture of this beam is decomposed into small beamlets. In practice, each structure is discretized into a finite number of cubes, which are known as voxels. The dose at a voxel is the sum of dose contributions from all the beamlets from all beams,

\[ D_i = \sum_{j=1}^{M} d_{ij} x_j, \]

where \( D_i \) is the dose in the \( i \)th voxel, \( i = 1 \ldots N \); \( x_j \) is the decision variable to the model which represents the weight for the \( j \)th beamlet, \( j = 1 \ldots M \). All \( x_j \) should fulfill a physical constraint requirement, where only positive fluencies can be delivered, \( x_j \geq 0 \); \( d_{ij} \) is the dose contribution of the \( j \)th beamlet to the \( i \)th voxel at unit weight. The objective function is composed of multiple objectives corresponding to individual anatomic structures. Each objective is assigned a weight or penalty that reflects the relative importance of the structure compared with other ones. The selection of penalty parameters is not intuitively obvious to the treatment planner and the values are often selected by trial and error. The objective function to be minimized in our model is defined as

\[
F(x) = \sum_{i=1}^{N} \frac{c_i}{2} (P_i - D_i)^2. \tag{1}
\]

Here \( x^T = (x_1, x_2, \ldots, x_M) \) denotes a vector of beamlet weights, \( N \) is the total number of voxels in the targets and OARs, \( P_i \) is the prescription radiation dose at voxel \( i \), and \( c_i \) is the associated penalty factor.

For the sake of simplicity and clarity of presentation, we rewrite problem (1) as the following bound-constrained quadratic programming (BCQP) problem:

\[
\min_x \frac{1}{2} \|y - Ax\|^2, \quad \text{s.t.} \quad x \geq 0, \tag{2}
\]

where \( y^T = (y_1, y_2, \ldots, y_N) \), \( y_i = \sqrt{c_i P_i} \), \( A \) is an \( N \times M \) matrix, and \( a_{ij} = \sqrt{c_i d_{ij}} \). Furthermore, problem (2) can be written in more standard BCQP form

\[
\min_x ( -A^T y )^T x + \frac{1}{2} x^T A^T A x = F(x), \quad \text{s.t.} \quad x \geq 0. \tag{3}
\]

The following subsection describes how we solve problem (3) using GP methods.
2.2 Gradient projection algorithm
For the GP method, the search path from each iterate is obtained by projecting the negative-gradient direction onto the feasible set. Suppose that $x^{(k)}$ is the beamlet weight vector (i.e., the intensity matrix) to be optimized at the $k$th iteration step. The optimization process at the $(k+1)$th iteration can be described as follows:

1. Choose some scalar parameter $\alpha^{(k)} > 0$ and set $w^{(k)} = (x^{(k)} - \alpha^{(k)} \nabla F(x^{(k)}))^+.$
2. Choose a second scalar $\lambda^{(k)} \in [0, 1]$ and set $x^{(k+1)} = x^{(k)} + \lambda^{(k)} (w^{(k)} - x^{(k)}).$

There are many variations of GP methods according to the way they choose $\alpha^{(k)}$ and $\lambda^{(k)}.$ The GP algorithms suggested by [9], basic and Barzilai-Borwein (BB) gradient projections, are employed in this paper. In these algorithms, $\nabla F(x)$ denotes the gradient of the objective function and is computed as follows:

$$\nabla F(x) = -A^T y + A^T Ax$$ \hspace{1cm} (4)

In the basic approach, we search through each iterate along the negative gradient $-\nabla F(x)$, projecting onto the nonnegative orthant, and performing a backtracking line search until a sufficient decrease is attained in $F$. Algorithm 1 shows the pseudocode for the basic GP algorithm. In contrast with the basic approach, the BB approach does not check the decrease in the objective function at every iteration. Algorithm 2 shows the pseudocode for the Barzilai-Borwein GP algorithm. These algorithms terminate if (1) the maximum number of iterations is exceeded, or (2) the change in objective function value is very small.

**Algorithm 1 Basic Gradient Projection**

1: Generate an initial solution $x^{(0)}$;
2: Choose parameters $\beta \in (0, 1)$ and $\mu \in (0, 1/2)$; $k := 0$;
3: while Stopping criteria are not met do
4: Define the vector $g^{(k)}$ by
5: 
6: Compute $\alpha_0$ as
7: 
8: Choose $\alpha^{(k)}$ to be the first number in the sequence $\alpha_0, \beta \alpha_0, \beta^2 \alpha_0, \ldots$ such that
9: 
10: Set $x^{(k+1)} = (x^{(k)} - \alpha^{(k)} \nabla F(x^{(k)}))^+$;
11: $k \leftarrow k + 1$;
12: end while
13: return $x^{(k)}$ as approximate solution.

2.3 Starting condition: initial beamlet weights
In our study, for each cancer case, the FMO models are optimized by GP methods using different initial sets of beamlet weights. These initial weights are as follows:

1. Uniform. All beamlet weights are set to be uniform.
2. Random. All beamlet weights are chosen randomly.

3. Numerical Experiments and Results
The GP methods were tested on two clinical cases: one prostate and one head and neck case. The volumes of interest (VOIs) and the number of voxels within each volume for each case are displayed in Table 1. Table 2 lists the dose-volume requirements for all VOIs. For each cancer case, we determined a set of objective function parameters to
Algorithm 2 Barzilai-Borwein Gradient Projection

1: Generate an initial solution $x^{(0)}$;
2: Choose parameters $\alpha_{\text{min}}$, $\alpha_{\text{max}}$, $\alpha^{(0)} \in [\alpha_{\text{min}}, \alpha_{\text{max}}]$; $k := 0$;
3: while Stopping criteria are not met do
4: Compute step: $\delta^{(k)} = (x^{(k)} - \alpha^{(k)} \nabla F(x^{(k)}))_{+} - x^{(k)}$
5: Calculate scalar
\[
\lambda^{(k)} = \text{mid} \left\{ 0, \left( \frac{(\delta^{(k)})^T \nabla F(x^{(k)})}{(\delta^{(k)})^T A^T A \delta^{(k)}}, 1 \right) \right\};
\]
6: Set $x^{(k+1)} = x^{(k)} + \lambda^{(k)} \delta^{(k)}$;
7: (update $\alpha$): compute $\gamma^{(k)} = (\delta^{(k)})^T A^T A \delta^{(k)}$
\[
\alpha^{(k+1)} = \begin{cases} 
\alpha_{\text{max}}, & \text{if } \gamma^{(k)} = 0 \\
\text{mid} \left\{ \alpha_{\text{min}}, \frac{||\delta^{(k)}||^2}{\gamma^{(k)}}, \alpha_{\text{max}} \right\}, & \text{otherwise};
\end{cases}
\]
8: $k \leftarrow k + 1$;
9: end while
10: return $x^{(k)}$ as approximate solution.

satisfy typical clinical requirements. For all cases, the penalty for the PTV was high relative to the OARs. This is because our highest priority is to satisfy tumor coverage requirements. Three beams were used to design the proton treatment plans for both cases.

Plans were analyzed using dose-volume histograms (DVHs) calculated for each plan and for the target and main OARs. In our analysis, in order to make a fair comparison among different methods and initial conditions and to be consistent with the common practice of treatment planning, all plans were normalized in such a way that 95% of the PTV receives the prescription dose. For uniform initial condition, we designed proton plans using different uniform initial intensity distributions. For random initial conditions, plans were optimized several times using different randomly sampled initial intensity distributions, and the mean and standard deviation of the results were reported. The same objective function parameters were used for all initial conditions.

Table 1: VOIs and number of voxels within each volume for different cases

<table>
<thead>
<tr>
<th>Case</th>
<th>Volume</th>
<th>Number of voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate</td>
<td>PTV</td>
<td>2,310</td>
</tr>
<tr>
<td></td>
<td>OAR (rectum)</td>
<td>2,462</td>
</tr>
<tr>
<td></td>
<td>OAR (bladder)</td>
<td>4,565</td>
</tr>
<tr>
<td>Head and Neck</td>
<td>PTV</td>
<td>5959</td>
</tr>
<tr>
<td></td>
<td>OAR (oral cavity)</td>
<td>2703</td>
</tr>
<tr>
<td></td>
<td>OAR (spinal cord)</td>
<td>342</td>
</tr>
<tr>
<td></td>
<td>OAR (optic chiasm)</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>OAR (left parotid)</td>
<td>439</td>
</tr>
<tr>
<td></td>
<td>OAR (right parotid)</td>
<td>523</td>
</tr>
</tbody>
</table>

3.1 Case A. Prostate

Tables 3 shows final objective function values for the GP methods using different initial conditions for the prostate case. We observed small differences in objective function values. The BB method turned out to be more robust than the basic method since there were less deviations in the objective function values of the achieved plans. Uniform initial conditions with the smaller initial weights outperformed those with higher initial weights and random initial conditions. Figure 1 compares the DVHs obtained using different GP methods with different initial weights. All treatment
Table 2: Dose volume requirements for the VOIs of different cases

<table>
<thead>
<tr>
<th>Case</th>
<th>Volume Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prostate</strong></td>
<td><strong>Prescription: 76 Gy</strong></td>
</tr>
<tr>
<td>PTV</td>
<td>Volume receiving at least the prescription dose ≥ 95%</td>
</tr>
<tr>
<td>Rectum</td>
<td>Volume receiving doses higher than 60 Gy: ≤ 40%</td>
</tr>
<tr>
<td>Rectum</td>
<td>Volume receiving doses higher than 70 Gy: ≤ 25%</td>
</tr>
<tr>
<td>Bladder</td>
<td>Volume receiving doses higher than 70 Gy: ≤ 25%</td>
</tr>
<tr>
<td><strong>Head and neck</strong></td>
<td><strong>Prescription: 74 Gy</strong></td>
</tr>
<tr>
<td>PTV</td>
<td>Volume receiving at least the prescription dose ≥ 95%</td>
</tr>
<tr>
<td>Spinal Cord</td>
<td>Max dose: 45 Gy</td>
</tr>
<tr>
<td>Optic chiasm</td>
<td>Max dose: 54 Gy</td>
</tr>
</tbody>
</table>

plans met clinical dose-volume constraints and produced similar target coverage and OARs sparing (Figure 1).

Table 3: Comparison of the final objective function values resulting from the use of GP algorithms with different initial conditions for the prostate case

<table>
<thead>
<tr>
<th>Method</th>
<th>Initial weights</th>
<th>Objective value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>Uniform-0</td>
<td>0.1766</td>
</tr>
<tr>
<td></td>
<td>Uniform-0.2</td>
<td>0.1766</td>
</tr>
<tr>
<td></td>
<td>Uniform-0.4</td>
<td>0.1772</td>
</tr>
<tr>
<td></td>
<td>Uniform-0.6</td>
<td>0.1778</td>
</tr>
<tr>
<td></td>
<td>Uniform-0.8</td>
<td>0.1790</td>
</tr>
<tr>
<td></td>
<td>Uniform-1</td>
<td>0.1807</td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td>0.1785 (0.0001)</td>
</tr>
<tr>
<td>BB</td>
<td>Uniform-0</td>
<td>0.1765</td>
</tr>
<tr>
<td></td>
<td>Uniform-0.2</td>
<td>0.1766</td>
</tr>
<tr>
<td></td>
<td>Uniform-0.4</td>
<td>0.1766</td>
</tr>
<tr>
<td></td>
<td>Uniform-0.6</td>
<td>0.1766</td>
</tr>
<tr>
<td></td>
<td>Uniform-0.8</td>
<td>0.1772</td>
</tr>
<tr>
<td></td>
<td>Uniform-1</td>
<td>0.1772</td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td>0.1773 (0.0001)</td>
</tr>
</tbody>
</table>

For random cases, a (b) denotes mean (standard deviation) of results.
In Uniform cases, numbers show the initial beamlet weights.

3.2 Case B. Head-and-neck
Both GP methods initiated with different beamlet weights were applied to a head-and-neck case. The final objective function values are tabulated in Table 4. For this case, again BB methods consistently produced better quality solutions with smaller deviations than the basic method in terms of objective function values. Both GP methods generated lower objective function values with small uniform initial weights when compared to those of large uniform or random choices of initial weights. The DVHs of all plans are displayed in Figure 2. The DVHs were nearly the same regardless of the GP algorithms and initial conditions used to optimize the plans. In all plans, the spinal cord and optic chiasm were within their tolerance limits of 45 and 54 Gy, respectively.

4. Conclusion
In order to address the inconsistent convergence issue of QP algorithms found in the medical physics literature, we presented two types of gradient projection algorithms (basic and Barzilai-Borwein) to solve the fluence map optimization problem in radiation therapy planning. The methods were implemented and compared on two separate clinical cancer cases. Different initial starting conditions were considered to investigate the robustness of the algorithms. The
algorithms showed better performance in terms of objective function value when they are initiated with small uniform weights, than the large uniform or random weights. Comparing the results of two GP methods on two clinical cancer cases demonstrated that the Barzilai-Borwein is more robust as it delivered more consistent solutions. Despite small differences in the objective function values, both GP methods produced similar target coverage and OAR sparing, even when different initial starting conditions were used.
Figure 2: Cumulative DVHs of structures from the solution of FMO problem employing the BB and basic GP methods (head-and-neck case)

References


