Noname manuscript No.
(will be inserted by the editor)

Bi-objective optimisation over a set of convex sub-problems

Guillermo Cabrera-Guerrero · Matthias Ehrgott · Andrew J. Mason · Andrea Raith

Received: date / Accepted: date

Abstract During the last decades, research in multi-objective optimisation (MO) has seen considerable growth. However, this activity has been focused on linear, non-linear, and combinatorial optimisation with multiple objectives. Multi-objective mixed integer (linear or non-linear) programming has received considerably less attention. In this paper we propose an algorithm to compute a finite set of non-dominated points/efficient solutions of a bi-objective mixed binary optimisation problems for which the sub-problems obtained when fixing the binary variables are convex, and there is a finite set of feasible binary variable vectors.

Our method uses bound sets and exploits the convexity property of the sub-problems to find a set of efficient solutions for the main problem. Our algorithm creates and iteratively updates bounds for each vector in the set of feasible binary variable vectors, and uses these bounds to guarantee that a set of exact non-dominated points is generated. For instances where the set of feasible binary variable vectors is too large to generate such provably optimal solutions within a reasonable time, our approach can be used as a matheuristic by heuristically selecting a promising subset of binary variable vectors to explore.

This investigation is motivated by the problem of beam angle optimisation arising in radiation therapy planning, which we solve heuristically to provide numerical results.

Guillermo Cabrera-Guerrero
Pontificia Universidad Católica de Valparaíso, Chile
E-mail: guillermo.cabrera@pucv.cl

Matthias Ehrgott
Lancaster University, United Kingdom
E-mail: m.ehrgott@lancaster.ac.uk

Andrew J. Mason
University of Auckland, New Zealand
E-mail: a.mason@auckland.ac.nz

Andrea Raith
University of Auckland, New Zealand
E-mail: a.raith@auckland.ac.nz

Keywords Multiple objective programming · convex optimisation · mixed binary programming · intensity modulated radiation therapy

1 Introduction

The aim of multi-objective optimisation (MO) is to find efficient solutions of an optimisation problem in the presence of multiple conflicting objectives. A solution is efficient if it cannot be improved in any criterion without degradation of some other criteria (Ehrgott, 2005). Several strategies to find efficient solutions have been proposed in the literature. While both continuous and combinatorial optimisation with multiple objectives have been widely studied (Ehrgott, 2005; Ehrgott et al., 2016; Miettinen, 1999; Wiecek et al., 2016), multi-objective mixed integer programming (MOMIP) problems have received considerably less attention. Teghem and Kunsch (1986) present a survey on interactive methods, which aim to find most preferred solutions to MOMIP problems. Mavrotas and Diakoulaki (1998, 2005) propose a branch and bound algorithm for bi-objective mixed binary linear programming problems. Vincent et al. (2013) present a branch and bound algorithm that corrects and improves the results obtained by the one in (Mavrotas and Diakoulaki, 1998, 2005). Belotti et al. (2013) also propose a branch and bound algorithm for bi-objective mixed integer problems. They focus on fathoming rules that exclude sub-problems that are guaranteed not to contain efficient solutions and solve sub-problems using a parametric simplex algorithm. Stidsen et al. (2014) present an algorithm for the special class of problems where one of the two objectives has only integer variables. Boland et al. (2014, 2015) present an exact method called the triangle splitting method for bi-objective mixed integer programming problems. Their algorithm works in objective space and maintains a diverse set of non-dominated points as an approximation of the non-dominated set in case of early termination.

In Fattahi and Turkay (2018), a one direction search algorithm that is able to solve biobjective mixed-binary linear problems exactly is introduced. Soylu (2018) published an algorithm which, in each iteration, identifies sub-problems that contribute to the non-dominated set and excludes them from future consideration using tabu constraints. The non-dominated set of the problem is computed by applying an upper envelope finding algorithm to the identified sub-problems.

Some authors have modelled problems arising in logistics as MOMIPs (Özceylan and Paksoy, 2014; Rezaei and Davoodi, 2011; Shirazi et al., 2014). In these works, the MOMIPs are addressed using general heuristic methods. Cacchiani and D'Ambrosio (2017) presented a branch-and-bound based heuristic to approximately solve convex multi-objective mixed integer non-linear programming problems and de Santis et al. (2019) propose a branch-and-bound algorithm to solve a convex MOMIP by using lower bounds generated in an adaptive way.

In this paper we consider a bi-objective version of a mixed integer non-linear programming (MINLP) problem that has both binary and continuous decision variables, and is convex in the continuous variables. We call this problem the BOMINLP (bi-objective mixed integer non-linear programming) problem. Although in this paper we only consider non-linear objective functions, we note that the method we propose in this work can also be applied to BOMINLP problems with

linear objective functions and constraints. The BOMINLP problem we consider in this paper is of the form:

$$\begin{aligned}
\text{BOMINLP: } \min_{x,y} \; & \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix} \\
\text{s.t.} \quad & g_j(x,y) \leqq b_j \quad \text{for } j = 1,\dots,m && \text{(1a)} \\
& g^x_j(x) \leqq b^x_j \quad \text{for } j = 1,\dots,m^x && \text{(1b)} \\
& g^y_j(y) \leqq b^y_j \quad \text{for } j = 1,\dots,m^y && \text{(1c)} \\
& y \in \{0,1\}^p && \text{(1d)} \\
& x \in \mathbb{R}^n, && \text{(1e)}
\end{aligned}$$

where $f(x,y) = (f_1(x,y), f_2(x,y))^\intercal$ is a vector-valued objective function with two conflicting objectives, m is the number of inequality constraints on x and y, m^x is the number of inequality constraints on x, m^y is the number of inequality constraints on y, p is the number of binary decision variables and n is the number of continuous decision variables.

We assume that functions $f_1(x,y)$, $f_2(x,y)$, $g_j(x,y)$ and $g^x_j(x)$ are continuously differentiable convex functions in x. We also impose regularity conditions to ensure Lagrangian dual values can be found, as discussed in Section 3.1. Note that although all our integer variables are binary valued, our approach extends naturally to more general bounded integer variables.

BOMINLP (1) may have an infinite number of non-dominated points due to the continuous variables. To our knowledge, the only method to compute an explicit description of the non-dominated set of continuous MOPs is a very recent one presented in Jayasekara et al. (2019). In their approach, which they demonstrate on problems with a small number of variables, the authors are able to solve convex MOP problems where all the objectives are quadratic ones. (This is in contrast to BOMILPs, where the continuous parts of the non-dominated set can be expressed as line segments, which algorithms listed at the beginning of Section 1 exploit). Therefore, we have to be content with finding only a finite subset of non-dominated points. Furthermore, having binary variables means BOMINLP (1) is not convex, and so scalarisation methods (see e.g. Wiecek et al. (2016)) are a natural strategy for generating the solution. In this paper, we choose to use the well-known ε-constraint method, which solves a number of single objective MINLPs to create the non-dominated points.

Our work is motivated by the *multi-objective beam angle optimisation* (MOBAO) problem in intensity modulated radiation therapy (IMRT) planning, which we describe in more detail in Section 5 and which we use as a case study for the application of our algorithm. Unfortunately, as we show in Section 5, it is not practical to apply the ε-constraint method to our IMRT BOMINLP problem because the single objective MINLP sub-problems take too long to solve. Hence we will instead pursue a decomposition strategy in which we create sub-problems, each of which has y fixed to one of a finite set of known solutions for y. We solve the ε-constraint scalarisations of these sub-problems, and combine the solutions to form a final non-dominated set. As we show in Section 5, our method is able to exploit convexity and differentiability to significantly reduce the number of problem instances that need to be solved. Our method is conceived as generic as this speed up does not depend on any other problem-specific features.

We mentioned that our approach assumes that a finite set of binary solutions for y is known. If this set can be explored within the available computation time, then the solutions found by our algorithm are efficient. For our IMRT problem, however, the number of binary solutions is too large for our algorithm to consider them all in a reasonable time. However, as is often the case for practical problems, we can exploit problem-specific knowledge to exclude 'bad' values for y that are unlikely to be part of an optimal solution. This approach of combining a heuristic selection of promising y with a subsequent optimisation over this set gives us a new matheuristic (Fischetti and Fischetti, 2016) approach for our problem. Although this compromises the guarantee of optimality in the sense that our solutions are no longer provably efficient, it gives us a practical approach that, as we see in Section 5, gives the same solution quality as existing approaches but in about one tenth of the time.

In the next section, we formally introduce our decomposition and define our sub-problems. The concept of *bound sets* (as introduced by Ehrgott and Gandibleux (2007)) is then presented in Section 3. Our proposed algorithm is presented in Section 4. We use a small bi-objective example to illustrate how our method proceeds. Then, in Section 5 the MO-BAO problem and its mathematical formulation are outlined. We use the MO-BAO problem as an application of our method and show computational results for a real case. Finally in Section 6, some conclusions and final remarks are presented.

2 Decomposition and Sub-Problem Definition

In this section, we formally introduce our decomposition and present the sub-problems we will be solving.

In model (1) and elsewhere we use the following notation for comparing vectors. Given $w^1, w^2 \in \mathbb{R}^2$, we write $w^1 \leqq w^2$ if $w^1_k \leqq w^2_k$ for all $k = 1, 2$; $w^1 \leq w^2$ if $w^1 \leqq w^2$ but $w^1 \neq w^2$; and $w^1 < w^2$ if $w^1_k < w^2_k$ for all $k = 1, 2$. We say that $w^1 \in \mathbb{R}^2$ *dominates* $w^2 \in \mathbb{R}^2$ (or, equivalently, w^2 is *dominated by* w^1) if $w^1 \leq w^2$.

Definition 1 We say a vector $y \in \{0,1\}^p$ is feasible if there exists some $x \in \mathbb{R}^n$ for which (x, y) is feasible for BOMINLP (1). We let $\mathcal{Y} = \left\{y^1, y^2, \ldots, y^{|\mathcal{Y}|}\right\} \subseteq \{0,1\}^p$ be the set of all such feasible y. For each $y \in \mathcal{Y}$ we let $\mathcal{X}(y) \subseteq \mathbb{R}^n$ denote the set of x such that (x, y) satisfies (1a) and (1e). We also let $\bar{\mathcal{X}} \subseteq \mathbb{R}^n$ denote the set of x that satisfy (1b) and (1e). Moreover, we denote by $(\mathcal{X}, \mathcal{Y}) = \bigcup_{y \in \mathcal{Y}} \left\{(x, y) : x \in \bar{\mathcal{X}} \cap \mathcal{X}(y)\right\}$, the set of all feasible solutions of (1).

Formally, solving BOMINLP (1) requires finding the set of non-dominated points and a set of efficient solutions, the image of which is the non-dominated set. A solution $(x^*, y^*) \in (\mathcal{X}, \mathcal{Y})$ is called an *efficient* solution if there is no $(x, y) \in (\mathcal{X}, \mathcal{Y})$ where $f(x, y)$ dominates $f(x^*, y^*)$. In this case $f(x^*, y^*)$ is called a *non-dominated* point.

Our decomposition strategy is motivated by the observation that BOMINLP (1) can be re-written as

$$\begin{aligned} \text{BOMINLP: } \min_{y} \quad & \left(\min_{x} \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix} \right) \\ \text{s.t.} \quad & x \in \bar{\mathcal{X}},\, x \in \mathcal{X}(y), \\ & y \in \mathcal{Y}. \end{aligned} \tag{2}$$

This formulation contains a sub-problem $SP(y)$ in x for every feasible $y \in \mathcal{Y}$ given by

$$\begin{aligned} SP(y): \min_{x} \quad & \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix} \\ \text{s.t.} \quad & x \in \bar{\mathcal{X}},\, x \in \mathcal{X}(y). \end{aligned} \tag{3}$$

Let $\mathcal{X}_E^y$ be the set of efficient solutions of sub-problem $SP(y)$. In the following we assume that $\mathcal{X}_E^y$ is non-empty.

Proposition 1 *1. If (x,y) is an efficient solution of (2) then $x \in \mathcal{X}_E^y \subseteq \bar{\mathcal{X}} \cap \mathcal{X}(y)$.*
2. $(\{f(x,y) : (x,y) \in (\mathcal{X},\mathcal{Y})\})_N = (\{f(x,y) : y \in \mathcal{Y}, x \in \mathcal{X}_E^y\})_N$, where $(S)_N$ denotes the non-dominated (sub)set of a set S.

Proof 1. This follows from the definition of all sets involved.
2. Both inclusions are easy to see by contradiction.

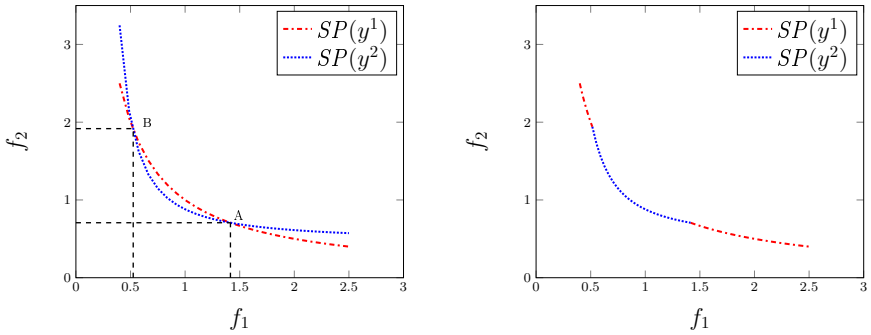
Proposition 1 states that if (x,y) is an efficient solution of BOMINLP (1), then x is also an efficient solution of sub-problem $SP(y)$. Thus, we can solve BOMINLP by finding all efficient solutions to $SP(y)$ (3) for each $y \in \mathcal{Y}$, and then performing a dominance analysis over these sub-problem solutions to create a final set of efficient solutions.

We observe that $SP(y)$ (3) is a bi-objective convex optimisation problem, and so it is computationally practical to compute efficient solutions to $SP(y)$; see, e.g., the survey of (Wiecek et al., 2016). We also note that because, for any given y, functions $f_1(x,y)$ and $f_2(x,y)$ in (3) are differentiable with respect to x, we can obtain Lagrangian (i.e. dual variable) values; we will use these to bound sub-problem solutions (see Section 3).

Example 1 *Let us consider an instance of BOMINLP (1) as follows.*

$$\begin{aligned} \min_{x,y} \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix} &= \begin{pmatrix} x \\ \frac{y_1}{x} + y_2 \left(0.2 + e^{1/x}\right) \end{pmatrix} \\ \text{s.t.} \quad x &\leqq 2.5 \\ -x &\leqq -0.4 \\ y_1 + y_2 &= 1 \\ y &\in \{0,1\}^2 \\ x &\in \mathbb{R}. \end{aligned}$$

In this example, $\mathcal{Y} = \{y^1, y^2\}$ with $y^1 = (1,0)$ and $y^2 = (0,1)$. Figure 1(a) shows the non-dominated points of sub-problems $SP(y^1)$ and $SP(y^2)$ where $A =$



(a) Sets of non-dominated points of sub-problems $SP(y^1)$ and $SP(y^2)$.

(b) Resulting set of non-dominated points of the BOMINLP (1).

Fig. 1: An illustration of the type of problem we address in this paper.

$(1.4151, 0.7071)$ *and* $B = (0.5150, 1.9172)$ *are non-dominated points of both sub-problems. For* $x \in [0.4, 0.5150]$ *and* $x \in [1.4151, 2.5]$*, non-dominated points of sub-problem* $SP(y^1)$ *are also non-dominated points of the BOMINLP problem, whereas for* $x \in [0.5150, 1.4151]$*, non-dominated points of* $SP(y^2)$ *are also non-dominated points of the BOMINLP problem. Figure 1(b) shows the set of non-dominated points of the BOMINLP problem. As we can see, although sub-problems* $SP(y^1)$ *and* $SP(y^2)$ *are convex, the BOMINLP problem may not be. A convex multi-objective optimisation problem, such as the subprobems here, has a* $\mathbb{R}^2_{\leqq}$*-convex non-dominated set, where set* S *is* $\mathbb{R}^2_{\leqq}$*-convex if* $S + \mathbb{R}^2_{\leqq}$ *is convex. It follows that convexity of* $SP(y)$ *does in general not imply convexity of* (1)*.*

By the assumption that sub-problem $SP(y)$ is convex for each $y \in \mathcal{Y}$, it follows that the set of non-dominated points of $SP(y)$ is $\mathbb{R}^2_{\leqq}$-convex Ehrgott (2005). Hence a natural decomposition approach to solving (1) is to determine the set of non-dominated points of the union (over elements of $\mathcal{Y}$) of the non-dominated sets of $SP(y)$. Furthermore, we do not expect all $y \in \mathcal{Y}$ to contribute to the non-dominated set of (1). Hence, by definition, solving (1) has three aspects.

(i) Finding those elements of $\mathcal{Y}$ for which the non-dominated points of sub-problem $SP(y)$ contribute to the non-dominated set of the BOMINLP problem.
(ii) Finding the non-dominated set of sub-problem $SP(y)$ for each y identified in *(i)*.
(iii) Finding the non-dominated set of BOMINLP from sets of non-dominated points of $SP(y)$.

Unfortunately, it is difficult to predict for which y the condition in *(i)* applies without solving sub-problem $SP(y)$ for all $y \in \mathcal{Y}$. If the computation times are not limited, then we can explore all elements of $\mathcal{Y}$ to identify the non-dominated points needed for steps *(ii)* and *(iii)*. However, this may not be practical if $\mathcal{Y}$ is very large or the computation times are limited. In Cabrera G. et al. (2018) and Cabrera-Guerrero et al. (2018), we have proposed heuristics that create a set $\hat{\mathcal{Y}} \subset \mathcal{Y}$ of "promising" feasible binary vectors. In this paper we therefore address

(ii) and *(iii)* for the case where $\mathcal{Y}$ is replaced by $\hat{\mathcal{Y}}$. We need to stress at this point that if $\hat{\mathcal{Y}} \neq \mathcal{Y}$, then the set of points our method finds for the BOMINLP problem is not necessarily equal to the set of non-dominated points of (1) the method would find if the entire set $\mathcal{Y}$ were considered. Thus, we describe the points obtained by our method when using a subset of $\mathcal{Y}$ as *approximately* non-dominated points. The selection of a set of promising vectors $\hat{\mathcal{Y}} \subseteq \mathcal{Y}$ is out of the scope of this paper. For the sake of simplicity, hereafter we assume that $\hat{\mathcal{Y}} = \mathcal{Y}$ and we shall refer to it simply as $\mathcal{Y}$.

As discussed above, the well-known ε-constraint method is a suitable approach for step (ii) above, i.e. for finding a non-dominated set for $SP(y)$. This approach requires forming the ε-constraint scalarisation $SP(\varepsilon, y)$ of $SP(y)$. This $SP(\varepsilon, y)$ sub-problem is a single-objective convex optimisation problem of the form:

$$\begin{aligned} SP(\varepsilon, y): \min_{x} \ & f_1(x, y) \\ \text{s.t. } & x \in \bar{\mathcal{X}},\ x \in \mathcal{X}(y), && \text{(4a)} \\ & f_2(x, y) \leqq \varepsilon. && \text{(4b)} \end{aligned}$$

We need to solve $SP(\varepsilon, y)$ for all $\varepsilon \in \mathcal{E}$, where $\mathcal{E}$ is a set of ε values that determines which non-dominated points the user wishes to find for BOMINLP.

A naïve approach to implementing steps (ii) and (iii) above would involve solving each of the $|\mathcal{Y}|$ sub-problems $|\mathcal{E}|$ times, i.e. solving $|\mathcal{Y}| \times |\mathcal{E}|$ instances of (4). As mentioned earlier, we will demonstrate in Section 5 that our method is able to exploit convexity and differentiability of $f_1(x, y)$ and $f_2(x, y)$ to significantly reduce the number of instances of (4) that are solved.

For each sub-problem $SP(y)$, it is useful to define an associated range of valid ε values, given by $\varepsilon \in [\varepsilon_y^{\min}, \varepsilon_y^{\max}]$, where $\varepsilon_y^{\min}$ is the smallest ε for which $SP(\varepsilon_y^{\min}, y)$ is feasible, i.e. $\varepsilon_y^{\min} = \min\{f_2(x, y) : x \in \bar{\mathcal{X}}, x \in \mathcal{X}(y)\}$. We determine $\varepsilon_y^{\max}$ by finding the largest f_2-value of any non-dominated point using the standard approach of calculating a lexicographically minimal solution $x^{(1,2)}$, as detailed in Section 4. The ε-constraint method allows us to find any efficient solution to $SP(y)$ by solving sub-problem $SP(\varepsilon, y)$ for some $\varepsilon_y^{\min} \leq \varepsilon \leq \varepsilon_y^{\max}$. It is easy to see that putting $\varepsilon < \varepsilon_y^{\min}$ cannot generate any efficient solutions, and so we never solve $SP(\varepsilon, y)$ for such ε values. Lemma 1 considers the case that constraint (4b) is not binding.

Lemma 1 *Let x^* be an optimal solution of $SP(\varepsilon, y)$ such that $f_2(x^*, y) < \varepsilon$. Then $SP(\varepsilon, y)$ has no efficient solution $\hat{x}$ such that $f_2(\hat{x}, y) > f_2(x^*, y)$.*

Proof Assume that x' is an efficient solution to $SP(\varepsilon, y)$ such that $f_2(x', y) > f_2(x^*, y)$. Then $f_1(x', y) < f_1(x^*, y)$. By optimality of x^* for $SP(\varepsilon, y)$ it follows that $f_2(x', y) > \varepsilon$. By continuity of f_2 and by application of the Intermediate Value Theorem, there exists some $\lambda \in (0, 1)$ such that for $\hat{x} = \lambda x' + (1-\lambda)x^*$ it holds that $f_2(\hat{x}, y) = \varepsilon$, i.e. $\hat{x}$ is feasible for $SP(\varepsilon, y)$. Due to convexity $f_1(\hat{x}, y) = f_1(\lambda x' + (1-\lambda)x^*, y) \leq \lambda f_1(x', y) + (1-\lambda) f_1(x^*, y) < f_1(x^*, y)$. This is a contradiction to x^* being an optimal solution of $SP(\varepsilon, y)$.

Lemma 1 confirms that we only need to consider $\varepsilon \leq \varepsilon_y^{\max}$. With $\varepsilon \in [\varepsilon_y^{\min}, \varepsilon_y^{\max}]$, constraint (4b) is binding and we obtain only efficient solutions; see Proposition 2.

Proposition 2 *For any fixed y, if $\varepsilon_y^{\min} \leq \varepsilon \leq \varepsilon_y^{\max}$ then any optimal solution x^* of $SP(\varepsilon, y)$ is an efficient solution of $SP(y)$ and satisfies $f_2(x^*, y) = \varepsilon$.*

Proof Due to our choice of $\varepsilon_y^{\min}, \varepsilon_y^{\max}$ we know that efficient solutions $\bar{x}$ and $\hat{x}$ exist with $f_2(\bar{x}, y) = \varepsilon_y^{\min}$ and $f_2(\hat{x}, y) = \varepsilon_y^{\max}$. For any ε with $\varepsilon_y^{\min} < \varepsilon < \varepsilon_y^{\max}$ Lemma 1 implies that there exists an optimal solution x^* of $SP(\varepsilon, y)$ such that $f_2(x^*, y) = \varepsilon$.

Assume that x^* is weakly efficient, but not efficient. Then there exists x' that dominates x^* with $f_1(x', y) = f_1(x^*, y)$ and $f_2(x', y) < f_2(x^*, y) = \varepsilon$. We consider a feasible solution $\tilde{x} = \lambda x' + (1 - \lambda)\hat{x}$ with $\lambda \in (0, 1)$. Due to convexity, and noting that $f_1(\hat{x}, y) < f_1(x', y)$, parameter $\lambda \in (0, 1)$ can be chosen so that $f_1(\tilde{x}, y) < f_1(x^*, y)$ and $f_2(\tilde{x}, y) < f_2(x^*, y)$. It follows that x' strictly dominates x^*, a contradiction to the weak efficiency of x^*.

We show, in Section 4, how the values $\varepsilon_y^{\min}$ and $\varepsilon_y^{\max}$ can be found by solving lexicographic optimisation problems. If either (or both) of these values is infinitely large, then we assume that the value is replaced by a new finite value that defines a range of interest to the user.

We assume that elements in $\mathcal{E}$ are indexed such that $\mathcal{E} = \{\varepsilon^1, \ldots, \varepsilon^L\}$ where $\varepsilon^1 < \varepsilon^2 < \cdots < \varepsilon^{L-1} < \varepsilon^L$. Later, in Section 4, we explain how to compute the ε values in set $\mathcal{E}$. To keep notation simple, we will only compute one set $\mathcal{E}$ which we use for all $y \in \mathcal{Y}$, and so $SP(\varepsilon, y)$, $\varepsilon \in \mathcal{E}, y \in \mathcal{Y}$ defines the full set of sub-problems considered by our algorithm. However, we note that because of the way the values are computed in the algorithm, the algorithm will only ever solve $SP(\varepsilon, y)$ for ε values satisfying $\varepsilon_y^{\min} \leq \varepsilon \leq \varepsilon_y^{\max}$. Solving $SP(\varepsilon, y)$ for all $\varepsilon \in \mathcal{E}$ means we compute (at most) L non-dominated points of sub-problem $SP(y)$. Hereafter, we will refer to this subset of non-dominated points of sub-problem $SP(y)$ obtained by means of the ε-constraint method simply as the set of non-dominated points of sub-problem $SP(y)$.

3 Bounds from Convexity

In this section, we demonstrate how the convexity of sub-problem $SP(y)$ can be used to give lower and upper bounds for $SP(\varepsilon, y)$ that we can use to efficiently search for non-dominated points of (1). The work in this section draws upon the ideas of lower bound sets and upper bound sets introduced in Ehrgott and Gandibleux (2007).

3.1 Sub-problem Lower Bounds

We wish to form lower bound sets on the non-dominated points of a sub-problem $SP(y)$ defined by some particular $y \in \mathcal{Y}$. Assume we have solved $SP(\varepsilon, y)$ for L' ε values $\varepsilon \in \{\hat{\varepsilon}^1, \hat{\varepsilon}^2, ..., \hat{\varepsilon}^{L'}\}$, where $\hat{\varepsilon}^1 < \hat{\varepsilon}^2 < \cdots < \hat{\varepsilon}^{L'}$ and $\hat{\varepsilon}^1 = \varepsilon_y^{\min}$ and $\hat{\varepsilon}^{L'} = \varepsilon_y^{\max}$. Let the optimal solution of $SP(\hat{\varepsilon}^i, y)$ be $x^*_{\hat{\varepsilon}^i}$. Because $\varepsilon_y^{\min} \leq \hat{\varepsilon}^i \leq \varepsilon_y^{\max}$ for all i, constraint (4b) is active at the optimal solution, and so $f_2\left(x^*_{\hat{\varepsilon}^i}, y\right) = \hat{\varepsilon}^i$. Then, our current set of non-dominated points of $SP(y)$ can be denoted by

$$\mathcal{S}_N^y = \left\{ \left(f_1\left(x^*_{\hat{\varepsilon}^1}, y\right), \hat{\varepsilon}^1 \right), \left(f_1\left(x^*_{\hat{\varepsilon}^2}, y\right), \hat{\varepsilon}^2 \right), \ldots, \left(f_1\left(x^*_{\hat{\varepsilon}^{L'}}, y\right), \hat{\varepsilon}^{L'} \right) \right\}. \quad (5)$$

We assume that each sub-problem $SP(\hat{\varepsilon}^i, y)$ satisfies some appropriate regularity conditions, and so we can determine $\lambda_{\hat{\varepsilon}^i}$, the value of the Lagrangian dual variable on the constraint (4b) at each optimal solution. We present an example of regularity conditions in Section 5 for our radiation therapy planning problem. The Lagrangian dual value $\lambda_{\hat{\varepsilon}^i}$ gives the rate of change in the optimal objective value of $SP(\hat{\varepsilon}^i, y)$ (i.e. the $f_1(x, y)$ value of some non-dominated point) as we increase $\hat{\varepsilon}^i$ (i.e. increase $f_2(x, y)$). This means that the Lagrangian duals are interpreted as trade-offs between the two objectives of sub-problem $SP(y)$. This issue has been discussed in Haimes and Chankong (1979) and Chankong and Haimes (1983, Chapter 4.9). They give sufficient conditions for $SP(y)$ under which this interpretation is correct.

Thus, because the non-dominated set is $\mathbb{R}^2_{\geqq}$-convex, we can follow an approach similar to that of Benders (1962), and use $\mathcal{S}_N^y$ to calculate a lower bound set $\mathcal{L}(\mathcal{S}_N^y)$ on the non-dominated points of $SP(y)$ as follows:

$$\mathcal{L}(\mathcal{S}_N^y) = \left(\bigcap_{i=1}^{L'} \left\{ (f_1', f_2') \in \mathbb{R}^2 : f_1' \geqq f_1\left(x^*_{\hat{\varepsilon}^i}, y\right) + \lambda_{\hat{\varepsilon}^i}\left(f_2' - \hat{\varepsilon}^i\right) \right\} \right)_N, \tag{6}$$

where $(f_1', f_2') \in \mathbb{R}^2$ is a point in objective space.

We are interested in a lower bound on the objective function value of $SP(\varepsilon, y)$ (i.e. on $f_1(x, y)$) for a new ε value, $\varepsilon_y^{\min} \leq \varepsilon \leq \varepsilon_y^{\max}$, (i.e. for a new $f_2(x, y)$ value) for which $SP(\varepsilon, y)$ is as-yet unsolved. This bound, denoted by $f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y)$, is given by $f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y) = f_1'$ where $(f_1', f_2') \in \mathcal{L}(\mathcal{S}_N^y)$ is the unique element of $\mathcal{L}(\mathcal{S}_N^y)$ for which $f_2' = \varepsilon$. To compute this efficiently, we find i such that $\hat{\varepsilon}^i \leq \varepsilon < \hat{\varepsilon}^{i+1}$, if it exists. We then have

$$f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y) = \max\left\{ f_1\left(x^*_{\hat{\varepsilon}^i}, y\right) + \lambda_{\hat{\varepsilon}^i}\left(\varepsilon - \hat{\varepsilon}^i\right), f_1\left(x^*_{\hat{\varepsilon}^{i+1}}, y\right) + \lambda_{\hat{\varepsilon}^{i+1}}\left(\varepsilon - \hat{\varepsilon}^{i+1}\right) \right\}. \tag{7}$$

If $\varepsilon < \hat{\varepsilon}^1$ or $\hat{\varepsilon}^{L'} < \varepsilon$ then solving $SP(\varepsilon, y)$ cannot generate any non-dominated points for $SP(y)$, which we indicate by putting $f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y) = +\infty$.

Figure 2 shows, for some sub-problem $SP(y)$, a set of non-dominated points $\mathcal{S}_N^y$ and (in dashed lines) the resulting lower bound set $\mathcal{L}(\mathcal{S}_N^y)$, i.e., the value of $f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y)$ for any ε value (i.e. for any value of f_2). Each dashed line is the tangent line to a non-dominated point $\left(f_1(x^*_{\hat{\varepsilon}^i}, y), \hat{\varepsilon}^i\right) \in \mathcal{S}_N^y$ with a slope equal to $\lambda_{\hat{\varepsilon}^i}$. This figure also shows upper bounds, which we discuss next.

3.2 Sub-problem Upper Bounds

We can compute upper bounds for a sub-problem $SP(y)$ as follows. Assume, as before, that we have $\mathcal{S}_N^y$ as given by (5). By convexity, an upper bound set $\mathcal{U}(\mathcal{S}_N^y)$ for the non-dominated points of $SP(y)$ is given by the non-dominated frontier of the convex hull of $\mathcal{S}_N^y$:

$$\mathcal{U}(\mathcal{S}_N^y) = (\text{conv}(\mathcal{S}_N^y))_N. \tag{8}$$

As with the lower bound, we can efficiently compute elements in $\mathcal{U}(\mathcal{S}_N^y)$ for some new ε, $\varepsilon_y^{\min} < \varepsilon < \varepsilon_y^{\max}$, by finding i such that $\hat{\varepsilon}^i \leq \varepsilon < \hat{\varepsilon}^{i+1}$ and then calculating

$$f_{\text{upper}}(\varepsilon, \mathcal{S}_N^y) = (1 - \alpha) f_1\left(x^*_{\hat{\varepsilon}^i}, y\right) + \alpha f_1\left(x^*_{\hat{\varepsilon}^{i+1}}, y\right) \tag{9}$$

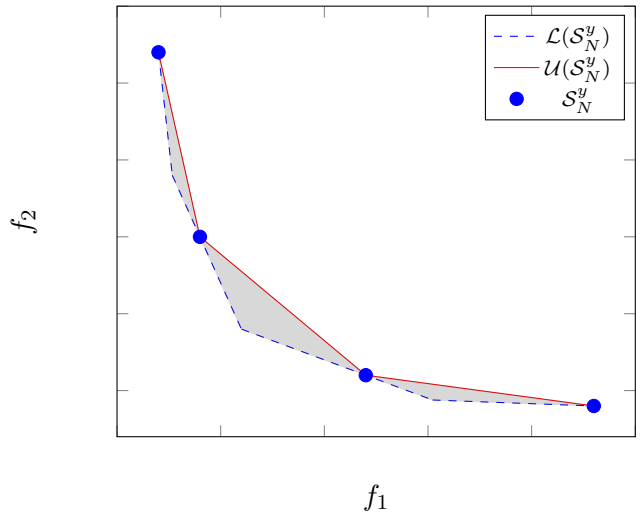


Fig. 2: Lower and upper bound sets $\mathcal{L}(\mathcal{S}_N^y)$ and $\mathcal{U}(\mathcal{S}_N^y)$ for a set of non-dominated points $\mathcal{S}_N^y$ of $SP(y)$.

where $\alpha = (\varepsilon - \hat{\varepsilon}^i)/(\hat{\varepsilon}^{i+1} - \hat{\varepsilon}^i)$. If $\varepsilon < \hat{\varepsilon}^1$ or $\hat{\varepsilon}^{L'} < \varepsilon$ then we put $f_{\text{upper}}(\varepsilon, \mathcal{S}_N^y) = +\infty$.

Figure 2 shows in solid lines the upper bound set $\mathcal{U}(\mathcal{S}_N^y)$. Due to the convexity of $SP(y)$, non-dominated points of $SP(y)$ are located in $(\mathcal{L}(\mathcal{S}_N^y) + \mathbb{R}^2_{\leqq}) \setminus (\mathcal{U}(\mathcal{S}_N^y) + \mathbb{R}^2_{>})$, the shaded area in Figure 2.

3.3 Joint Bounds

In general, given some $\mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}}$, we can form joint lower and upper bound sets that give us information on the non-dominated points of BOMINLP (1) as follows. A lower bound set for (1) is given by

$$\mathcal{L}(\mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}}) = \left(\bigcup_{y \in \mathcal{Y}} \left(\mathcal{L}(\mathcal{S}_N^y) + \mathbb{R}^2_{\leqq} \right) \right)_N . \tag{10}$$

Similarly, an upper bound set for (1) is given by

$$\mathcal{U}(\mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}}) = \left(\bigcup_{y \in \mathcal{Y}} \left(\mathcal{U}(\mathcal{S}_N^y) + \mathbb{R}^2_{\leqq} \right) \right)_N . \tag{11}$$

The non-dominated points of (1) must lie in the region between the lower and upper bound sets. An example of such a region is given by the shaded area in Figure 4d.

Our algorithm uses these bounds as follows. Given some particular ε, we calculate the following joint upper bound:

$$f_{\text{upper}}(\varepsilon, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}}) = \min_{y \in \mathcal{Y}} f_{\text{upper}}(\varepsilon, \mathcal{S}_N^y). \tag{12}$$

Then, we know that solving $SP(\varepsilon, y)$ for any (ε, y) pair with $f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y) > f_{\text{upper}}(\varepsilon, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$ is not required as the point $(f_1(x_\varepsilon^*, y), \varepsilon)$ we will compute will have $f_1(x_\varepsilon^*, y) > f_{\text{upper}}(\varepsilon, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$, and thus the point cannot improve the current set of non-dominated points of (1).

4 Proposed Method

A naïve method to address the BOMINLP problem is to use the ε-constraint method to compute a set of non-dominated points of sub-problem $SP(y)$ for each $y \in \mathcal{Y}$, and to then perform a dominance analysis over the union of these sets to obtain a set of pairwise non-dominated points of (1). We will use the following example to illustrate the performance of this naïve method and contrast it with our more efficient approach.

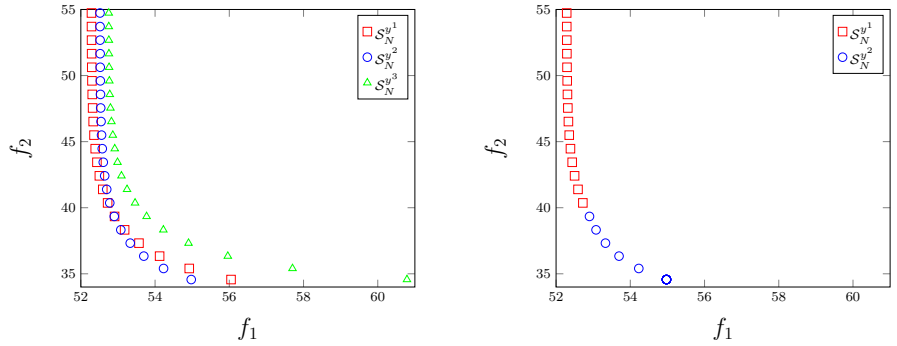
Example 2 *Let $\mathcal{Y} = \{y^1, y^2, y^3\}$. Let $\mathcal{E} = \{\varepsilon^1, \ldots, \varepsilon^L\}$, with $L = 21$. We wish to find 21 non-dominated points of (1), one for each $\varepsilon \in \mathcal{E}$. In fact, functions $f_1(x, y)$ and $f_2(x, y)$ are taken from the numerical example in Section 5, but their precise form is not important at this stage. If we apply the naïve method to this problem then $L = 21$ non-dominated points for each $y \in \mathcal{Y}$ must be computed. That is, we solve optimisation problem $SP_1(\varepsilon, y)$ $L \times |\mathcal{Y}| = 63$ times. Figure 3(a) shows the set of non-dominated points of $SP(y)$ for each $y \in \mathcal{Y}$. Figure 3(b) shows the resulting set of non-dominated points after performing the dominance analysis over the sets shown in Figure 3(a). As we can see, most of the points computed by the naïve method turn out to be dominated by at least one other point. For instance, all the non-dominated points found for $SP(y^3)$ are dominated and, therefore, none of them contributes to the final set of non-dominated points of (1) in Figure 3(b). Thus, if either obtaining a single non-dominated point of $SP(y)$ requires a considerable amount of time or the number of non-dominated points to be computed is large, the naïve method can be impractical.*

We propose a method, outlined in Algorithm 1, that significantly reduces the number of non-dominated points computed for each sub-problem. Our algorithm finds all the non-dominated points of (1) that would have been found by the naïve method but it uses bounding information to reduce the number of sub-problems that need to be solved. The proposed method generates the bounds by exploiting the convexity of sub-problem $SP(y)$.

In order to determine the range $[\varepsilon_y^{\min}, \varepsilon_y^{\max}]$ of ε values to be used when evaluating $SP(\varepsilon, y)$ we first find lexicographically non-dominated points of $SP(y)$ for all $y \in \mathcal{Y}$ (lines 2-5 in Algorithm 1).

As demonstrated in Ehrgott (2005), a lexicographically optimal solution in $\bar{\mathcal{X}} \cap \mathcal{X}(y)$ is also an efficient solution of $SP(y)$. We denote our lexicographic optimisation problems $SP^{(k,l)}(y)$, where $(k, l) \in \{(1, 2), (2, 1)\}$ indicates that objective k is most important, followed by objective l, as follows:

$$\begin{aligned} SP^{(k,l)}(y): \min_x \quad & f_l(x, y) \\ \text{s.t.} \quad & f_k(x, y) = f_k^*(y), \\ & x \in \bar{\mathcal{X}},\ x \in \mathcal{X}(y). \end{aligned}$$



(a) Sets of non-dominated points computed by the naïve method for $SP(y)$, for all $y \in \mathcal{Y} = \{y^1, y^2, y^3\}$.

(b) Set of non-dominated points of (1) with sub-problems $SP(y^1)$, $SP(y^2)$ and $SP(y^3)$. Only y^1, y^2 lead to non-dominated points of (1).

Fig. 3: The naïve method applied to Example 2. The final set of non-dominated points of (1) found by the naïve method given $\mathcal{E}$ is obtained by finding a set of L non-dominated points of $SP(y)$ for all $y \in \mathcal{Y}$. A dominance analysis is performed over the union of these sets of points.

Here $f_k^*(y)$ is obtained as the optimal objective value of an optimisation problem that minimises the most important objective k, i.e. $f_k^*(y) = \min_x\{f_k(x, y) : x \in \bar{\mathcal{X}}, x \in \mathcal{X}(y)\}$.

We solve $SP^{(1,2)}(y)$ and $SP^{(2,1)}(y)$ for each $y \in \mathcal{Y}$ to obtain lexicographically optimal solutions $x^{(1,2)}(y)$ and $x^{(2,1)}(y)$ for $SP^{(1,2)}(y)$ and $SP^{(2,1)}(y)$, respectively. These define the lower and upper limits $\varepsilon_y^{\min} = f_2(x^{(2,1)}(y), y)$ and $\varepsilon_y^{\max} = f_2(x^{(1,2)}(y), y)$ on ε for each sub-problem $SP(\varepsilon, y)$. This information is recorded by using these solutions to initialise the set of non-dominated points S_N^y, giving $S_N^y = \{f(x^{(1,2)}(y), y), f(x^{(2,1)}(y), y))\}$, for all $y \in \mathcal{Y}$ (line 4 in Algorithm 1). We also use these points to determine the first and last elements in $\mathcal{E}$, ε^1 and ε^L, as follows (lines 6 and 7 in Algorithm 1, respectively):

$$\varepsilon^1 = \min_{y \in \mathcal{Y}} f_2(x^{(2,1)}(y), y), \tag{13}$$

$$\varepsilon^L = \max_{y \in \mathcal{Y}} f_2(x^{(1,2)}(y), y), \text{ with } f(x^{(1,2)}(y), y) \text{ non-dominated within } \bigcup_{y \in \mathcal{Y}} \mathcal{S}_N^y. \tag{14}$$

If we want to focus on non-dominated points within a specific part of the objective space, we can limit ε values to be within a range $[\varepsilon^{\min}, \varepsilon^{\max}]$, where both $\varepsilon^{\min}$ and $\varepsilon^{\max}$ should be given by the decision maker and $[\varepsilon^{\min}, \varepsilon^{\max}] \subseteq [\varepsilon^1, \varepsilon^L]$. It is important to note that Equation (14) is restricted to lexicographic points that are not dominated by other lexicographic points. This ensures that ε^L will not be set to unnecessarily large values.

Once ε^1 and ε^L have been computed, values of $\varepsilon^i \in \mathcal{E}$, with $i = 2, \cdots, (L-1)$, are determined such that $\varepsilon^1 < \varepsilon^2 < \cdots < \varepsilon^{L-1} < \varepsilon^L$ (line 8 in Algorithm 1). These values are arbitrary values that will be determined depending on the problem that

is being solved. In this paper we choose them to be uniformly distributed over the range $\left[\varepsilon^1, \varepsilon^L\right]$.

We are now ready to perform the main iterative steps in Algorithm 1. The iterations proceed by identifying some $\varepsilon \in \mathcal{E}$ (line 10 in Algorithm 1) and some $y \in \mathcal{Y}$ for which $f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y) < f_{\text{upper}}(\varepsilon, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$ (line 12 in Algorithm 1). We call the sub-problem $SP(\varepsilon, y)$ a *candidate sub-problem* in the sense that it may produce a new non-dominated point in our final solution, and thus it needs to be solved. (Sub-problems $SP(\varepsilon, y)$ with $f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y) \geq f_{\text{upper}}(\varepsilon, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$ are *bounded sub-problems* and do not need solving as they cannot contribute to our final solution.) There will often be many candidate problems that could be solved next; we discuss shortly a strategy for choosing a good ε. If there are multiple candidate problems $SP(\varepsilon, y)$ for this ε, then we choose $SP(y)$ with the best lower bound by putting $y = \arg\min_{y \in \mathcal{Y}} f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y)$. We then solve $SP(\varepsilon, y)$ to find a new non-dominated point (line 13 in Algorithm 1), which is added to S_N^y (line 14 in Algorithm 1). After this step, we have strictly increased $f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y)$ or strictly decreased $f_{\text{upper}}(\varepsilon, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$ (or both). These iterations continue until $f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y) \geq f_{\text{upper}}(\varepsilon, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$ for all $\varepsilon \in \mathcal{E}$, $y \in \mathcal{Y}$. At this point, we merge the sets $\mathcal{S}_N^y$ to give $\mathcal{S}_N := \left(\bigcup_{y \in \mathcal{Y}} \mathcal{S}_N^y\right)_N$ (line 17 in Algorithm 1) as the set of non-dominated points of (1) obtained by our method.

As noted above, each iteration decreases the gap between the best lower bound and the upper bound for some $\varepsilon \in \mathcal{E}$. These bounds must become equal after solving no more than $|\mathcal{Y}|$ instances of $SP(\varepsilon, y)$ for each ε. Thus, our algorithm must eventually terminate after no more than $L|\mathcal{Y}|$ iterations. Our experience is that the bounding process results in far fewer iterations being required in practice.

4.1 Selection of Next ε – $\texttt{getBest}\varepsilon\left(\mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}}\right)$

We observed above that there is often a choice of $\varepsilon \in \mathcal{E}$ for which to solve $SP(\varepsilon, y)$ next. Therefore, we need to design a strategy to choose the next ε so that points obtained by solving $SP(\varepsilon, y)$ make changes in the lower and/or upper bounds that lead to more sub-problems becoming bounded. Unfortunately, determining the best strategy is difficult as it is dependent on problem-specific features. Initial experimental results show that, for instance, choosing an ε value close to previously explored ε values is a very inefficient strategy. We can see this in Figure 2: Let us assume that ε values are for f_2. Then, assume we chose an ε value that is very close to an ε value used to compute one of the blue points. If we now solve $SP(\varepsilon, y)$ with the chosen ε the new generated point is likely to be very close to the blue point. Thus, the new point will only produce a small change in the lower and/or upper bounds. Clearly, small changes in the bounds are not very helpful as they are unlikely to lead to many candidate sub-problems becoming bounded. Thus, on line 10 Algorithm 1 includes a function $\texttt{getBest}\varepsilon(\mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$ that aims to choose an ε value that leads to large changes in the bounds. In our experiments, choosing the ε value as the midpoint between the current ε value, $\varepsilon^{current}$, and a smaller ε^s value, where ε^s has already been explored and all the ε values between ε^s and $\varepsilon^{current}$ remain unexplored, has performed well. In case there is no ε^s that meets these conditions, then the function $\texttt{getBest}\varepsilon(\mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$ looks for

an explored ε^a value starting from ε^L, such that ε^{a-1} remains unexplored. Once an ε^a is found, it is set as the new $\varepsilon^{current}$ value. Then, the ε^s is selected as described above and the midpoint ε value between $\varepsilon^{current}$ and ε^s is set as the new ε value to be explored. If we cannot find an ε^a, then it means that all ε values have been explored and, therefore, our algorithm stops. When choosing the new ε, this function also ensures that there is at least one y for which $\varepsilon_y^{\min} \leq \varepsilon \leq \varepsilon_y^{\max}$, and so there is at least one candidate sub-problem $SP(\varepsilon, y)$ to solve.

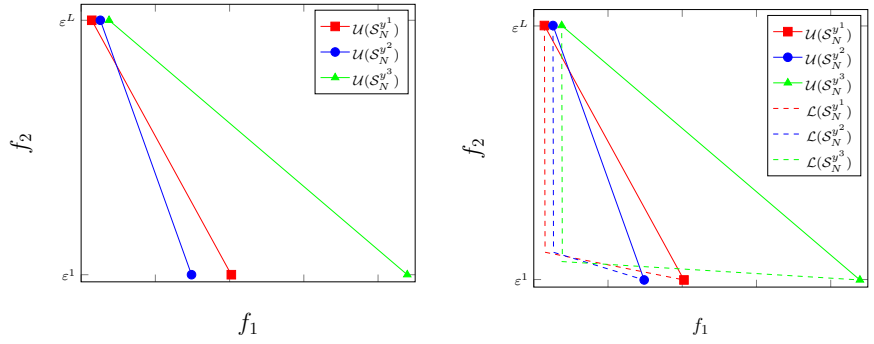
In the example we present next, it is useful to visualise our candidate (and bounded) sub-problems by viewing their bounds in objective space. To this end, we define a set Γ which contains *candidate points* denoted by $\gamma^{\varepsilon,y} \in \Gamma$, where $\gamma^{\varepsilon,y} = (f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y), \varepsilon)$. We construct one candidate point for each $y \in \mathcal{Y}$, $\varepsilon \in \mathcal{E}$ for which $f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y) < f_{\text{upper}}(\varepsilon, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$. We let $\Gamma_y = \{\gamma^{\varepsilon,y'} \in \Gamma : y' = y\}$, be the set of candidate points associated with each $y \in \mathcal{Y}$.

Algorithm 1: An algorithm to find a set of efficient solutions of the BOMINLP problem (1)

Input: $\mathcal{Y}$ (Set of feasible vectors y)
Input: L (Number of ε values to use in the ε-constraint method)
Output: $\mathcal{S}_N$ (Set of non-dominated points of (1))

begin
  foreach $y \in \mathcal{Y}$ **do**
    $\{x^{(1,2)}(y), x^{(2,1)}(y)\} \leftarrow$ `computeLexicographicSols`(y) ;
    $\mathcal{S}_N^y \leftarrow \{f(x^{(1,2)}(y), y), f(x^{(2,1)}(y), y)\}$;
  end
  $\varepsilon^1 = \min_{y \in \mathcal{Y}} f_2(x^{(2,1)}(y), y)$;
  $\varepsilon^L =$
    $\max_{y \in \mathcal{Y}} f_2(x^{(1,2)}(y), y)$, with $f(x^{(1,2)}(y), y)$ non-dominated within $\bigcup_{y \in \mathcal{Y}} \mathcal{S}_N^y$;
  $\mathcal{E} \leftarrow$ `compute`$\mathcal{E}\left(\varepsilon^1, \varepsilon^L, L\right)$;
  while $\exists\, \varepsilon \in \mathcal{E}, y \in \mathcal{Y} : f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y) < f_{\text{upper}}(\varepsilon, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$ **do**
    $\varepsilon \leftarrow$ `getBest`$\varepsilon(\mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$;
    repeat
      $y \leftarrow \arg\min_{y \in \mathcal{Y}} f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y)$;
      $(x^*, \lambda) \leftarrow$ `solve`$(SP(\varepsilon, y))$;
      $\mathcal{S}_N^y \leftarrow \mathcal{S}_N^y \cup f(x^*, y)$;
    until $\min_{y \in \mathcal{Y}} f_{\text{lower}}(\varepsilon, \mathcal{S}_N^y) \geq f_{\text{upper}}(\varepsilon, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$;
  end
  $\mathcal{S}_N \leftarrow \left(\bigcup_{y \in \mathcal{Y}} \mathcal{S}_N^y\right)_N$;
end

Example 3 *Consider again the BOMINLP instance given in Example 2. Figure 4(a) shows the lexicographically non-dominated points and the corresponding upper bound sets $\mathcal{U}(\mathcal{S}_N^y)$ for each $y \in \mathcal{Y}$. Because we have three sub-problems $SP(y^1)$, $SP(y^2)$ and $SP(y^3)$, six lexicographically non-dominated points are computed; these are shown as filled markers. In this example, these have been constrained to a user-specified area of interest, and thus the lexicographic solutions share common f_2 values. Figure 4(b) shows lower and upper bound sets for each $y \in \mathcal{Y}$. Figure 4(c) shows the area the non-dominated points of sub-problem $SP(y^1)$*



(a) Lexicographically non-dominated points of sub-problem $SP(y)$ for all $y \in \mathcal{Y}$.

(b) Lower bound sets (dashed) of sub-problem $SP(y)$ for all $y \in \mathcal{Y}$.

(c) Area delimited by lower and upper bound sets of sub-problem $SP(y^1)$.

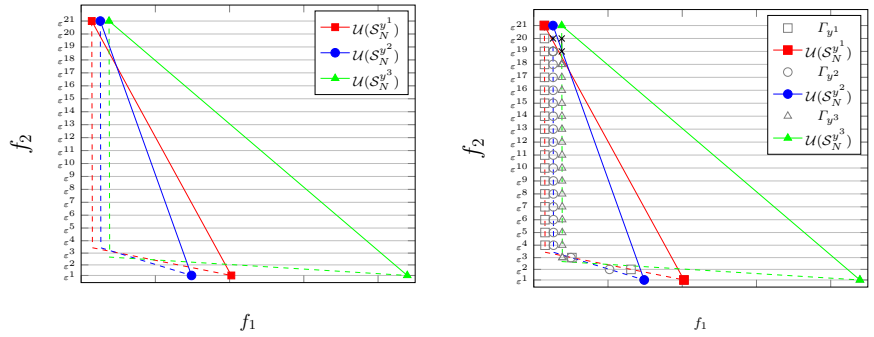
(d) All non-dominated points of BOMINLP problem must fall in the shaded area given the lexicographically non-dominated points computed for each $y \in \mathcal{Y}$.

Fig. 4: The lexicographically non-dominated points of $SP(y)$ in Example 2 are computed as well as corresponding bound sets $\mathcal{L}(\mathcal{S}_N^y)$ and $\mathcal{U}(\mathcal{S}_N^y)$ (dashed and solid lines respectively).

must fall in, which is limited by the lower and upper bound sets of $SP(y^1)$ *in Equations* (6) *and* (8)*, respectively. Figure 4(d) shows the area the non-dominated points of BOMINLP problem in Example 2 must fall in, which is limited by the joint lower and upper bound sets of BOMINLP in Equations* (10) *and* (11)*, respectively.*

After the lexicographically non-dominated points are computed, $\mathcal{E}$ *is defined. In Figure 5(a) one line parallel to the* f_1 *axis is drawn for each of the 21* ε *values in* $\mathcal{E}$*. Figure 5(b) shows the set of candidate sub-problems (represented by their candidate points, shown as unfilled markers). We note that three sub-problems* $SP(\varepsilon, y)$ *defined by* $(\varepsilon, y) \in \{(\varepsilon^{20}, y^2), (\varepsilon^{20}, y^3), (\varepsilon^{19}, y^3)\}$ *are bounded and thus are not candidates that need solving.*

We now choose some candidate sub-problem $SP(\varepsilon, y)$ *to solve next. Based on the procedure described above, we choose* $\varepsilon = \varepsilon^{11}$ *as this is the element evenly spaced between the two previously explored* ε *values, namely* ε^1 *and* ε^{21}*. Because* $f_{\text{lower}}(\varepsilon^{11}, \mathcal{S}_N^{y^1}) < f_{\text{lower}}(\varepsilon^{11}, \mathcal{S}_N^{y^2}) < f_{\text{lower}}(\varepsilon^{11}, \mathcal{S}_N^{y^3})$*, we solve the sub-problem*



(a) Lines parallel to the f_1 axis show the 21 ε values in $\mathcal{E}$ used to solve the ε-constraint problems $SP(\varepsilon, y)$.

(b) The candidate sub-problems, represented by their associated candidate points. Three sub-problems are bounded (indicated by "×").

Fig. 5: The set $\mathcal{E}$ is computed, and a set of candidate sub-problems is determined (shown by the sets of associated candidate points Γ_y). Three sub-problems are bounded.

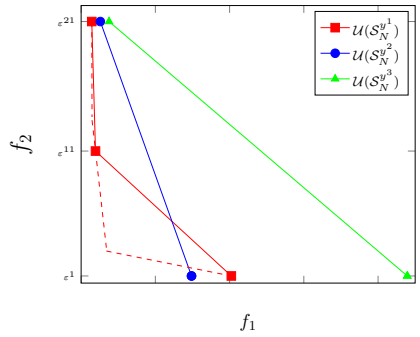
$SP(\varepsilon^{11}, y^1)$. The resulting non-dominated point is shown in Figure 6(a). This non-dominated point is added to $\mathcal{S}_N^{y^1}$. This gives new bound sets $\mathcal{U}(\mathcal{S}_N^{y^1})$ and $\mathcal{L}(\mathcal{S}_N^{y^1})$ (dashed lines in Figure 6(a)), new candidate points Γ_{y^1} (see Figure 6(b)) and, as a result of bounding, a reduced set of candidate points for y^2 and y^3 (compare Figure 6(c) with Figure 5(b)). We observe that $f_{\text{lower}}(\varepsilon^{11}, \mathcal{S}_N^y) \geq f_{\text{upper}}(\varepsilon^{11}, \mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$ for all $y \in \mathcal{Y}$, and so sub-problems $SP(\varepsilon^{11}, y^2)$ and $SP(\varepsilon^{11}, y^3)$ are now bounded and do not need to be solved.

As indicated in Algorithm 1, we now choose the next ε to be explored. Since ε^{11} is our current ε value, we need to find an explored ε^s value, with $s < 11$, such that all $\varepsilon \in \{\varepsilon^{s+1}, \ldots, \varepsilon^{10}\}$ remain unexplored. In this case, we found that $\varepsilon^s = \varepsilon^1$. Thus, the midpoint between ε^{11} and ε^1 is ε^6. The sub-problem with the smallest lower bound is y^2, and so $SP(\varepsilon^6, y^2)$ is solved. The non-dominated point obtained is added to $\mathcal{S}_N^{y^2}$ resulting in updated bounds and updated candidate points for y^2 (see Figure 7(a)).

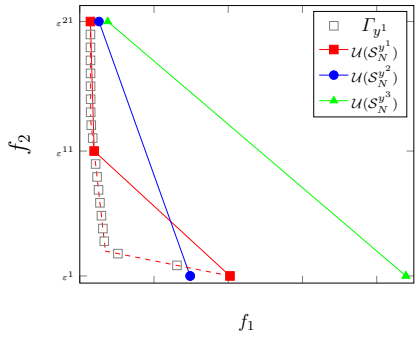
Unlike in the previous iteration, solving $SP(\varepsilon^6, y^2)$ did not result in the other sub-problems ($SP(\varepsilon^6, y^1)$ and $SP(\varepsilon^6, y^3)$) becoming bounded (see Figure 7(a)). We next solve $SP(\varepsilon^6, y^1)$ as it has the better lower bound. Solving this does not result in $SP(\varepsilon^6, y^3)$ becoming bounded, and so $SP(\varepsilon^6, y^3)$ is solved next (see Figure 7(b)).

The algorithm continues in this fashion by choosing a new ε to explore (in this case ε^{16}), solving the sub-problems, and so on, until all $SP(\varepsilon, y)$ sub-problems are bounded.

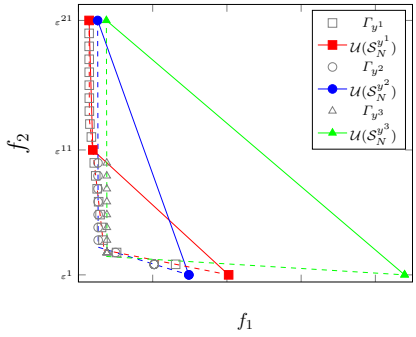
Figure 8(a) shows the set of non-dominated points computed by Algorithm 1 for Example 3. While the naïve method needs to compute 63 points to find a set of non-dominated points of (1), only 28 points were computed using our method. Both algorithms end up with $L = 21$ non-dominated points in the final set. Although we have not solved all possible sub-problems, the final set of non-



(a) A new non-dominated point is computed by solving $SP(\varepsilon^{11}, y^1)$, added to the set $\mathcal{S}_N^{y^1}$ and the new bound sets $\mathcal{U}(\mathcal{S}_N^{y^1})$ (solid line) and $\mathcal{L}(\mathcal{S}_N^{y^1})$ (dashed lines) are computed.



(b) This results in a new set of candidate points associated with y^1.



(c) The candidate point sets Γ_{y^2} and Γ_{y^3} have been reduced in size as a result of some candidate sub-problems becoming bounded.

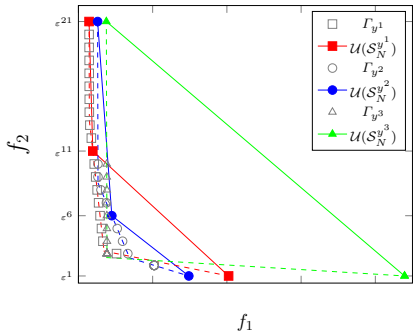
Fig. 6: A new non-dominated point is computed which results in new bound sets and candidate points.

dominated points shown in Figure 8(b) is identical to the set obtained by the naïve method as shown in Figure 3(b).

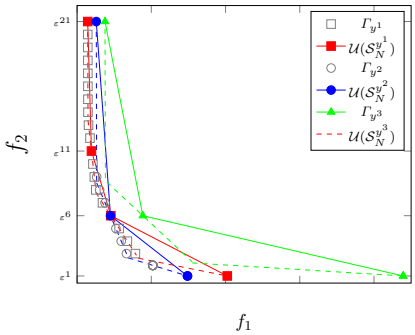
The next section shows an application of the procedure in Algorithm 1 on a practical problem arising in radiation therapy for cancer treatment.

5 Intensity Modulated Radiation Therapy: Application and Computational Results

Intensity modulated radiation therapy (IMRT) is a common technique in external radiation therapy. Its goal is to damage tumour cells by delivering ionising radiation from an external source to the tumour or *planning target volume* (PTV) without compromising surrounding normal tissue and organs at risk (OARs). Unfortunately, because of the physics of radiation delivery, there is a trade-off between



(a) We have chosen to explore ε^6 next. First, we solve $SP(\varepsilon^6, y^2)$ and update bound sets and candidate points for $SP(y^2)$. Then $SP(\varepsilon^6, y^1)$ is solved as is the candidate with the better lower bound. Again bound sets and candidate points for $SP(y^2)$ are updated.

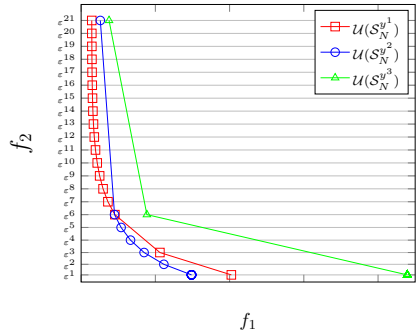


(b) Solving $SP(\varepsilon^6, y^1)$ did not result in $SP(\varepsilon^6, y^3)$ being bounded, and so $SP(\varepsilon^6, y^3)$ has now also been solved. We note that no candidate points remain for sub-problem $SP(y^3)$ as these have now all been bounded.

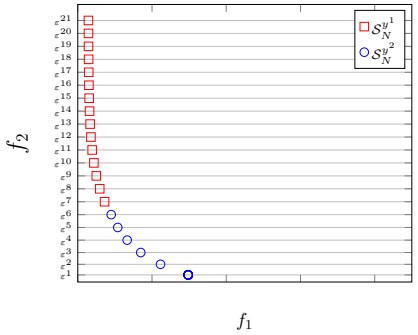
Fig. 7: Three non-dominated points are computed for ε^6 (one for each $y \in \mathcal{Y}$).

tumour control and protecting normal tissue and OARs. IMRT requires the generation of a *fluence map* which leads to a dose distribution that delivers radiation to the tumour while sparing surrounding OARs.

Due to the complexity of designing a treatment plan, the IMRT planning problem is usually divided into three sequential problems, namely, beam angle optimisation (BAO), fluence map optimisation (FMO) and multi-leaf collimator sequencing (Ehrgott et al., 2009). A solution of the BAO problem determines the directions of radiation beam angles, i.e. a beam angle configuration (BAC) that is selected from a discrete set of available beam angles. Then, the optimal fluence



(a) The set of 28 non-dominated points for sub-problems generated when using Algorithm 1.



(b) Final set of non-dominated points obtained by Algorithm 1. This set is identical to the set in Figure 3(b).

Fig. 8: Final set of generated points using the approach proposed in this study.

of radiation for each beam angle needs to be computed (FMO problem) to give a high quality dose distribution in the organs. The goal will be to find a fluence map that maximises the dose delivered to the tumour and, at the same time, minimises the dose deposited at each OAR.

In the context of Section 4 we are solving a multi-objective version of the BAO problem (MO-BAO) aiming to identify an efficient set of BACs, hence MO-BAO corresponds to BOMINLP problem (2) with $\mathcal{A} \in \hat{\mathcal{A}}$ representing a BAC and $\hat{\mathcal{A}} = \mathcal{Y}$ the set of all BACs. Each BAC $\mathcal{A}$ is evaluated by solving a multi-objective FMO problem (MO-FMO) to identify efficient fluence maps that achieve a required dose to the PTV while minimising the radiation dose received by the OARs. In the context of IMRT, problem (3) is *MO-FMO*($\mathcal{A}$) for BAC $\mathcal{A}$, and its ε-constraint scalarisation (4) is denoted *MO-FMO*($\mathcal{A}, \varepsilon$).

Section 5.1 outlines the details of an IMRT model for the interested reader. Otherwise, the reader can skip over Section 5.1 and continue with our computational experiments in Section 5.3, from page 23, instead.

5.1 Intensity Modulated Radiation Therapy Details

The single objective BAO problem is usually stated as a mixed integer (or binary) optimisation problem (Bangert et al., 2012; Ehrgott and Johnston, 2003; Preciado-Walters et al., 2004, 2006), as the set of all possible angles K is discrete and only a subset of them is considered in a BAC $\mathscr{A}$. Similar formulations have been used for its multi-objective counterpart (Breedveld et al., 2012; Schreibmann et al., 2004). In general, the MO-BAO problem is non-convex with possibly many local optima (Ehrgott and Johnston, 2003; Ehrgott et al., 2008; Lim and Cao, 2012; Pugachev et al., 2001). Furthermore, when MO-BAO and MO-FMO are posed together (i.e. when evaluating a BAC requires solving the MO-FMO problem), the set of feasible solutions of the MO-FMO is highly enlarged (Bortfeld and Schlegel, 1993; Pugachev and Xing, 2001). Thus, it is not possible to explore the entire set of feasible solutions within a reasonable computation time. Hence, only a subset of "promising" BACs is usually considered.

Cabrera G. et al. (2018) and Cabrera-Guerrero et al. (2018) propose a strategy to approximately solve the MO-BAO problem that consists of two phases. In the first phase a set of "promising" beam angle configurations is generated. In Cabrera G. et al. (2018) these promising BACs correspond to a set of locally optimal BACs found by a steepest descent algorithm that resulted to be not only locally optimal but also non-dominated. In Cabrera-Guerrero et al. (2018) the steepest descent algorithm in Cabrera G. et al. (2018) is extended to a Pareto local search strategy that finds a set of locally non-dominated BACs. Other strategies can also be used to find these promising BACs, though. Once the set of promising BACs has been found,a large set of non-dominated points for each sub-problem is generated and a dominance analysis is performed over the entire set of generated points to find the set of (approximately) non-dominated points of the main MO-BAO problem (second phase). As we pointed out in Section 1, such a strategy is quite inefficient as it computes a large number of points that will eventually be shown to be dominated. The method presented in this article can be used to improve the procedure used in the second phase in Cabrera G. et al. (2018) by reducing the number of non-dominated points that need to be generated.

In the context of the method proposed in this paper, the MO-BAO problem is the BOMINLP problem (1). In the MO-BAO problem we seek a set of efficient BACs $\mathcal{A} \subseteq \mathcal{P}^N(K)$, where $\mathcal{P}^N(K)$ is the set of all N-element subsets of K and K is the set of all possible beam angles around the patient, with $K = \{k\pi/180 : k = 0, 1, 2, \ldots, 359\}$. $\mathscr{A} \in \mathcal{P}^N(K)$ is a BAC and $N > 0$ the a priori determined number of angles. We denote the i-th angle of $\mathscr{A}$ by $\mathscr{A}_i$ for $i = 1, \ldots, N$. Moreover, for a given fixed BAC $\mathscr{A} \in \mathcal{P}^N(K)$ we obtain the MO-FMO problem for the given BAC as a sub-problem. The goal of the MO-FMO problem is to find a set of efficient fluence maps $\mathcal{X}_E^{\mathscr{A}} \subseteq \mathcal{X}(\mathscr{A})$ for the given BAC $\mathscr{A} \in \mathcal{P}^N(K)$, where $\mathcal{X}(\mathscr{A})$ is the set of feasible fluence maps for BAC $\mathscr{A}$ (being, simply, those fluence maps which have zero fluence for all beamlets associated with un-used angles). Efficient solutions $x \in \mathcal{X}_E^{\mathscr{A}}$ can be found using the ε-constraint method mentioned before.

In this paper we consider the generalised equivalent uniform dose (*gEUD*) as the basis of our objective functions. The *gEUD* can be defined as the biologically equivalent dose that, if delivered uniformly, would lead to the same response as the actual non-uniform dose distribution (Niemierko, 1997).

Before we introduce the mathematical expression of the *gEUD* and the objective functions we are going to use in this work, we need to present some general notation in IMRT.

Let $x \in \mathbb{R}^n$ denote a vector of beamlet fluences, where n corresponds to the total number of beamlets over all pre-determined beam angles and $x_i \geqq 0$ is the fluence at beamlet i. All mathematical optimisation models in IMRT are based on the radiation dose deposited into each voxel j of PTV and OARs by fluence map x. This dose distribution d is calculated using the expression

$$d_j^r(x) = \sum_{i=1}^{n} A_{ji}^r x_i \text{ for all } j = 1, 2, ..., m^r, \tag{15}$$

where $r \in R = \{O_1, \ldots, O_Q, T\}$ is an element in the index set of regions (we use the term region to denote either the tumour, any organ at risk, or normal tissue), with the tumour indexed by $r = T$ and the organs at risk and normal tissue indexed by $r = O_q$ with $q = 1, \ldots, Q$. m^r is the total number of voxels in region r, j corresponds to a specific voxel in region r, $d^r \in \mathbb{R}^{m^r}$ is a dose vector and its elements d_j^r give the total dose delivered to voxel j in region r by the fluence map $x \in \mathcal{X}(\mathscr{A})$. Here, the dose deposition matrix $A^r \in \mathbb{R}^{m^r \times n}$ is a given matrix where $A_{ji}^r \geqq 0$ defines the rate at which radiation dose along beamlet i is deposited into voxel j in region r.

The mathematical expression for *gEUD* is

$$gEUD^r(x) = \left(\frac{1}{m^r} \sum_{j=1}^{m^r} \left(d_j^r(x) \right)^{a^r} \right)^{1/a^r}, \tag{16}$$

where a^r is a region-dependent parameter and $d_j^r(x)$ is as given in Equation (15). For the tumour ($r = T$), we put $a^T < 0$, whereas for OARs ($r = O_q$) we choose $a^{O_q} > 1$.

Using *gEUD*-based objective functions we have that the associated MO-FMO problem is convex which allows us, on the one hand, to find efficient solutions to

the MO-FMO problem and, on the other hand, to apply the strategy proposed in Section 4 to reduce the number of non-dominated points computed to obtain a set of (approximately) non-dominated points of the MO-BAO problem.

The objective $gEUD^r(x)$ has several desirable properties for optimisation (Choi and Deasy, 2002; Romeijn et al., 2004). If $a^r \geqq 1$, $gEUD^r(x)$ is a norm and thus is a convex function. If, on the other hand, $a^r < 0$, then $gEUD^r(x)$ is no longer a norm, but is a concave function of x (Choi and Deasy, 2002). Moreover, $gEUD^r(x)$ is positively homogeneous for all a^r (Cabrera G. et al., 2014). As demonstrated in Cabrera G. et al. (2014), a positively homogeneous multi-objective optimisation problem with p objective functions can be converted into a multi-objective optimisation problem with $p-1$ objectives. This can be done by transforming one of the objectives into a constraint. Hence the MO-FMO problem can be solved with $p-1$ objectives and a constraint specifying that the p-th objective takes some specified value. (To maintain convexity, we implement this as an inequality constraint applied to the tumour's $gEUD$, and observe that this constraint will always be satisfied at equality in an optimal solution.) Having solved the problem for one particular value of the p-th objective, all other efficient solutions can be obtained by scaling (Cabrera G. et al., 2014).

The MO-BAO we address in this paper is as follows:

$$\begin{aligned} \text{MO-BAO:} \quad \min_{\mathscr{A}} \left(\min_{x} \begin{pmatrix} gEUD^{O_1}(x) \\ gEUD^{O_2}(x) \end{pmatrix} \right) \\ \text{s.t.} \qquad gEUD^T(x) \geqq t \\ x \in \mathcal{X}(\mathscr{A}) \\ \mathscr{A} \in \hat{\mathcal{A}}, \end{aligned} \tag{17}$$

the solution of which is a set containing all efficient BACs which use exactly N of $|K|$ available angles, where t is equal to the prescribed $gEUD$ of the tumour and $\hat{\mathcal{A}} = \mathcal{P}^N(K)$.

5.2 Comparison with ε-constraint scalarization for the full MO-BAO problem

Solving the MO-BAO problem in (17), that is, finding all efficient BACs which use exactly N angles, is extremely difficult. This is, on the one hand, because of the large number of possible BACs that exist and, on the other hand, because of the computation time needed to produce non-dominated points for each of those BACs.

We demonstrate this by solving a single ε-constraint scalarisation of (17) with a reduced number of available beam angles. In Table 1 we report the computation time required to solve a single MINLP ε-constraint scalarisation instance for the 'CERR' problem instance (see Section 5.3) as the number of available beam angles $|K|$ increases, i.e. each row gives the computation time for a single scalarised instance of (17) that selects the best $N = 5$ beam angles from $|K| \in \{12, 15, 18, 20\}$ available beam angles. To create this scalarisation, we set the value of ε for the gEUD of the bladder constraint to 37 and minimised the gEUD to the rectum. The prescribed tumour gEUD was set to $t = 70$ for this and all further numerical tests. It is important to note that we use the same computer and solver to solve

Table 1: Computation time needed to find a single efficient 5-angle BAC and treatment plan with a small number $|K|$ of available beam angles by solving a single scalarisation instance of MO-BAO (17).

Available Angles	Angle Step size	CPU time (s)
$\|K\| = 12$	30°	5,069
$\|K\| = 15$	24°	19,715
$\|K\| = 18$	20°	56,450
$\|K\| = 20$	18°	105,131

these single ε-constraint scalarisations and the experiments presented in Section 5.3.

The last column of Table 1 shows that the computation time needed to solve a single ε-constraint scalarisation of (17) increases quickly as the number of available beam angles becomes larger. Clearly, the time needed for all $|K| = 360$ integer beam angles would be enormous. Since the time listed in Table 1 is the time to find a single non-dominated point, the time required to find a set of non-dominated points would be many times the computation time reported in Table 1. It is clear that this is not practical from a clinical point of view. In Section 5.3 we consider instances with $|K| = 72$ available beam angles and 80 different values of ε. Solving this MO-BAO problem exactly using its ε-constraint scalarisation with currently available mathematical programming solvers and computers is simply not possible, and so efficient alternative approaches such as ours are required.

The MO-BAO problem (17) is an instance of BOMINLP (2). The corresponding sub-problem (3) under our decomposition approach is the MO-FMO problem:

$$\begin{aligned} \text{MO-FMO}(\mathscr{A}): \quad \min_{x} \ & \begin{pmatrix} z_1(x) \\ z_2(x) \end{pmatrix} = \begin{pmatrix} gEUD^{O_1}(x) \\ gEUD^{O_2}(x) \end{pmatrix} \\ \text{s.t.} \ & gEUD^{T}(x) \geqq t \\ & x \in \mathcal{X}(\mathscr{A}). \end{aligned} \tag{18}$$

The solution to (18) is a set $\mathcal{X}_E^{\mathscr{A}}$ of efficient fluence maps for BAC $\mathscr{A}$. As before, we find this set using the ε-constraint method, which requires solving single objective constrained optimisation problems of the following form:

$$\begin{aligned} \text{MO-FMO}(\varepsilon, \mathscr{A}): \min_{x} \ & gEUD^{O_1}(x) \\ \text{s.t.} \ & gEUD^{T}(x) \geqq t \\ & x \in \mathcal{X}(\mathscr{A}) \\ & gEUD^{O_2}(x) \leqq \varepsilon. \end{aligned} \tag{19}$$

As we discussed above, we require Langrangian dual values to exist at optimal solutions to this MO-FMO($\varepsilon, \mathscr{A}$) sub-problem. For practical problem instances, this sub-problem satisfies the well-known linear independence constraint qualification (LICQ) (see A), and so these Lagrangian values can be found using standard optimisation solvers such as the Ipopt solver (Wächter and Biegler, 2006) we use to run all our experiments.

In the next section, we use Algorithm 1 to solve our MO-BAO problem (17), where we observe that the sub-problem (3) is given by our MO-FMO sub-problem

(18), and the sub-problem (4) required for the ε-constraint method is given by (19).

5.3 Computational Experiments

In this section we apply Algorithm 1 to the MO-BAO problem (17) for 5 prostate cases obtained from Clinica Alemana de Santiago (Chile) for anonymous former patients (referred to as 'TRT001' to 'TRT005') and also for a prostate case extracted from Deasy et al. (2003) (referred to as 'CERR'). We perform all the experiments presented in this paper on an Intel i7 processor with 32GB of memory running on Ubuntu 18.

This case study considers three regions, namely, PTV, indexed by T, and the rectum and the bladder as OARs O_{rectum} and $O_{bladder}$. The total number of voxels is about 56,000 and the number of beamlets n over the five beams is around 320 depending on the considered BAC.

The single objective problem $MO\text{-}FMO(\mathscr{A}, \varepsilon)$ is solved exactly using the *IPOPT* solver from COIN-OR (Wächter and Biegler, 2006). IPOPT implements a primal-dual interior point method that uses first and second derivative information to find locally optimal solutions. To find $\mathcal{E}$, we first compute lexicographically optimal solutions for all $\mathscr{A} \in \hat{\mathcal{A}}$. We bound the objective functions such that maximum dose ($gEUD$) values for both the rectum and the bladder are not larger than 65. The obtained range of objective values for $O_{bladder}$ is $[\varepsilon^1, \varepsilon^L] = [31.735, 65]$. After computing $\varepsilon^1, \varepsilon^L$, we compute set $\mathcal{E}$, which consists of 80 equally spaced ε values within the range $[\varepsilon^1, \varepsilon^{80}]$.

As mentioned in Section 5.1, the MO-BAO problem is too difficult to solve exactly within a reasonable computation ´time frame, and so we replace the set of all BACs $\hat{\mathcal{A}}$ in (18), corresponding to $\mathcal{Y}$ in (17), by a set of 15 promising BACs consisting of $N = 5$ beam angles each. These promising BACs were obtained using the two-phase approach presented in Cabrera G. et al. (2018).

Table 2 shows a summary of the results obtained for each instance. The number of generated points (column 'Num Points Generated') is significantly smaller than the $50 \times 80 = 1200$ points the naïve approach (see Section 2) needs to obtain the same set of non-dominated points for the problem. Column 'Fraction of Naïve Approach' shows, for each instance, the fraction of points we require when compared with the naïve approach, which is on average only 12.42%, i.e., our approach is about ten times faster than the naïve approach.

One measure of inefficiency in our algorithm is the number of points that are generated but then discarded because they are dominated. Assuming no points are discarded, then the minimum number of points any approach needs to compute to produce the same set of non-dominated points is 109 (2 lexicographic points for each of the 15 BACs plus one point for each of the 80 ε values, less one point because the $\varepsilon = 65$ value coincides with a lexicographic point). The column 'Num Points Discarded' shows the extra points that were generated beyond this minimum and then discarded as a result of being dominated. We observe that we are discarding, on average, about 26.6% of the points we generate. Finally, the computation of each non-dominated point takes about 80 seconds on average. Thus, to produce the entire finite set of non-dominated points for $|K| = 72$ available beam angles our approach will take no more than 3.5 hours, which is much shorter than

Table 2: Results obtained for all 6 instances tested in this study.

Instance	Num Points Generated	Fraction of Naïve Approach	Num Points Discarded	Fraction Discarded
CERR	146	12.2%	37	25.3%
TRT001	148	12.3%	39	26.4%
TRT002	134	11.2%	25	18.7%
TRT003	156	13.0%	47	30.1%
TRT004	148	12.3%	39	26.4%
TRT005	162	13.5%	53	32.7%
Average	**149**	**12.4%**	**39**	**26.6%**

the 29 hours the solver takes to obtain only one non-dominated point using the ε-constraint scalarisation when considering $|K| = 20$ available beam angles (see Table 1).

Figure 9 shows the results obtained for the CERR instance. The final set of non-dominated points consists of the 80 points shown in Figure 10. Points in this set belong to only 3 of the 15 BACs ($\mathscr{A}^1, \mathscr{A}^2, \mathscr{A}^3$ in Figure 10).

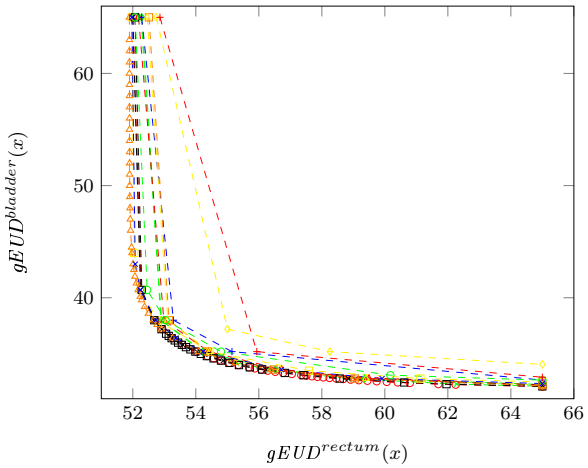
6 Conclusions

In this study a method to efficiently search for a finite set of non-dominated points of bi-objective mixed binary non-linear optimisation problems (BOMINLP), where sub-problems obtained when fixing integer variables are convex, is proposed. We take advantage of the convexity property of the sub-problem to iteratively discard points that will result in dominated points of the bi-objective mixed binary optimisation problem. Using our method, the vast majority of the dominated points from the sub-problem are discarded without any quality impairment on the final set of non-dominated points of the main problem.

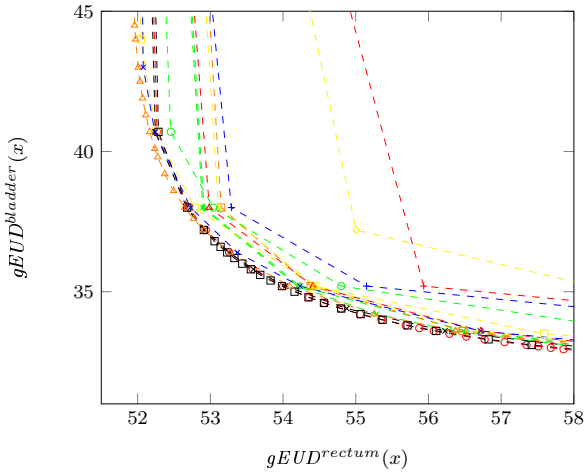
Our experiments show that our matheuristic can generate 80 approximately non-dominated points for our test problem in 3.5 hours, while the traditional epsilon-constraint method applied to the largest problem we could model (with just 20 of the 72 angles considered by the matheuristic) took 29 hours to find just one approximately non-dominated point. It is also 10 times faster than a naïve enumeration implementation. Further improvements may be possible, for example through better choices in our $\texttt{getBest}\varepsilon(\mathcal{S}_N^{y^1}, \mathcal{S}_N^{y^2}, \ldots, \mathcal{S}_N^{y^{|\mathcal{Y}|}})$ function.

We applied the proposed method to a problem arising in radiation therapy optimisation, namely the multi-objective beam angle optimisation problem and obtained very good results. To further strengthen the support for the quality of our algorithm it will be worth trying our proposed approach on other BOMINLP problems such as the bi-objective joint facility location and the network design problem addressed in Carrano et al. (2007). Further, extending our algorithm to MO problems with more than 2 objectives can also be explored.

To make further progress in algorithms for solving BOMINLP problems, it will be important to develop methods that can identify start and end points of sections of the non-dominated set of the BOMINLP generated by specific sub-problems $SP(y)$, i.e. by addressing point *(i)* on page 7. Such algorithms can then benefit from progress in exact methods in bi-objective convex optimisation or use the algorithm proposed here as a sub-routine.



(a) The non-dominated points generated while using Algorithm 1.



(b) Figure 9(a) zoomed in.

Fig. 9: Set of non-dominated points of sub-problem (MO-FMO) computed by Algorithm 1 for each $\mathscr{A} \in \hat{\mathcal{A}}$.

Acknowledgments

The authors would like to thank FONDECYT REDI 170036 for partially supporting this research. We would also like to thank Clínica Alemana de Santiago for sharing anonymous data from their former patients. G. Cabrera-Guerrero would like to thank FONDECYT Iniciación 11170456 for partially supporting this research.

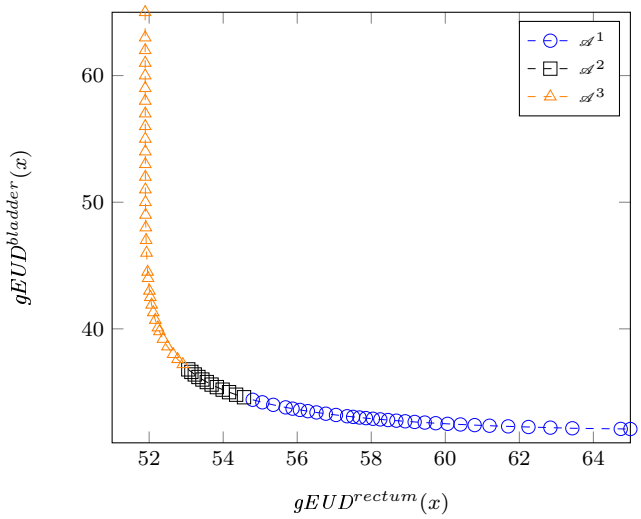


Fig. 10: Final set of (approximately) non-dominated points of the main MO-BAO problem obtained using the proposed method.

A Lagrangian Multiplier Existence

Assume we have some optimal solution x^* to MO-FMO$(\varepsilon, \mathscr{A})$ for some $\varepsilon, \mathscr{A}$. We wish to show that optimal dual variables (i.e. Lagrangian multipliers) exist for MO-FMO$(\varepsilon, \mathscr{A})$ at x^*. To show this, we show that solution x^* satisfies the well known linear independence constraint qualification (LICQ), i.e. that the gradients of the active inequality constraints are linearly independent at the optimal solution x^* to MO-FMO$(\varepsilon, \mathscr{A})$. Note that the constraints $x \in \mathcal{X}(\mathscr{A})$ simply indicate that all fluence variables belonging to beam angles not in $\mathscr{A}$ are zero, and therefore are not present in the problem; see Section 5. Therefore (19) effectively has only two $gEUD$ constraints.

If only one of the two constraints is active, then the result follows immediately. Therefore, we focus on the case where both $gEUD$ inequality constraints are satisfied at equality.

Consider some organ $r \in \{T, O_2\}$, and some beamlet k with intensity x_k. The kth component in $\nabla gEUD^r(x)$ is given by

$$\frac{\partial}{\partial x_k} gEUD^r(x) = \frac{1}{m^r} \left[\sum_{j=1}^{m^r} A^r_{jk} \left(\sum_{i=1}^{n} A^r_{ji} x_i \right)^{a^r - 1} \right] \left[\frac{1}{m^r} \sum_{j=1}^{m^r} \left(\sum_{i}^{n} A^r_{ji} x_i \right)^{a^r} \right]^{\frac{1}{a^r} - 1} \tag{20}$$

If $\nabla gEUD^T(x^*)$ and $\nabla gEUD^{O_2}(x^*)$ are not linearly independent, then there exists some $h > 0$ for which $\frac{\partial}{\partial x_k} gEUD^T(x^*) = h \cdot \frac{\partial}{\partial x_k} gEUD^{O_2}(x^*)$ for all $k = 1, 2, \ldots, n$. Our experiments show that each optimal solution x^* to MO-FMO$(\varepsilon, \mathscr{A})$ has many dozens of non-zero x^*_i values, giving many dozens of non-zero values in $\nabla gEUD^T(x^*)$ and $\nabla gEUD^{O_2}(x^*)$. Furthermore, recall that $A^r_{ji} \geqq 0$ is a real-valued constant describing the rate at which radiation dose along beamlet i is deposited into voxel j in region r, and thus each non-zero A^r_{ji} value is typically unique. Given that we also have $a^T \neq a^{O_2}$, we conclude that for any practical problem, no such h will exist, and so $\nabla gEUD^T(x)$ and $\nabla gEUD^{O_2}(x)$ will be independent, showing that the desired result will hold in practice.

References

Bangert M, Ziegenhein P, Oelfke U (2012) Characterizing the combinatorial beam angle selection problem. Physics in Medicine and Biology 57(20):6707–6723

Belotti P, Soylu B, Wiecek M (2013) A branch-and-bound algorithm for biobjective mixed-integer programs. Tech. rep., Optimization Online

Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. Numerische Mathematik 4(1):238–252, DOI 10.1007/BF01386316, URL https://doi.org/10.1007/BF01386316

Boland N, Charkhgard H, Savelsbergh M (2014) The triangle splitting method for biobjective mixed integer programming. In: Lee J, Vygen J (eds) IPCO 2014, Springer International Publishing, Cham, pp 162–173

Boland N, Charkhgard H, Savelsbergh M (2015) A criterion space search algorithm for biobjective mixed integer programming: The triangle splitting method. INFORMS Journal on Computing 27(4):597–618

Bortfeld T, Schlegel W (1993) Optimization of beam orientations in radiation therapy: some theoretical considerations. Physics in Medicine and Biology 38(2):291–304

Breedveld S, Storchi P, Voet P, Heijmen B (2012) iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar imrt plans. Medical Physics 39(1):951–963

Cabrera G G, Ehrgott M, Mason A, Philpott A (2014) Multi-objective optimisation of positively homogeneous functions and an application in radiation therapy. Operations Research Letters 42(4):268–272

Cabrera G G, Ehrgott M, Mason AJ, Raith A (2018) A matheuristic approach to solve the multiobjective beam angle optimization problem in intensity-modulated radiation therapy. International Transactions in Operational Research 25(1):243–268

Cabrera-Guerrero G, Mason A, Raith A, Ehrgott M (2018) Pareto local search algorithms for the multi-objective beam angle optimisation problem. Journal of Heuristics 24(2):205–238

Cacchiani V, D'Ambrosio C (2017) A branch-and-bound based heuristic algorithm for convex multi-objective MINLPs. European Journal of Operational Research 260(3):920–933

Carrano EG, Takahashi RHC, Fonseca CM, Neto OM (2007) Bi-objective combined facility location and network design. In: Obayashi S, Deb K, Poloni C, Hiroyasu T, Murata T (eds) Evolutionary Multi-Criterion Optimization 2007, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 486–500

Chankong V, Haimes Y (1983) Multiobjective Decision Making: Theory and Methodology. Elsevier Science Publishing Co., New York

Choi B, Deasy J (2002) The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning. Physics in Medicine and Biology 47(20):3579–3589

Deasy J, Blanco A, Clark V (2003) CERR: A computational environment for radiotherapy research. Medical Physics 30(5):979–985

Ehrgott M (2005) Multicriteria Optimization (2. ed.). Springer

Ehrgott M, Gandibleux X (2007) Bound sets for biobjective combinatorial optimization problems. Computers & Operations Research 34(9):2674 – 2694

Ehrgott M, Johnston R (2003) Optimisation of beam directions in intensity modulated radiation therapy planning. OR Spectrum 25(2):251–264

Ehrgott M, Holder A, Reese J (2008) Beam selection in radiotherapy design. Linear Algebra and its Applications 428(5):1272–1312

Ehrgott M, Güler C, Hamacher HW, Shao L (2009) Mathematical optimization in intensity modulated radiation therapy. Annals of Operations Research 175(1):309–365

Ehrgott M, Gandibleux X, Przybylski A (2016) Exact methods for multi-objective combinatorial optimisation. In: Greco S, Ehrgott M, Figueira RJ (eds) Multiple Criteria Decision Analysis: State of the Art Surveys, International Series in Operations Research & Management Science, Springer New York, New York, NY, chap 19, pp 817–850

Fattahi A, Turkay M (2018) A one direction search method to find the exact nondominated frontier of biobjective mixed-binary linear programming problems. European Journal of Operational Research 266:415–425

Fischetti M, Fischetti M (2016) Matheuristics. In: Martí R, Panos P, Resende M (eds) Handbook of Heuristics, vol 1, Springer, pp 710–713

Haimes Y, Chankong V (1979) Kuhn-Tucker multipliers as trade-offs in multiobjective decision-making analysis. Analytica 15:59–72

Jayasekara PL, Adelgren N, Wiecek MM (2019) On convex multiobjective programs with application to portfolio optimization. Journal of Multi-Criteria Decision Analysis Accepted for Publication

Lim G, Cao W (2012) A two-phase method for selecting IMRT treatment beam angles: Branch-and-prune and local neighborhood search. European Journal of Operational Research 217(3):609–618

Mavrotas G, Diakoulaki D (1998) A branch and bound algorithm for mixed zero-one multiple objective linear programming. European Journal of Operational Research 107(3):530–541

Mavrotas G, Diakoulaki D (2005) Multi-criteria branch and bound: A vector maximization algorithm for mixed 0-1 multiple objective linear programming. Applied Mathematics and Computation 171(1):53–71

Miettinen K (1999) Nonlinear Multiobjective Optimization, International Series in Operations Research and Management Science, vol 12. Kluwer Academic Publishers, Dordrecht

Niemierko A (1997) Reporting and analyzing dose distributions: A concept of equivalent uniform dose. Medical Physics 24(1):103–110

Özceylan E, Paksoy T (2014) Interactive fuzzy programming approaches to the strategic and tactical planning of a closed-loop supply chain under uncertainty. International Journal of Production Research 52(8):2363–2387

Preciado-Walters F, Rardin RL, Langer M, Thai V (2004) A coupled column generation, mixed integer approach to optimal planning of intensity modulated radiation therapy for cancer. Mathematical Programming 101(2):319–338

Preciado-Walters F, Langer M, Rardin R, Thai V (2006) Column generation for IMRT cancer therapy optimization with implementable segments. Annals of Operations Research 148(1):65–79

Pugachev A, Xing L (2001) Pseudo beams-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy. International Journal of Radiation Oncology Biology Physics 51(5):1361–1370

Pugachev A, Li JG, Boyer A, Hancock S, Le Q, Donaldson S, Xing L (2001) Role of beam orientation optimization in intensity-modulated radiation therapy. International Journal of Radiation Oncology Biology Physics 50(2):551–560

Rezaei J, Davoodi M (2011) Multi-objective models for lot-sizing with supplier selection. International Journal of Production Economics 130(1):77–86

Romeijn H, Dempsey J, Li J (2004) A unifying framework for multi-criteria fluence map optimization models. Physics in Medicine and Biology 49(10):1991–2013

de Santis M, Eichfelder G, Niebling J, Rocktäschel S (2019) Solving multiobjective mixed integer convex optimization problems. Tech. rep., Technische Universität Ilmenau

Schreibmann E, Lahanas M, Xing L, Baltas D (2004) Multiobjective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy. Physics in Medicine and Biology 49(5):747–770

Shirazi H, Kia R, Javadian N, Tavakkoli-Moghaddam R (2014) An archived multi-objective simulated annealing for a dynamic cellular manufacturing system. Journal of Industrial Engineering International 10(2):1–17

Soylu B (2018) The search-and-remove algorithm for biobjective mixed-integer linear programming problems. European Journal of Operational Research 268:281–299

Stidsen T, Andersen K, Dammann B (2014) A branch and bound algorithm for a class of biobjective mixed integer programs. Management Science 60(4):1009–1032

Teghem J J, Kunsch P (1986) Interactive methods for multi-objective integer linear programming. In: Fandel G, Grauer M, Kurzhanski A, Wierzbicki A (eds) Large-Scale Modelling and Interactive Decision Analysis, Lecture Notes in Economics and Mathematical Systems, vol 273, Springer Berlin Heidelberg, pp 75–87

Vincent T, Seipp F, Ruzika S, Przybylski A, Gandibleux X (2013) Multiple objective branch and bound for mixed 0-1 linear programming: Corrections and improvements for the biobjective case. Computers & Operations Research 40(1):498–509

Wächter A, Biegler LT (2006) On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. Mathematical Programming 106(1):25–57

Wiecek MM, Ehrgott M, Engau A (2016) Continuous multiobjective programming. In: Greco S, Ehrgott M, Figueira JR (eds) Multicriteria Decision Analysis: State of the Art Surveys, International Series in Operations Research & Management Science, vol 233, Springer, New York, chap 18, pp 739–815