

Panos M. Pardalos · Pando G. Georgiev  
Petraq Papajorgji · Britta Neugaard  
*Editors*

# Systems Analysis Tools for Better Health Care Delivery

# Springer Optimization and Its Applications

---

VOLUME 74

---

## *Managing Editor*

Panos M. Pardalos (University of Florida)

## *Editor–Combinatorial Optimization*

Ding-Zhu Du (University of Texas at Dallas)

## *Advisory Board*

J. Birge (University of Chicago)

C.A. Floudas (Princeton University)

F. Giannessi (University of Pisa)

H.D. Sherali (Virginia Polytechnic and State University)

T. Terlaky (Lehigh University)

Y. Ye (Stanford University)

## *Aims and Scope*

Optimization has been expanding in all directions at an astonishing rate during the last few decades. New algorithmic and theoretical techniques have been developed, the diffusion into other disciplines has proceeded at a rapid pace, and our knowledge of all aspects of the field has grown even more profound. At the same time, one of the most striking trends in optimization is the constantly increasing emphasis on the interdisciplinary nature of the field. Optimization has been a basic tool in all areas of applied mathematics, engineering, medicine, economics, and other sciences.

The series *Springer Optimization and Its Applications* publishes undergraduate and graduate textbooks, monographs and state-of-the-art expository work that focus on algorithms for solving optimization problems and also study applications involving such problems. Some of the topics covered include nonlinear optimization (convex and nonconvex), network flow problems, stochastic optimization, optimal control, discrete optimization, multiobjective programming, description of software packages, approximation techniques and heuristic approaches.

For further volumes:

[www.springer.com/series/7393](http://www.springer.com/series/7393)



Panos M. Pardalos • Pando G. Georgiev  
Petraq Papajorgji • Britta Neugaard  
Editors

# Systems Analysis Tools for Better Health Care Delivery

 Springer

*Editors*

Panos M. Pardalos  
Department of Industrial and Systems  
Engineering  
University of Florida  
Gainesville, FL, USA

Petraq Papajorgji  
Ministry of Education and Sciences  
National Agency for Exams  
Tirana, Albania

Pando G. Georgiev  
Center for Applied Optimization  
Department of Industrial and Systems  
Engineering  
University of Florida  
Gainesville, FL, USA

Britta Neugaard  
James A. Haley VA Medical Center  
Tampa, FL, USA

ISSN 1931-6828

ISBN 978-1-4614-5093-1

ISBN 978-1-4614-5094-8 (eBook)

DOI 10.1007/978-1-4614-5094-8

Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012953471

Mathematics Subject Classification (2010): 11T71, 14G50, 34L25, 34A30, 34A34, 37L30, 37N40, 41A05, 41A20, 41A21, 42A16, 51M04, 51N20, 60G25, 60J10, 60K15, 65C20, 65C50, 65D18, 65S05, 65T60, 68P25, 68P30, 68U20, 74J20, 78A46, 81U40, 90B10, 90B50, 90C11, 90C29, 90C90, 94A05, 94A15, 94A60, 97R30, 97R60, 97U70

© Springer Science+Business Media New York 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

This book presents some recent systems engineering and mathematical tools for health care along with their real-world applications by health care practitioners and engineers. Advanced approaches, tools, and algorithms used in operating room scheduling and patient flow are covered. State-of-the-art results from applications of data mining, business process modeling, and simulation in health care, together with optimization methods, form the core of the book. It illustrates the increased need of partnership between engineers and health care professionals.

In what follows, we present a brief outline of the contributed papers in this volume, which are collected in an alphabetical order of the contributors.

In Chap. 1, Dionne M. Aleman, Hamid R. Ghaffari, Velibor V. Mišić, Michael B. Sharpe, Mark Ruschin, and David A. Jaffray present a semi-infinite linear programming approach to solve high-resolution, convex quadratic optimization treatment problems in a reasonable amount of time. They also devise several computational improvements to the commonly used projected gradient algorithm that provide significant time savings when optimizations must be performed iteratively. Their approaches allow previously unwieldy treatment planning problems to be solved in a clinically viable amount of time.

In Chap. 2, Nebil Buyurgan and Nabil Lehlou present a study on the analysis of portable asset management strategies in hospitals. The problem addressed here is the unavailability of the portable assets when they are needed due to lost or hoarding, which lead to significant amount of staff time for search and underutilization of the assets. A simulation-based decision support tool is constructed to analyze the different processes and the impact of Radio Frequency Identification (RFID) technology on the widely adopted portable asset management models. The results suggest that the substantial gain could be realized by implementing RFID systems.

In Chap. 3, Camilo Mancilla and Robert H. Storer presents several data mining tools that can be used to investigate health outcomes, and provides a sample analysis of health care data to demonstrate their use. The tools include market basket analysis, text analysis, and predictive modeling. These tools can investigate also cancer treatments. The need to analyze real data is particularly necessary with the increased prominence of comparative effectiveness analysis.

In Chap. 4, Anastasius Moutzoglou and Anastasia Kastania review different ways that stochastic integer programming has been used to improve efficiency and efficacy in health care delivery. For the purpose of this study health care delivery is divided in two areas: resource allocation and operations. In each area the stochastic components are identified and the algorithms and solution techniques that have been proposed in the literature are described. Current challenges and open questions are stated.

In Chap. 5, Neng Fan, Syed Mujahid, Jicong Zhang, Pando Georgiev, Petraq Papajorgji, Ingrida Steponavice, Britta Neugaard, and Panos M. Pardalos present a survey on e-health management. Changes in health care delivery have become so widespread and numerous that the idea of e-health has become one of excitement and prediction rather than intervention. On the other hand, the endorsement of e-health is spreading slowly. Few companies focus on population-oriented e-health tools partly because of perceptions about the viability and capacity of the market. Moreover, developers of e-health resources are a highly diverse group with differing skills and resources while a common problem for developers is finding the balance between risk and outcome. On the other hand, e-health presents risks to patient health information that involve not only appropriate protocols but also laws, regulations, and appropriate safety culture. Breaches of network security and international viruses have elevated the public awareness of online information and computer security, although the overwhelming majority of security breaches do not directly involve health-related data. Finally, as we believe in the implications of the genetic components of disease, we expect a significant increase in the genetic information of clinical records. The future vision is mobile personalized e-health in a patient-centered and patient-safety context.

In Chap. 6, Patricia Cerrito. We use a binary integer programming model to formulate and solve a nurse scheduling problem (NSP) which maximally satisfies nurse preferences. In a practical application of a VA hospital, besides considering the scheduling of two types of nurses (registered nurses and licensed practical nurses), two other types of employees (nursing assistants and health care techs), one nurse manager and a clinical nurse leader are also included in our model. Most of these employees are working full-time. In addition, we distinguish the schedule of weekdays and weekends with different requirements and different preferences for employees. Besides the requirements for each shift, we consider requirements for specific employees in some shifts in practical situations. The seven shifts do not necessarily have the same length in our model. Vacation time of employees is also considered in our model. Thus, the requirements for nurse scheduling are complicated and the objective is to maximize the satisfaction of preferred schedules of all these employees, including both nurses and other staffs. The presented model is complex, but efficiently solvable, satisfying the set of requirements in a particular application in a VA hospital.

In Chap. 7, Jennifer A. Pazour and Russell D. Meller study the pharmaceutical supply chain from a pharmaceutical distributor to a patient. The authors make

comparisons between a traditional distribution center and a hospital pharmacy and discuss the technologies used in both facilities, with special emphasis on the order-fulfillment process. The authors review analytical models for order-fulfillment technologies prevalent in pharmaceutical distribution, including models for A-Frame systems, carousel systems, picking machines, unit-dose repackaging technologies, and automated dispensing cabinets. Finally, the authors provide conclusions and future research directions.

In Chap. 8, Elina Rönnberg, Torbjörn Larsson, and Ann Bertilsson describe automatic scheduling of nurses. The intention of this chapter is to provide a piece of practical experience that can help bridge the gap between advanced method development and the use of automatic nurse scheduling in practice. The approach described here is to take account of a real-life problem with all its details, and to use a straightforward meta-heuristic in order to deliver automatically generated schedules. The contribution of this chapter is based on the result of two case studies, which will provide insights into real-world examples, including evaluation and feedback from the wards.

We wish to express our deepest appreciation to the above-named authors who contributed their papers for publication in this volume. In addition, we are also very thankful to Springer Publishing Company for their generous support for this publication.

Gainesville, FL, USA  
Gainesville, FL, USA  
Tirana, Albania  
Tampa, FL, USA

Panos M. Pardalos  
Pando G. Georgiev  
Petraq Papajorgji  
Britta Neugaard





# Contents

<b>1</b>	<b>Optimization Methods for Large-Scale Radiotherapy Problems</b> .....	<b>1</b>
	Dionne M. Aleman, Hamid R. Ghaffari, Velibor V. Mišić, Michael B. Sharpe, Mark Ruschin, and David A. Jaffray	
<b>2</b>	<b>Portable Asset Management in Hospitals</b> .....	<b>21</b>
	Nebil Buyurgan and Nabil Lehlou	
<b>3</b>	<b>Stochastic Integer Programming in Healthcare Delivery</b> .....	<b>37</b>
	Camilo Mancilla and Robert H. Storer	
<b>4</b>	<b>An Expository Discourse of E-Health</b> .....	<b>49</b>
	Anastasius Moumtzoglou and Anastasia Kastania	
<b>5</b>	<b>Nurse Scheduling Problem: An Integer Programming Model with a Practical Application</b> .....	<b>65</b>
	Neng Fan, Syed Mujahid, Jicong Zhang, Pando Georgiev, Petraq Papajorgji, Ingrida Steponavice, Britta Neugaard, and Panos M. Pardalos	
<b>6</b>	<b>Clinical Data Mining to Discover Optimal Treatment Patterns</b> .....	<b>99</b>
	Patricia Cerrito	
<b>7</b>	<b>Exploring the Parallels Between a Hospital Pharmacy and a Distribution Center</b> .....	<b>131</b>
	Jennifer A. Pazour and Russell D. Meller	
<b>8</b>	<b>Automatic Scheduling of Nurses: What Does It Take in Practice?</b> .....	<b>151</b>
	Elina Rönnerberg, Torbjörn Larsson, and Ann Bertilsson	



# Optimization Methods for Large-Scale Radiotherapy Problems

Dionne M. Aleman, Hamid R. Ghaffari, Velibor V. Mišić,  
Michael B. Sharpe, Mark Ruschin, and David A. Jaffray

## 1 Introduction

Mathematical models have been widely applied to the problem of designing highly customized radiotherapy treatment plans [1], but these previous approaches have almost exclusively focused on relatively moderate-sized treatments. The treatments previously studied include site-specific (e.g., head-and-neck, breast, prostate) *intensity modulated radiation therapy* (IMRT)—a treatment modality that allows for each beam in the treatment to have a unique distribution of radiation in order to deliver highly accurate dose—and Gamma Knife® treatments for small targets in the brain. For large-scale treatments, such as *total body irradiation* or very high-resolution Gamma Knife® Perfexion™ treatments, the optimization methods previously employed are no longer viable.

---

The work of this author was supported in part by The Canada Foundation for Innovation. The work of this author was supported in part by the Natural Sciences and Engineering Research Council Undergraduate Summer Research Award.

D.M. Aleman (✉) • H.R. Ghaffari • V.V. Mišić  
Department of Mechanical and Industrial Engineering, University of Toronto,  
5 King's College Road, Toronto, ON Canada M5S 3G8  
e-mail: aleman@mie.utoronto.ca; ghaffari@mie.utoronto.ca; velibor.misic@utoronto.ca

M.B. Sharpe • D.A. Jaffray  
Department of Radiation Oncology, Princess Margaret Hospital, University of Toronto,  
610 University of Avenue, Toronto, ON Canada M5G 2M9  
e-mail: michael.sharpe@rmp.uhn.on.ca; david.jaffray@rmp.uhn.on.ca

M. Ruschin  
Department of Radiation Oncology, Odette Cancer Centre, University of Toronto,  
2075 Bayview Ave, Toronto, ON Canada M4N 3M5  
e-mail: mark.ruschin@sunnybrook.ca

We investigate optimization methods to address the computational difficulties present in large-scale radiation therapy treatment optimization. We specifically focus on improving total body irradiation using IMRT and applying IMRT optimization mathematical models to Gamma Knife® Perfexion™. We consider the radiotherapy optimization process to consist of two sub-problems, a common approach in IMRT literature (e.g., [2–6]). These problems must be solved sequentially rather than simultaneously in large-scale treatment problems due to the massive data requirements in storing the effects of radiation delivery configurations on the patient’s tissues. The first problem is to determine the relative positions of the radiation beams with respect to the patient’s body. Once those beam orientations are obtained, the general radiotherapy optimization model is formulated as

$$\begin{aligned} & \text{minimize} && \phi(\mathbf{z}) && \text{(RT - OPT)} \\ & \text{subject to} && \mathbf{z} = d(\mathbf{x}) \\ & && \mathbf{x} \geq \mathbf{0} \end{aligned}$$

where  $\mathbf{x}$  represents the radiation intensity of the variables within our control;  $d(\mathbf{x})$  is a function relating radiation intensities  $\mathbf{x}$  to delivered dose  $\mathbf{z}$ ; and  $\phi(\mathbf{z})$  is a quantitative measure of treatment plan quality, where smaller values correspond to better treatments.

Define  $z_{js}$  as the dose delivered to voxel  $j$  in structure  $s \in S$ , which has  $v_s$  voxels. A voxel is a cube used to discretize the patient’s body. In our approach,  $\phi(\mathbf{z})$  is comprised of convex quadratic penalties  $F_s(z_{js})$  that weight the over- and underdosage of each voxel  $j$  in structure  $s \in S$ . Letting  $(\cdot)_+$  represent  $\max\{\cdot, 0\}$ , the penalty function for a single voxel is

$$F_s(z_{js}) = \frac{1}{v_s} \left[ w_s \left( \underline{T}_s - z_{js} \right)_+^2 + \bar{w}_s \left( z_{js} - \bar{T}_s \right)_+^2 \right]$$

where  $w_s$  is the weight assigned to penalize any dose received under  $\underline{T}_s$ , and  $\bar{w}_s$  is the weight assigned to penalize any dose received over  $\bar{T}_s$ . In previous approaches using penalty-based objectives (e.g., [2, 3, 6]), the thresholds at which over- and underdosing are the same. Our formulation allows for unique values at which to penalize overdose and underdose, and therefore yields increased flexibility by providing for “sweet spots” of radiation for structures at which no penalty is assigned. Graphically, the penalty can be represented as shown in Fig. 1 for a given structure, where T-u and T-o indicate the under- and overdosage thresholds, respectively. The RT-OPT objective is then to minimize the total penalty:

$$\phi(\mathbf{z}) = \sum_{s \in S} \sum_{j=1}^{v_s} F_s(z_{js}).$$

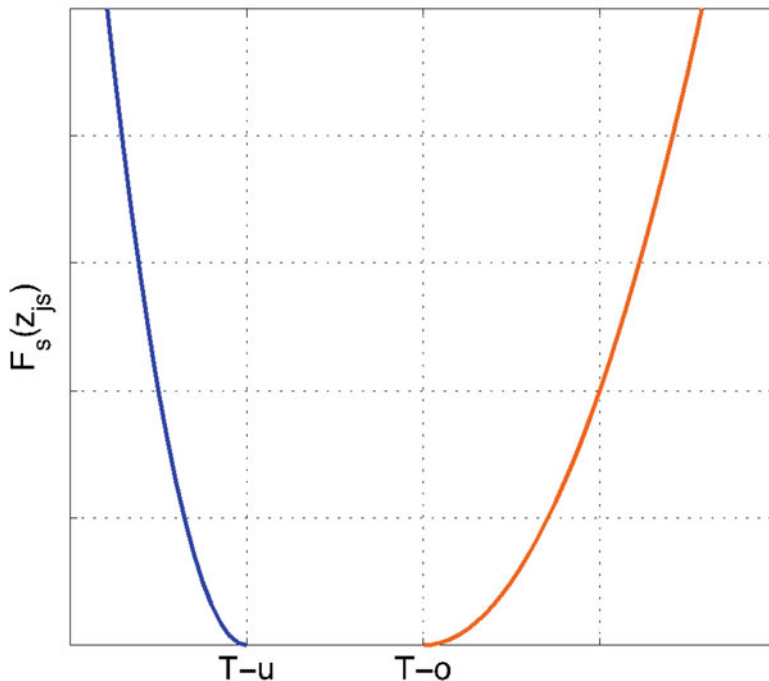


Fig. 1 Convex quadratic dose penalty with un-penalized sweet spot

## 2 Total Marrow Irradiation

Prior to receiving a bone marrow transplant, the patient's existing bone marrow must be eradicated in order to allow the donated stem cell transplant to successfully integrate with the body. One method of eliminating the bone marrow is through total body irradiation. Failure to destroy all the bone marrow will result in a transplant failure, so it is crucial to deliver enough radiation that the bone marrow is sufficiently eradicated.

The necessity of delivering high dose levels comes with more frequent and more severe toxic effects in healthy tissue [7], which can lead to a lower chance of a successful transplant. Further complications to the patient's health are caused by the method of radiation delivery. Because such a large area must be treated, clinicians typically place the patient far away from the isocenter—the central focus point of the beams of radiation—so that the area covered by each beam is large. The uncertainty in delivered dose results in the need to deliver high levels of radiation to ensure the bone marrow receives appropriate dose. It also prevents the use of conformal treatments that target just the bone marrow while sparing healthy organs.

*Total marrow irradiation* (TMI), a treatment that irradiates only the bone marrow while avoiding healthy tissues, can be achieved using IMRT. There has been limited research into the application of IMRT to TMI, mainly due to the computational issues present in designing complex treatments for such a large area. Using standard commercial planning systems, large reductions in dose to some organs can potentially be achieved [8, 9]. However, important organs such as the spinal cord are not considered. TMI has been considered using tomotherapy, and similarly shows that the dose organs can be significantly reduced [10, 11]. In contrast to these studies, we consider the TMI problem within the mathematical framework RT-OPT that we have successfully applied to TMI [12, 13]; we also consider non-coplanar beams (beams obtained from the movement of more than one linear accelerator component), which are necessary to deliver radiation to the patient at a standard 100 cm isocenter distance [14] so that analytical dose approximations are valid.

The two major subproblems in TMI are *beam orientation optimization* (BOO) and *fluence map optimization* (FMO). BOO determines the optimal beams from which to deliver radiation. Once these beams are obtained, FMO determines the optimal distribution of radiation for each beam. The distribution of radiation in each beam is delivered by considering each beam as being comprised of many smaller beamlets, each of which can deliver a radiation dose, called a *fluence*, independent of the other beamlets. The fluences for a set of beams are called a *fluence map*.

## 2.1 *Beam Orientation Optimization*

Beam orientation optimization has been well studied in the literature using a variety of approaches from genetic and evolutionary algorithms (e.g., [15–18]) to simulated annealing (e.g., [19–24]) to beam’s-eye-view techniques (e.g., [21, 22, 25–29]) and more. Despite the large amount of research done in BOO, only a relatively small number of studies (e.g., [2, 3, 16, 18, 30–32]) have used the optimal fluence maps resulting from a set of beams to inform the selection beam orientations to use in the treatment plan. This simplification is largely due to the computational difficulties associated with having the FMO problem as the objective function of the BOO problem. These difficulties are highlighted by the fact that mixed integer approaches to combined BOO and FMO can only be performed if the beam solution space is restricted to a very small candidate set [33, 34].

The computational difficulties in addressing BOO and FMO simultaneously are even more evident in TMI, where the patient is on the order of 10 times larger than in the previously studied site-specific treatments, and 10 times more beams with 100 times more beamlets per beam are required to deliver an accurate treatment. Further, the patient size requires the use of non-coplanar beams, which increases the beam solution space to the point where little previous research has been able to consider non-coplanar beams [3, 21, 23, 30, 35–37]. Of these, all but [3] considered only a handful of non-coplanar beams.

Because it is widely accepted that the optimal solution to the FMO problem presents the most relevant measure of a beam set's quality [17–20, 23, 24, 33, 34, 36–46], and because it is essential to deliver a high quality treatment plan in TMI (more so than in traditional site-specific treatments), we seek to combine BOO and FMO. However, the TMI treatment planning problem is more difficult to solve due to the large patient size and beam solution space, so an algorithm that can move quickly through the solution space is desirable.

Thus, in order to solve the BOO problem, we employ the Add/Drop neighborhood search approach developed by [2] for coplanar beam selection. We extend the Add/Drop method to address the non-coplanar beam space necessary for TMI as described in [13]. In the Add/Drop method, each beam is analyzed in turn and replaced with an improving neighboring beam. Our enhancement redefines a beam's neighborhood as not simply a collection of all beams within a certain proximity, but only as nearby beams obtained from moving a single linear accelerator component. Therefore, each beam has multiple neighborhoods. The neighborhood examined in an iteration depends on historical improvements of that beam-component pair and probabilistic expectations of treatment plan improvement. Further details about the non-coplanar Add/Drop method are provided in [13].

Because each iteration of the Add/Drop method requires enumeration of the FMO solutions for each neighbor in the selected neighborhood, the computation time of the algorithm is dependent on the speed of the FMO optimization. Thus, we focus our efforts on improving the speed of the FMO optimization.

## 2.2 *Fluence Map Optimization*

As previously stated, IMRT optimization for TMI is more difficult than site-specific treatments due to the size of the patient. While a typical head-and-neck treatment contains  $\approx 80,000$  voxels, a TMI treatment contains  $\approx 760,000$  voxels. Empirically, our work has indicated that 30 beams are necessary to deliver a clinically acceptable TMI treatment, and each beam typically has  $\approx 3,000$  active beamlets. This results in 90,000 beamlet intensities that must be optimized in the fluence map optimization. The number of decision variables makes the use of Hessian-based algorithms, such as interior point methods that have been shown to yield optimal high-quality IMRT treatments [47], prohibitively expensive in terms of time and potentially numerically unstable.

We therefore apply a standard projected gradient algorithm with an Armijo line search [48, 49] to solve the FMO problem. Although projected gradient methods cannot guarantee an optimal solution, such methods are known to be fast and empirically return good solutions. Because the FMO formulation given in RT-OPT is a convex quadratic with only nonnegativity constraints, the local optimum approached by the projected method is the globally optimal solution. In practice, projected gradient methods have been shown to return quality treatment plans in



IMRT optimization [2, 3], even applied to TMI [12, 13]. However, while these solutions are obtained very quickly for site-specific treatments [47], about 45 min is required to return a solution for TMI.

Because the FMO must be performed for each neighbor in the Add/Drop algorithm used to select beam orientations, lengthy FMO computations significantly impact the solution space that can be searched in the fixed 12-h treatment planning limit imposed by clinicians. To speed up the projected gradient algorithm, we examine alternate line search techniques and a warm start approach.

### 2.2.1 Alternate Line Search Techniques

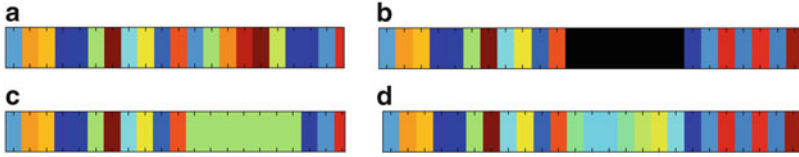
The projected gradient algorithm starts from a near-zero solution (in this case, a solution with all beamlets delivering almost zero intensity) and then picks a direction to move to obtain the maximum improvement in the objective function. This direction is the gradient of the objective function; specifically, since our FMO is a minimization problem, the direction is the negative gradient. Once this descent direction is obtained, the algorithm moves from the current point some distance in this direction. The determination of this distance is called a line search because the algorithm must search the descent direction for a distance to move, called a step length, so that the new solution yields at least some minimum amount of improvement in the objective function.

Most commonly, the majority of computation time spent in projected gradient methods is searching the descent direction for an appropriate distance to move the current solution. We explore three methods of performing the line search to determine which technique most improves the computation time of the algorithm.

First, we employ a traditional backtracking method [48]. In backtracking, we select some initial step length  $\Delta$ , and if the solution located at that point provides insufficient objective function improvement, we decrease the step length by a factor  $\delta$  and examine the resulting solution. This process is repeated until an appropriate step length is found.

After examining the step lengths taken by the traditional backtracking line search, we observed that for our specific optimization model, the final step lengths taken in the first few iterations were far larger than in most subsequent iterations. We therefore designed a modified backtracking line search where the initial step length is drastically reduced after a fixed number of iterations.

We also observed that the final step lengths taken in most iterations were generally very small. Yet, the backtracking algorithm was started with a fairly large value of  $\Delta$ . So, the third line search method we attempted was a simple forward line search operating in the same manner as the backtracking line search, with the exception that we start with a very small step length and increase the step length by a factor of  $\delta$  until a solution with sufficient objective function improvement is found.



**Fig. 2** Warm start illustration. (a) Initial solution. (b) Old beam replaced with new beamlets. (c) Average fluence initialization. (d) Least-squares fluence initialization

### 2.2.2 Warm Start Approaches

Although projected gradient methods start at near-zero solutions, the iterative process taken by the Add/Drop method supplies useful information about fluences of each beam in each iteration. Rather than start at a near-zero solution, we can use this information to start a near-optimal solution, which will significantly reduce the number of iterations needed to converge to a solution.

From one iteration to the next, only one beam changes in the Add/Drop method, meaning that only a subset of beamlet variables are new. Figure 2a illustrates beamlets values for some solution, which then become the values shown in Fig. 2b when one set of beamlets is replaced by new variables. The beam that changes is moved to a neighboring beam, which is near to the original beam. Therefore, it is likely that the fluences of the other beams in the solutions will be similar in the new solution to their values in the previous solutions.

Rather than discard this intuitive knowledge about the beamlet values from one iteration to the next, we use the previous optimal beamlet values as a warm start to the projected gradient algorithm. The beamlets from the old beam are replaced by new beamlets from the new beam, but, because these beamlets are new, we devise two methods of assigning their values in the initial projected gradient solution.

First, we initialize each of the new beamlets to have the average value of the beamlets of the old beam (Fig. 2c). This averaging is a computationally inexpensive method of exploiting the fact that the old beam and new beam will likely have similar overall intensity. Second, we initialize the values of the new beamlets to values that will most closely approximate the dose delivered by the original beamlets (Fig. 2d). This approximation is done using a least squares optimization.

If the new beam is the  $k$ th beam in the solution ( $\theta_k$ ) and the set of beamlets in this beam is  $B_{\theta_k}$  with intensities  $\tilde{x}_i$ , then the least squares optimization is given by

$$\begin{aligned}
 & \text{minimize} && \sum_{s \in S} \sum_{j=1}^{v_s} \left( \tilde{z}_{js}^{(k)} - z_{js}^{(k)} \right)^2 && \text{(LSQ)} \\
 & \text{subject to} && \tilde{z}_{js}^{(k)} = \sum_{i \in B_{\theta_k}} D_{ijs} \tilde{x}_i && s \in S, j = 1, \dots, v_s \\
 & && \tilde{x}_i \geq 0 && i \in B_{\theta_k}
 \end{aligned}$$

where  $\tilde{z}_{js}^{(k)}$  is the dose delivered by the new beamlets to voxel  $j$  in structure  $s$  and  $z_{js}^{(k)}$  is the dose delivered by the previous beamlets to voxel  $j$  in structure  $s$ . Although using a least squares optimization to initialize the new fluence values is potentially computationally intensive, it is possible that the reduction in the number of iterations performed by the projected gradient algorithm will still result in a faster solution.

### 2.3 Results

The algorithms were tested on a single TMI patient from The Princess Margaret Hospital (Toronto, ON, Canada). Each type of projected gradient method was executed on a Dell Intel Core 2 Duo with a 2.4 GHz CPU and 8 GB of RAM. Because the Add/Drop algorithm returns a locally optimal solution, the quality of the solution may be affected by the starting point. For robustness, we test the algorithms using 10 different randomly generated 30-beam solutions. The initial fluences of all of the beams were set to 0.3 Gy. Each variant of the projected gradient method terminated when successive iterations resulted in a relative objective function improvement of less than 0.01. For the reduced step projected gradient implementation, the step length was reduced after three iterations. For both the backtracking and reduced step implementations, the initial step length was 50 and  $\delta=0.25$ . For the forward line search implementation, the initial step length was 3 and  $\delta=10$ .

Table 1 illustrates the performance of each of the line search methods. In terms of the number of iterations and computation time, the reduced step variant performed the best, while the forward line search performed the worst. The poor computational speed of the forward line search results from the fact that, despite the small step size in a majority of iterations, the line searches with large step sizes required more computation time than what was gained by the small step sizes.

Somewhat surprisingly, the reduced step method resulted in the worst objective function values when used in the Add/Drop algorithm. One possibility for poor solution quality is that from our observations of the backtracking method, after the

**Table 1** Comparison of the line search methods

		Backtracking	Reduced step	Forward
Add/drop iterations	Mean	16	13.7	16.1
	St. dev.	2.94	6.52	5.47
	Minimum	12	7	8
	Maximum	20	23	23
Time (min)	Mean	43.74	30.94	47.85
	St. dev.	9.04	15.87	16.41
	Minimum	31.70	13.40	22.70
	Maximum	57.30	51.70	68.10
FMO obj.fn. value	Mean	13,298.29	17,458.25	14,695.36
	St. dev.	2,062.89	4,657.02	2,750.92
	Minimum	11,822.90	12,230.20	11,724.50
	Maximum	18,492.00	23,452.30	19,184.00