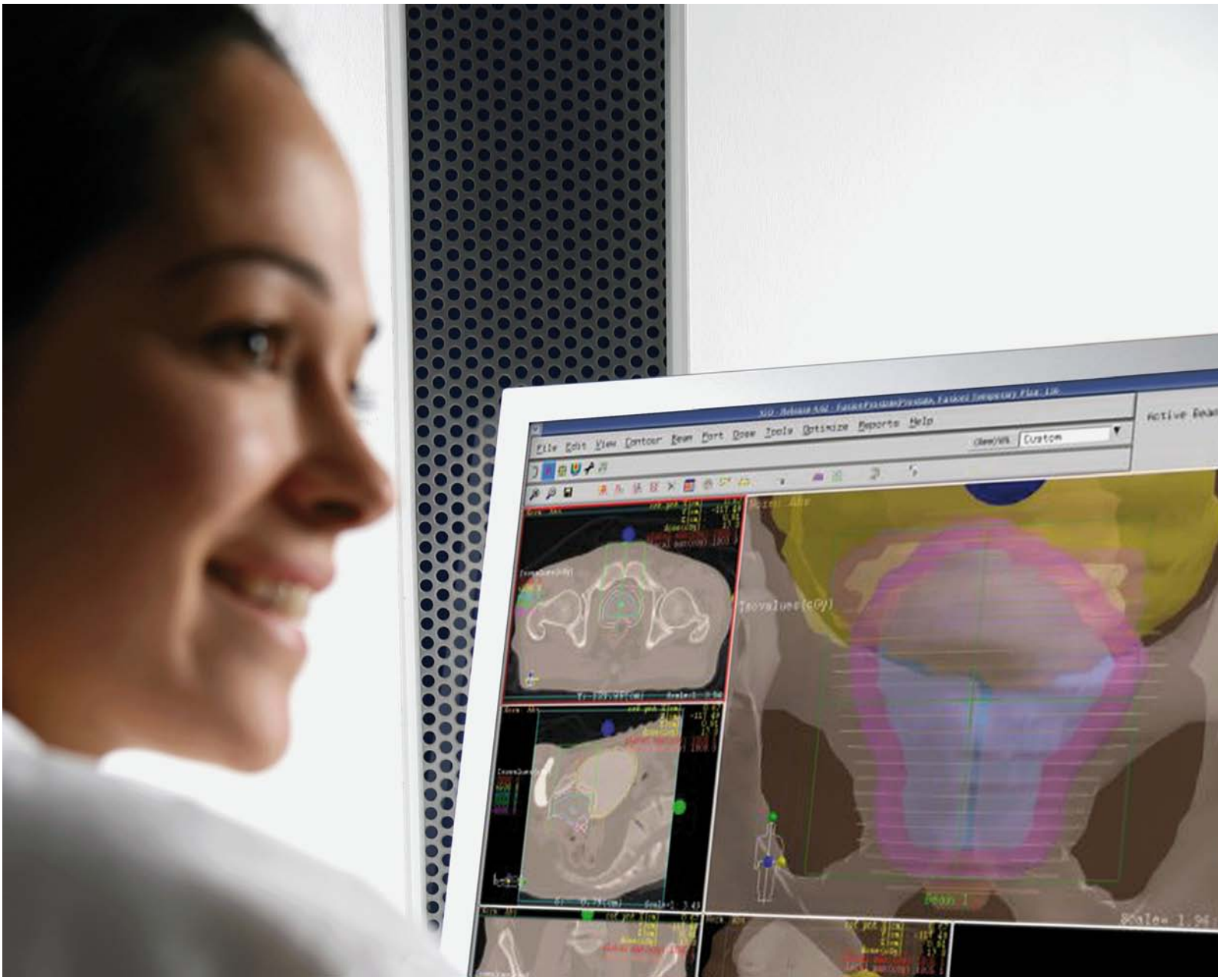
XiO®

Intensity Modulated Radiation Therapy (IMRT) Technical Reference



Document ID: LRMXIO0006
Language: English

ELEKTA

XiO® Intensity Modulated Radiation Therapy (IMRT) Technical Reference

Version 1.00

Copyright statement

Use of trademarks and trade names statement

Acknowledgement of other trademarks

Referenced documents

Elekta does not supply all documents that we refer to in this document with the equipment. Elekta reserves the right to make the decision on which of the documents it supplies with the equipment.

Contact information

GLOBAL ELEKTA SOFTWARE SUPPORT

Online Support: www.elekta.com/supportplus
Contact information (phone, email, webex): www.elekta.com/software-support
Training Calendar: www.elekta.com/training

ELEKTA REGIONAL OFFICES

13723 Riverport Drive, Suite 100
Maryland Heights, MO 63043
Sales
Toll Free: +1 800 878 4276
Fax: +1 314 993 0075
Email: sales-na@elekta.com
Customer Support
Phone: +1 800 878 (4276)
Email: support@elekta.com

Sunnyvale, California
100 Mathilda Place, Fifth Floor
Sunnyvale, CA 94086
Phone: +1 408 830 8000
Fax: +1 408 830 8003
Email: info@elekta.com

Europe, Latin America, Africa and Middle East Contact Information
Elekta, Inc.
Linac House
Fleming Way, Crawley, West Sussex RH10 9RR
Phone: +44 1293 544 422
Fax: +44 1293 654 321

CE 0086

Table of Contents

1. Overview

Inverse planning for intensity modulated radiation therapy (IMRT) is a process by which the user of a treatment planning system interacts with the system through the use of an optimization algorithm and a cost function to produce a radiation treatment plan that optimizes radiation dose to a patient. In XiO IMRT, which uses a conjugate gradient optimizer, the cost function is composed of the sum of objective functions. Each objective function, or simply "objective", is an anatomy-specific function that establishes dose goals (*e.g.,* the target should get at least 60 Gy and no more than 66 Gy and the cord should get no more than 45 Gy) or dose-volume goals (*e.g.,* no more than 40% of the liver should get more than 50 Gy). The result of inverse planning and IMRT is often a treatment plan that delivers dose to targets and spares normal tissues in a way superior to using conventional conformal radiation treatment methods.

It must be noted that inverse planning often appears to produce astounding dosimetric results with little effort, skill, or forethought. Product demonstrations that support this impression are common. The demonstrator arranges five evenly spaced coplanar beams around an isocentric treatment point, specifies objectives for treatment from a well-known clinical protocol, and presses a "start" button to begin the dose optimization process. A few seconds later, ideal beamlet intensity maps for all beams are produced. Upon inspection, the ideal doses look good in both dose-volume histograms and isodose displays. Ideal beamlet intensity maps are then segmented for delivery by multileaf collimator (MLC) or converted to a form deliverable by compensators. Due to the potentially limited resolution of delivery devices, the final deliverable dose estimates may be somewhat degraded from the ideal doses. But, typically, the final dose distribution is acceptable.

Inverse planning is not quite this easy in practice, but neither does it have to be much more complicated! XiO IMRT was designed to make inverse planning relatively simple. The key, from an algorithmic standpoint, is its *effective* objective functions, which, among other improvements, lessen the role of importance weights in inverse planning. If you have used inverse planning software before and been stymied by the proliferation of sometimes unexplainable importance weights, you will be pleased at what XiO IMRT has to offer – importance weights that relate directly to treatment goals and are easy to understand. From a software design standpoint, XiO IMRT offers the same great visualization and dose analysis tools that users have come to expect from all IMPAC products.

By reading this technical reference, you will familiarize yourself with important general concepts in inverse planning and specific features and limitations of XiO IMRT. In this document, exact numerical values and numerical limits are rarely mentioned, in favor of general concepts. Information about implementation-specific parameters and the general use of XiO IMRT can be found in the XiO Online Help, the "IMRT Training Guide", "QA of IMRT Beams – Technical Considerations", and the "Beam Modeling Guide." In addition, you may want to read Ezell *et al,* 2003, whose purpose is "to guide and assist the clinical medical physicist in developing and implementing a viable and safe IMRT program. This report, while not prescribing specific procedures, provides the framework and guidance to allow clinical radiation oncology physicists to make judicious decisions in implementing a safe and efficient IMRT program in their clinics."

This training guide is valid for XiO IMRT Release 4.2 and higher. Do not rely on this technical reference for information regarding earlier XiO IMRT implementations.

2. Optimization – Basic Concepts

This chapter is an introduction to optimization concepts and optimization algorithms, also known as "optimizers." The discussion is facilitated through the use of a simple patient composed of two volume elements, or "voxels."

The Two Voxel Patient

Figure 2-1 shows that the patient has a voxel for a target (T) and a voxel for normal anatomy (N), also called an organ at risk (OAR). Let D_T represent the dose to the target and D_N represent the dose to the OAR from the four "beamlets" aimed at the patient. In this case, a beamlet, which is normally one of many subdivisions of a beam, is equivalent to a beam, but in practice this is never the case. Each beam is normally subdivided into many beamlets that project from the machine source to a plane defined at the treatment isocenter. For example, a 10 cm x 10 cm beam with beamlets that project to 5 mm x 5 mm at isocenter would have 400 beamlets. Notice also in the figure that no beamlets intersect the OAR without also intersecting the target. It is typical for an inverse planning system to select beamlets that intersect targets whether or not they intersect OARs. It is the job of the optimizer to find the best combination of beamlet weights that can deliver tumorcidal dose to targets while limiting dose to OARs.

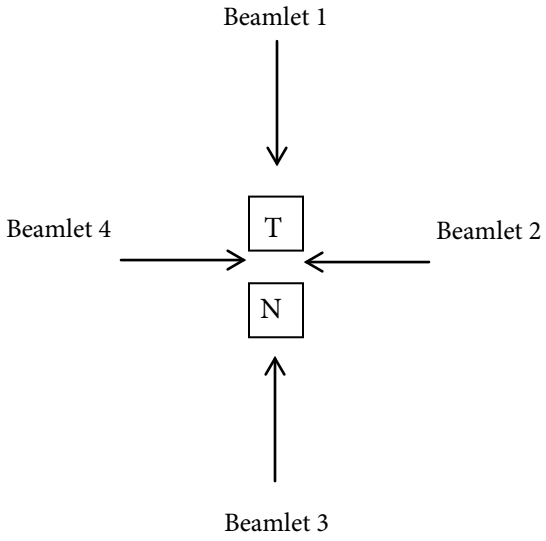


Figure 2-1: Two voxel patient with four beamlets

Cost Functions and Objectives

As mentioned in the overview, cost functions are composed of the sum of one or more objectives.

$$F = \sum_{\sigma=1}^{n} O_\sigma \qquad \textbf{(2-1)}$$

The above notation means that the cost function is the sum of a number of objectives, indexed over σ. For example, if there were three objectives

$$F = \sum_{\sigma=1}^{3} O_\sigma = O_1 + O_2 + O_3 + ADS \qquad (2\text{-}2)$$

Objectives for the two voxel patient, given the above notation, could be expressed as:

$$O_T(D_T) = (D_T - D_{GT})^2 \quad \textbf{- objective for the target} \qquad \textbf{(2-3)}$$

$$O_N(D_N) = (D_N - D_{GN})^2 \quad \textbf{- objective for the OAR} \qquad \textbf{(2-4)}$$

where D_{GT} is a goal dose for the target and D_{GN} is a goal dose for the OAR for some prescription. Notice, not surprisingly, that the objectives are a function of dose.

Assuming smoothing is turned off, the total cost function, or simply "cost function", for the patient could be expressed as:

$$F(D_T, D_N) = O_T(D_T) + O_N(D_N) = (D_T - D_{GT})^2 + (D_N - D_{GN})^2 \qquad (2\text{-}5)$$

From Equation 2-4, it is clear that if either the target or OAR voxel dose were to approach its goal dose, the evaluation of its respective objective would approach zero[1].

Normally, given this type of objective formulation, the goal dose for the target would be some value in the middle of its acceptable range and the goal dose for the OAR would be zero. Assuming that D_{GT} is 64 Gy and D_{GN} is 0 Gy, and smoothing is not being applied, the cost function becomes

$$F(D_T, D_N) = (D_T - 64)^2 + (D_N - 0)^2 = (D_T - 64)^2 + {D_N}^2 \qquad (2\text{-}6)$$

See IMRT_TUT_Sec13.ppt for an overview of the smoothing function.

Cost Function Minimization

Cost functions used in optimization are normally constructed so that the closer they are to achieving their goals, the lower the evaluation of the cost function. For this reason, optimization is also called function minimization or just minimization. Cost functions that maximize exist, but they tend to be more complicated and more unstable than cost functions that minimize. For example, the cost function could be reformulated as:

---

1 As we will see later, these are not very good cost function objectives, although this type of objective is used in a number of academic and commercial IMRT systems. XiO uses alternate "effective" cost functions [Wiesmeyer, 2004].

$$F(D_T, D_N) = \frac{1}{(D_T - 64)^2 + {D_N}^2} \qquad (2\text{-}7)$$

As the target dose and the OAR dose approach their goal doses, the evaluation of Equation 2-7 quickly approaches infinity.

Beamlets and Dose

A beamlet, as has already been shown, affects dose to the patient. It is a way of ideally representing the relationship between a rectangular fluence cross section and the dose to the voxels affected by the fluence projecting through that cross section. In the case of delivery by MLC, the cross section unit is defined by the MLC leaf width in one direction and some, perhaps smaller, measure in the direction of leaf movement. In XiO, in the case of delivery by compensator, the cross section is independent of the compensator milling precision, and there is an interpolation process that converts beamlet intensities to compensator thicknesses. During the inverse planning optimization process, beamlet weights are systematically changed in order to simulate the dosimetric effect of increasing or decreasing fluence for different beamlets[2].

In XiO IMRT, and other inverse planning systems, initial beamlet weights before optimization begins are set to unity (1.0). Most IMRT treatments are isocentric and, at isocenter, relative beamlet voxel doses are typically normalized so that the combination of beamlet voxel doses is correlated with the dose goals of the target. In XiO, beamlets are normalized so that at the isocenter beamlet doses sum to the lowest goal dose of any active target.

Table 2-1 shows possible voxel doses for the two voxel patient. The prescription for our example specifies a goal of 64 Gy for the target. Therefore, the four beamlets, when weighted 1.0, should each equally contribute 16 Gy to the dose at voxel *T*. Of the two beamlets that intersect voxel *N*, the dose from beamlet 1 is 15 Gy and the dose from beamlet 3 is 17 Gy. Note that these are physically reasonable, but arbitrary, dose values.

Voxel	Beamlet 1	Beamlet 2	Beamlet 3	Beamlet 4
T	16	16	16	16
N	15	-	17	-

Table 2-1: Beamlet voxel doses for two-voxel patient

Given these beamlet definitions

2 Beamlets are not recalculated during the optimization process as it is assumed that there is a strict linear relationship between units of fluence and dose. Of course, this is not exactly true. Combinations of beamlets, which represent different MLC apertures, can affect the apparent dose output. This is one of the reasons that a final dose calculation is important. Notice also that beamlet weights, which relate to fluence differences in a treatment field, are not the same as importance weights, which relate to the relative importance of objectives.

beamlet的权重和importance的权重是不一样的，beamlet与治疗野的通量差异有关，而importance是与目标的重要性有关系。

$$D_T = 16 \cdot w_1 + 16 \cdot w_2 + 16 \cdot w_3 + 16 \cdot w_4 \qquad (2\text{-}8)$$

$$D_N = 15 \cdot w_1 + 17 \cdot w_3 \qquad \textbf{(2-9)}$$

where w_i is a beamlet weight.

Gradient Descent Optimization

XiO IMRT uses a "conjugate gradient" optimization algorithm, which is a special type of "gradient descent" optimization algorithm [Nocedal and Wright, 2000; Press *et al.,* 2002]. Both conjugate gradient and gradient descent methods use the negative gradient of a cost function, with respect to input parameters, to systematically find the minimum of a cost function.

An inverse planning cost function may be defined to be a sum of dose objectives that is modulated by a collection of beamlet weights. The gradient for such a cost function, according to its mathematical definition, is a normalized vector composed of partial derivatives of the cost function objectives with respect to the beamlet weights. The optimizer multiplies the gradient by -1 to define the direction in which to search for a minimum of the cost function.

Gradient descent optimization consists of a series of iterations. Each iteration requires a gradient calculation followed by a search in the direction of the negative gradient until a minimum is found. The search in the direction of the negative gradient is called a "line search" [Nocedal and Wright, 2000; Press *et al.,* 2002]. The minimum that is found during a line search is usually not the global minimum of the cost function. It is simply a minimum of the cost function in that direction. The optimizer iterates for a specific number of cycles or until the cost function converges. Convergence occurs when the current evaluation of the cost function is not significantly better than the last evaluation. In most optimizers, including that of XiO, the user controls the maximum number of iterations and the convergence criterion.

The difference between the conjugate gradient and gradient descent methods is in the sequence of search directions that is selected during optimization. Conjugate directions are influenced by the local gradient but usually do not exactly follow the negative gradient for reasons that will be explained in *Optimization – Advanced Concepts*. However, it should be noted that conjugate gradient algorithms always use the negative gradient direction as the first direction in which to search and may resort to the negative gradient in cases in which the conjugate direction does not yield a meaningful cost improvement.

Later in this chapter, when we use the two voxel patient to facilitate a discussion of the internals of the optimization process, we will model gradient descent optimization, avoiding the unnecessary complication of determining conjugate directions.

Cost Function Evaluation

Given our beamlet definitions and unit beamlet weights, the evaluation of the cost function for the two voxel patient is:

$$F(D_T, D_N) = (D_T - 64)^2 + (D_N)^2 \qquad (2\text{-}10)$$

$$= ((16 \cdot w_1 + 16 \cdot w_2 + 16 \cdot w_3 + 16 \cdot w_4) - 64)^2 + (15 \cdot w_1 + 17 \cdot w_3)^2 \qquad \textbf{(2-11)}$$

$$= (0)^2 + (32)^2 = 1024 \qquad \textbf{(2-12)}$$

Cost function evaluations, during an optimization process, would be similar, except that target and OAR doses would be affected by new beamlet weights.

Cost Function Derivatives

Derivatives tell us about relationships between quantities. Specifically, they tell us how much a change in one quantity will affect another related quantity. In inverse planning, the optimizer changes beamlet weights and evaluates changes in its cost function. Therefore, the derivatives used in inverse planning must describe how changes in beamlet weights affect the cost function.

For each beamlet weight, there is a derivative:

$$\frac{\partial F}{\partial w} = \sum_{\sigma} \frac{\partial O_{\sigma}}{\partial w} \qquad (2\text{-}13)$$

It expresses the derivative of the cost function with respect to a beamlet's weight as the sum of the derivatives of all objectives with respect to that beamlet's weight.

For our example we have, equivalently:

$$\frac{\partial F}{\partial w} = \frac{\partial O_T}{\partial w} + \frac{\partial O_N}{\partial w}. \qquad (2\text{-}14)$$

The derivative of each objective with respect to the beamlet's weight can be further expressed as:

$$\frac{\partial O_T}{\partial w} = \frac{\partial O_T}{\partial D_T} \cdot \frac{\partial D_T}{\partial w} \text{ and} \qquad (2\text{-}15)$$

$$\frac{\partial O_N}{\partial w} = \frac{\partial O_N}{\partial D_N} \cdot \frac{\partial D_N}{\partial w}, \qquad \textbf{(2-16)}$$

Taking derivatives with respect to dose of Equations 2-3 and 2-4, we have:

$$\frac{\partial O_T}{\partial D_T} = 2 \cdot (D_T - D_{GT}) = 2 \cdot (D_T - 65) \text{ and} \qquad (2\text{-}17)$$

$$\frac{\partial O_N}{\partial D_N} = 2 \cdot (D_N - D_{GN}) = 2 \cdot D_N \qquad \textbf{(2-18)}$$

The derivatives of dose with respect to beamlet weight:

$$\frac{\partial D_T}{\partial w} \textbf{ and } \frac{\partial D_T}{\partial w} \qquad \textbf{(2-19, 2-20)}$$

are just the beamlet voxel doses as shown in

Figure 2-1, since there is a linear relationship between beamlet weights and these doses.

Thus, the general cost function derivative with respect to any beamlet weight in our example can be expressed

$$\frac{\partial F}{\partial w} = \frac{\partial O_T}{\partial D_T} \cdot \frac{\partial D_T}{\partial w} + \frac{\partial O_N}{\partial D_N} \cdot \frac{\partial D_N}{\partial w} \qquad (2\text{-}21)$$

$$= 2 \cdot (D_T - 64) \cdot \frac{\partial D_T}{\partial w} + 2 \cdot D_N \cdot \frac{\partial D_N}{\partial w} \qquad \textbf{(2-22)}$$

Finding the Global Minimum

It is possible, but unlikely, with initial beamlet weights that a cost function evaluation will yield a global minimum; that is, yield a solution that is a unique best minimum cost function evaluation. For instance, with the two voxel patient, it can easily be shown that the initial cost function evaluation with unit beamlet weights is not a minimum. However, because the system described by the beamlets, doses, and objectives for the two voxel patient is such a simple one, beamlet weights that will obtain a global minimum of the cost function can be calculated using algebra alone.

Recalling the cost function, Equation 2-5, if beamlets 1 and 3 are turned off, only the part of the cost function related to dose from beamlets 2 and 4, which only deliver dose to the target, remains. Suppose that we want to find equal weights for beamlets 2 and 4 that minimize the cost function. The optimal weight for beamlets 2 and 4 could be calculated

$$w = \frac{D_{GT}}{2} \cdot \frac{w_0}{D} \qquad (2\text{-}23)$$

$$= \frac{64}{2} \cdot \frac{1}{16} = 2 \qquad \textbf{(2-24)}$$

where w_0 is the initial beamlet weight and D is the dose contribution per unit intensity to the target voxel for either beamlet. With beamlet weights of {0, 2, 0, 2} the evaluation of the cost function is 0.0, the smallest value that the cost function can have for any set of inputs, and is thus a global minimum.

Because beamlets 2 and 4 are able to deliver dose to the target without delivering any dose to the OAR, there are an infinite number of beamlet ratios that could yield a dose of 64 Gy to the target and 0 Gy to the OAR. In reality, inverse planning involves much more complicated geometries, more voxels, more beamlets, dose distribution tradeoffs, etc. As such is the case, "closed form" solutions are typically not available; hence, the need for optimization methods to minimize cost functions.

Optimization as a Search in a Space

The legal range of beamlet weights defines a search space in which an optimizer can search for a minimum of a cost function. If there are 1000 beamlet weights to be optimized, the optimizer searches in a 1000-dimensional space, each combination of beamlet weights defining a different point in the space[3].

3 Notice that there is no mention of the beam organization of beamlets in our discussion. The fact that beamlets belong to beams is only important with respect to the calculation of minimum beamlet intensities, which will be explained in Optimization – Advanced Concepts.

By a legal range, it is meant that, at very least, a beamlet weight cannot be negative. Allowing negative beamlet weights would be equivalent to allowing negative fluence. Not allowing negative beamlet weights is what is called a hard constraint by some authors [Spirou and Chui, 1998]. The XiO optimizer does not allow negative beamlet weights and also does not allow beamlets to be weighted zero if a nonzero minimum transmission for a beam has been specified [Ezzell *et al.,* 2003].

Other constraint-like conditions that can contribute to defining a search space are implemented in some systems. One of the more common is the objective to control the degree of modulation of beamlet weights [Alber, 2001, Spirou *et al.,* 2001]. Smoother or less "noisy" intensity maps generally improve field segmentation for MLC delivery and make compensators easier to fabricate. This improvement comes at the price of potentially-degraded dosimetric results and its utility is likely a function of the type of optimizer that is used.

Line Searches

The process of searching for a minimum in a specific direction in a function space is known as a "line search" [Nocedal and Wright, 2000; Press *et al.* 2002]. "Line search magnitude" is defined as a normalized "distance" in a line search direction from some starting point in the search space. The "line search minimum" is the distance at which a minimum is found during a line search.

There are many types of line search methods. XiO uses two: one for the initial line search and another for each subsequent line search. Implementation details of the XiO line search methods are discussed in *Optimization – Advanced Concepts*.

In the following example, we simulate a line search algorithm using a commercially-available spreadsheet program. The simulated line search will be informative, but not especially accurate. Real line search algorithms "home in" on a line search minimum with very high accuracy, successively trying values closer and closer to the true minimum until some convergence criterion is met.

The spreadsheet line search, just like any other inverse planning line search, updates beamlet weights and evaluates the effect of changing anatomy doses on the objective-based cost function. It requires access to the following:

- Beamlet definitions
- Current beamlet weights
- Search direction
- Dose update method
- Cost evaluation method

Looking ahead, Table 2-2 shows a spreadsheet for the spreadsheet line search. Column 1 shows the line search magnitude. Notice that the values are evenly spaced according to some arbitrary step size (in this case the step is 0.1). This is one difference between real line search algorithms and this spreadsheet line search. Real line search algorithms typically refine the step size as they converge to the minimum. In the spreadsheet line search, a fixed step size is selected by trial and error; that is, when we change the fixed

step size to a new value and we see the minimum, as in Figure 2-4, we have found both the minimum and the step size. Of course, the minimum that is found in this way is potentially in error by up to ½ of the step size, but this level of accuracy, as will later be shown, is sufficient for at least a few simulated optimization iterations.

Columns 2-5 and 7-10 show individual beamlet doses to the target and OAR as a function of the line search magnitude. Columns 6 and 11 show the sum of the beamlet doses to the target and OAR. Column 12 shows the cost as a function of the line search magnitude.

Thinking Like an Optimizer

In this section, we try to "think" like an optimizer, modeling a basic gradient descent optimizer for simplicity. We will learn how to calculate search directions and simulate a line search algorithm using a commercially-available spreadsheet program. As stated earlier, the difference between the conjugate gradient and gradient descent methods is in the search directions that are selected during optimization. This difference is critical for optimizer performance, but should not detract from your understanding of the basic optimization process. Even though the calculations that are used will not be exactly like those in XiO, this promises to be an informative and challenging exercise.

Initial Cost

The cost, given the initial unit beamlet weights, was previously calculated in the Section Beamlets and Dose to be 1024 (unitless). Figure 2-2 shows the cost per voxel with respect to dose for the objectives. Since there are only two voxels, the cost is simply the sum of O_N for 32 Gy and O_T for 64 Gy; that is, the values that can be read off of the quadratic curves on the graph at 32 Gy for the OAR and at 64 Gy for the target.

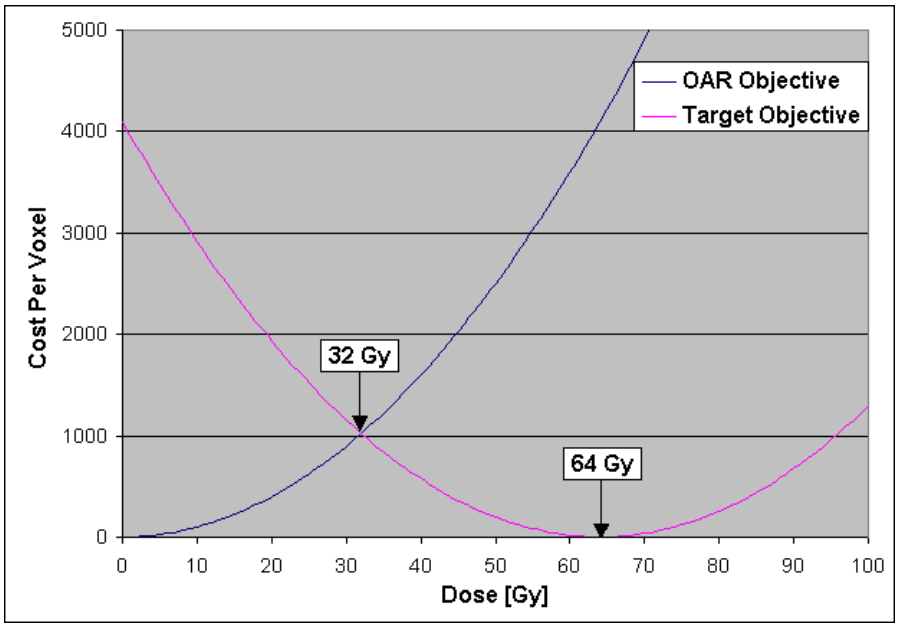


Figure 2-2: Quadratic objectives with initial OAR and target doses

Initial Search Direction

The search direction is determined from the evaluation of derivatives at the current point in the search space. Recall Equation 2-14b, the general derivative of cost with respect to any beamlet weight. With initial weights of 1.0 for all beamlets, D_T is 64 Gy and D_N is 32 Gy. The general derivative becomes

$$\frac{\partial F}{\partial w} = 2 \cdot (64 - 64) \cdot \frac{\partial D_T}{\partial w} + 2 \cdot 32 \cdot \frac{\partial D_N}{\partial w} \qquad (2\text{-}25)$$

$$= 2 \cdot 32 \cdot \frac{\partial D_N}{\partial w} \qquad \textbf{(2-26)}$$

Derivatives of cost with respect to beamlet weight can be calculated for each beamlet

$$\frac{\partial F}{\partial w_1} = 2 \cdot 32 \cdot 15 = 960 \qquad (2\text{-}27)$$

$$\frac{\partial F}{\partial w_2} = 2 \cdot 32 \cdot 0 = 0 \qquad \textbf{(2-28)}$$

$$\frac{\partial F}{\partial w_3} = 2 \cdot 32 \cdot 17 = 1088 \qquad (2\text{-}29)$$

$$\frac{\partial F}{\partial w_4} = 2 \cdot 32 \cdot 0 = 0 \qquad (2\text{-}30)$$

Only the derivatives for beamlet 1 and 3 are needed to define the initial direction of search for the minimum, since the evaluation of the derivatives for beamlets 2 and 4 is zero. Normalizing these derivatives, we obtain

$$\left|\frac{\partial F}{\partial w_1}\right| = \frac{960}{\sqrt{960^2 + 1088^2}} = 0.6616216 \qquad (2\text{-}31)$$

$$\left|\frac{\partial F}{\partial w_3}\right| = \frac{1088}{\sqrt{960^2 + 1088^2}} = 0.7498379 \qquad (2\text{-}32)$$

Thus, the initial direction, multiplying the gradient by –1, is

$$-\nabla F = -0.6616216\hat{w}_1 - 0.7498379\hat{w}_3 \qquad (2\text{-}33)$$

where $\hat{w}_1$ and $\hat{w}_3$ are unit vectors along the w_1 and w_3 "axes" in the search space.

Line Search Minimum

The above negative gradient direction tells us that, during the first line search, the weights for beamlets 1 and 3 will be relaxed, allowing dose to the OAR and the target to decrease. As the weights are relaxed, the cost for the OAR objective will improve (decrease) and the cost function for the target objective will worsen (increase) until the point at which any improvement in the OAR objective will be outweighed by a worsening of the target objective. Because of the relative simplicity of this problem, the line search

minimum value can be calculated algebraically. We will later compare this algebraic minimum to the line search minimum found using the spreadsheet line search.

Given the (negative) gradient that was calculated, beamlet weights vary with the line search magnitude (L) according to the following relationships:

$$w_1 = 1 - 0.6616216 \cdot L \qquad (2\text{-}34)$$

$$w_2 = 1 \qquad \textbf{(2-35)}$$

$$w_3 = 1 - 0.7498379 \cdot L \qquad \textbf{(2-36)}$$

$$w_4 = 1 \qquad \textbf{(2-37)}$$

Substituting Equations 2-34 – 2-37 into Equations 2-8 and 2-9 yields

$$D_T = 64 - 22.583352 \cdot L \text{ and} \qquad (2\text{-}38)$$

$$D_N = 32 - 22.671568 \cdot L \qquad \textbf{(2-39)}$$

Setting Equations 2-3 and 2-4 to be equal and substituting Equations 2-20a and 2-20b yields

$$O_T(D_T) = O_N(D_N) \qquad (2\text{-}40)$$

$$((64 - 22.583352 \cdot L) - 64)^2 = (32 - 22.671568 \cdot L)^2 \quad \textbf{(2-41)}$$

$$510.00779 \cdot L^2 = 1024 - 1450.9804 \cdot L + 514.00001 \cdot L^2 \quad \textbf{(2-42)}$$

Rearranging...

$$3.99222 \cdot L^2 - 1450.9804 \cdot L + 1024 = 0 \qquad \textbf{(2-43)}$$

We set the target and OAR objectives equal, above, because we were looking for the point along the line search direction where the objectives would "pull" equally against each other.

Equation 2-22d is what is known as a quadratic equation; that is, it is in the form

$$ax^2 + bx + c = 0 \qquad (2\text{-}44)$$

Quadratic equations can be solved by use of the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad (2\text{-}45)$$

Thus,

$$L = \frac{1450.9804 \pm \sqrt{(1450.9804)^2 - 4 \cdot 3.99222 \cdot 1024}}{2 \cdot 3.99222} \qquad \textbf{(2-46)}$$

$$= \frac{1450.9804 \pm 1445.3346}{7.98444} \qquad \textbf{(2-47)}$$

$$\textbf{= 0.7071003, 362.74491} \qquad \textbf{(2-48)}$$

Of the two solutions, the first, 0.7071003, is the one that we are looking for, because we know that the weights must decrease but not become less than zero (which would be outside of our legal search space). Solving for weights at the line search minimum:

$$w_1 = 1 - 0.6616216 \cdot L = 1 - 0.6616216 \cdot 0.7071003 = 0.5321672 \quad (2\text{-}49)$$

$$w_2 = 1 \quad \textit{(2-50)}$$

$$w_3 = 1 - 0.7498379 \cdot L = 1 - 0.7498379 \cdot 0.7071003 = 0.4697894 \quad \textit{(2-51)}$$

$$w_4 = 1 \quad \textit{(2-52)}$$

Again, substituting into Equations 2-8 and 2-9:

$$D_T = 16 \cdot 0.5321672 + 16 \cdot 1.0 + 16 \cdot 0.4697894 + 16 \cdot 1.0 = 48.031306 \quad (2\text{-}53)$$

$$D_N = 15 \cdot 0.5321672 + 17 \cdot 0.4697894 = 15.968928 \quad \textit{(2-54)}$$

Error! Reference source not found.shows the algebraically-calculated doses for the initial line search minimum. As predicted, the new doses pull more evenly against each other, as shown by their more equal objective costs.

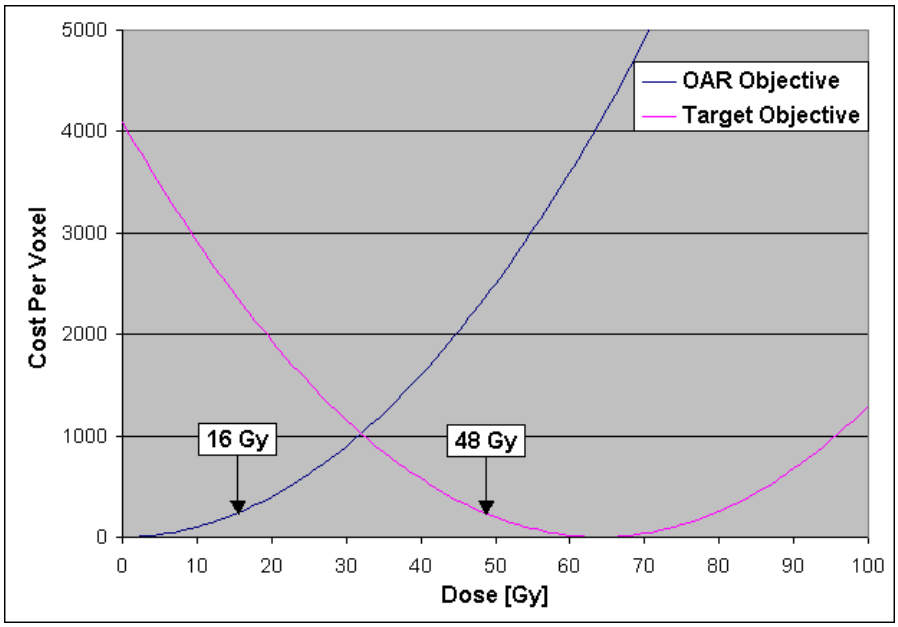


Figure 2-3: Quadratic objectives with OAR and target doses after initial line search.

The new objective evaluations are

$$O_T(48.031306) = (48.031306 - 64)^2 = 254.99919 \quad (2\text{-}55)$$

$$O_N(15.968928) = 15.968928^2 = 255.00666 \quad \textit{(2-56)}$$

Since the total cost is the sum of the individual objective costs, the new cost after a single line search would be approximately 510.00585. This is a cost improvement of roughly

$$\frac{1024 - 510.00585}{1024} \cdot 100 = 50.2\% \qquad (2\text{-}57)$$

The predicted cost improvement during the first iteration of the optimizer is typical. Approximately 95% of all improvement occurs in the first 30 iterations. Further iterations may be valuable, however, because it is often necessary to run an optimizer 100 or more iterations to get the most possible benefit out of the optimization process.

Initial Line Search

In theory, the spreadsheet line search algorithm should be able to find the minimum that was found in the previous algebraic calculations. In practice, the minimum that is found is a function of the smallest step and other criteria allowed by a line search. Table 2-2 shows the calculation of the spreadsheet line search. The row in bold italics is its best guess at a minimum for the cost function along the same negative gradient used in the algebraic computation of the minimum of the initial line search.

As mentioned earlier in our discussion of the spreadsheet line search, the spreadsheet is set up and then different step sizes are selected by trial and error until the right step size that finds the minimum is selected. The comparison of results is quite good.

- Final line search magnitude of 0.70 vs. 0.71 (1.4% error)
- Final target dose of 48.32 Gy vs. 48.03 Gy (-0.60 % error)
- Final OAR dose of 16.26 Gy vs. 15.97 (-1.7% error)
- Final cost of 520.38 vs. 510. 00 (-2.0 % error)

Getting close to the exact line search minimum is important for efficiency, but small errors do not deter optimizers from eventually finding a global minimum, or at least, getting very close to it.

Line Search Magnitude	Dose Per Beamlet (Target)				Target Dose	Dose Per Beamlet (OAR)				OAR Dose	Total Cost
	B1	B2	B3	B4		B1	B2	B3	B4		
0.00	16.00	16.00	16.00	16.00	64.00	15.00	0.00	17.00	0.00	32.00	1024.00
0.10	14.94	16.00	14.82	16.00	61.76	14.01	0.00	15.74	0.00	29.75	890.20
0.20	13.89	16.00	13.63	16.00	59.52	13.02	0.00	14.48	0.00	27.50	776.54
0.30	12.83	16.00	12.45	16.00	57.28	12.03	0.00	13.23	0.00	25.26	683.02
0.40	11.78	16.00	11.26	16.00	55.04	11.04	0.00	11.97	0.00	23.01	609.65
0.50	10.72	16.00	10.08	16.00	52.80	10.05	0.00	10.71	0.00	20.76	556.42
0.60	9.66	16.00	8.90	16.00	50.56	9.06	0.00	9.45	0.00	18.51	523.33
0.70	***8.61***	***16.00***	***7.71***	***16.00***	***48.32***	***8.07***	***0.00***	***8.19***	***0.00***	***16.26***	***510.38***
0.80	7.55	16.00	6.53	16.00	46.08	7.08	0.00	6.94	0.00	14.02	517.57
0.90	6.50	16.00	5.34	16.00	43.84	6.09	0.00	5.68	0.00	11.77	544.91
1.00	5.44	16.00	4.16	16.00	41.60	5.10	0.00	4.42	0.00	9.52	592.39
1.10	4.38	16.00	2.98	16.00	39.36	4.11	0.00	3.16	0.00	7.27	660.01

1.20	3.33	16.00	1.79	16.00	37.12	3.12	0.00	1.90	0.00	5.02	747.77
1.30	2.27	16.00	0.61	16.00	34.88	2.13	0.00	0.65	0.00	2.78	855.68
1.40	1.22	16.00	0.00	16.00	33.22	1.14	0.00	0.00	0.00	1.14	948.95
1.50	0.16	16.00	0.00	16.00	32.16	0.15	0.00	0.00	0.00	0.15	1013.81

Table 2-2: Line Search Values - Initial Search Direction

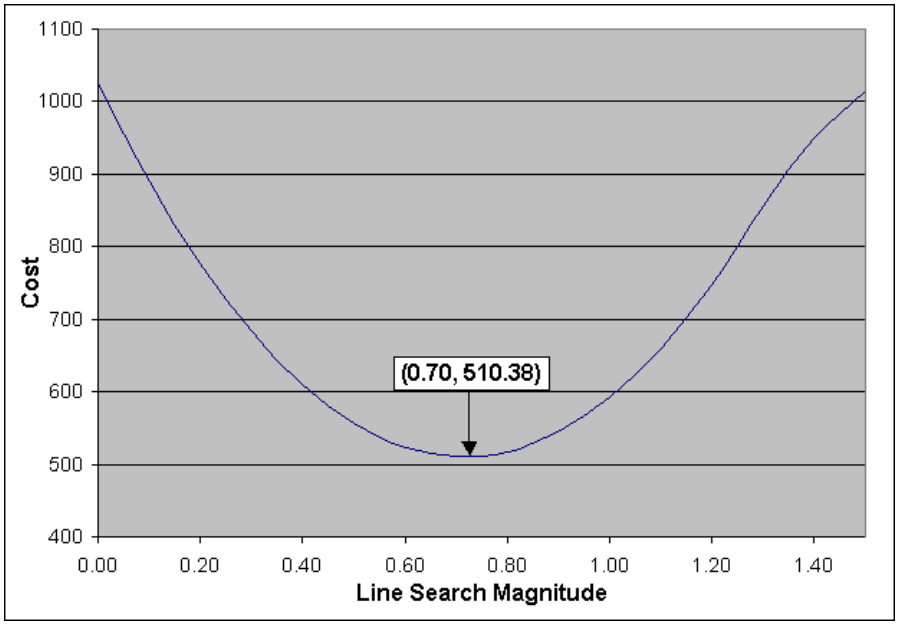


Figure 2-4: Line Search Costs – Initial Search Direction

Geometric Interpretation

Because only two of the beamlet weights are changed during the initial line search, it is easy to visualize the search process in a two-dimensional graphic, as shown in Figure 2-5. The line search is along the direction defined by the negative gradient calculated in Equation 2-33. The spreadsheet line search calculates all of the line search magnitudes between 0.0 and a line search distance of 1.5 in increments of 0.1.

Line Search Magnitude	$\frac{\partial F}{\partial w_1}$	$\frac{\partial F}{\partial w_2}$	$\frac{\partial F}{\partial w_2}$	$\frac{\partial F}{\partial w_4}$
	-0.66	0.00	-0.74	0.00
	w_1	w_2	w_3	w_4
0.00	1.000	1.000	1.000	1.000
0.10	0.934	1.000	0.926	1.000
0.20	0.868	1.000	0.852	1.000
0.30	0.802	1.000	0.778	1.000
0.40	0.736	1.000	0.704	1.000
0.50	0.670	1.000	0.630	1.000
0.60	0.604	1.000	0.556	1.000
0.70	***0.538***	***1.000***	***0.482***	***1.000***
0.80	0.472	1.000	0.408	1.000
0.90	0.406	1.000	0.334	1.000
1.00	0.340	1.000	0.260	1.000
1.10	0.274	1.000	0.186	1.000
1.20	0.208	1.000	0.112	1.000
1.30	0.142	1.000	0.038	1.000
1.40	0.076	1.000	0.000	1.000
1.50	0.010	1.000	0.000	1.000

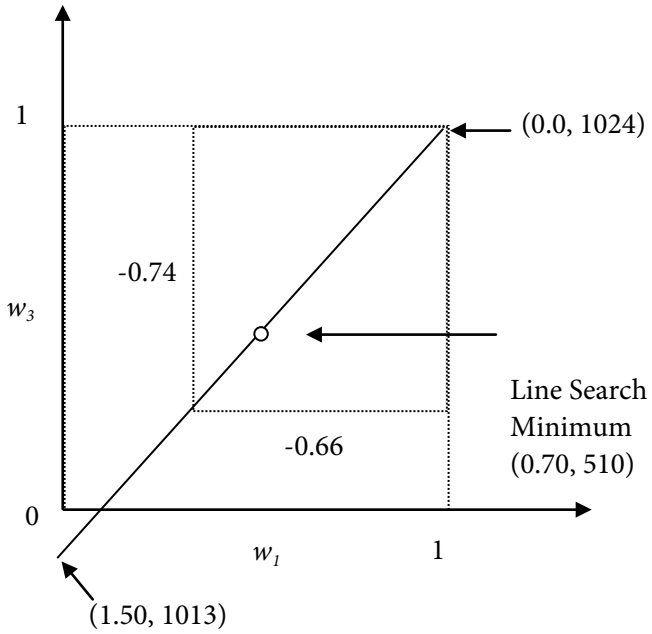


Figure 2-5: Two-dimensional representation of the initial line search

Second Iteration

With the new weights for the beamlets from the spreadsheet line search, rather than from the algebraic calculation, D_T is 48.32 Gy and D_N is 16.26 Gy. (We limit further calculations to two decimal digits accuracy for simplicity.) Derivatives of cost with respect to beamlet weights can *again* be calculated for each beamlet, defining a new direction in the search space.

$$\frac{\partial F}{\partial w} = 2 \cdot (48.32 - 64) \cdot \frac{\partial D_T}{\partial w} + 2 \cdot 16.26 \cdot \frac{\partial D_N}{\partial w} \qquad (2\text{-}58)$$

$$\frac{\partial F}{\partial w_1} = 2 \cdot (-15.68) \cdot 16.0 + 2 \cdot 16.26 \cdot 15.0 = -13.96 \qquad (2\text{-}59)$$

$$\frac{\partial F}{\partial w_2} = 2 \cdot (-15.68) \cdot 16.0 = -501.76 \qquad \textbf{(2-60)}$$

$$\frac{\partial F}{\partial w_3} = 2 \cdot (-15.68) \cdot 16.0 + 2 \cdot 16.26 \cdot 17.0 = 51.08 \quad \textbf{(2-61)}$$

$$\frac{\partial F}{\partial w_4} = 2 \cdot (-15.68) \cdot 16.0 = -501.76 \quad \textbf{(2-62)}$$

$$\sqrt{(-13.96)^2 + (-501.76)^2 + 51.08^2 + (-501.76)^2} = 711.57 \quad (2\text{-}63)$$

$$\left|\frac{\partial F}{\partial w_1}\right| = \frac{-13.96}{711.57} = -0.020 \quad (2\text{-}64)$$

$$\left|\frac{\partial F}{\partial w_2}\right| = \frac{-501.76}{711.57} = -0.71 \quad \textbf{(2-65)}$$

$$\left|\frac{\partial F}{\partial w_3}\right| = \frac{51.08}{711.57} = 0.072 \quad \textbf{(2-66)}$$

$$\left|\frac{\partial F}{\partial w_4}\right| = \frac{-501.76}{711.57} = -0.71 \quad \textbf{(2-67)}$$

Thus

$$-\nabla F = 0.20\hat{w}_1 + 0.71\hat{w}_2 - 0.0.072\hat{w}_3 + 0.71\hat{w}_4 \quad (2\text{-}68)$$

Line Search Magnitude	Dose Per Beamlet (Target)				Target Dose	Dose Per Beamlet (OAR)				OAR Dose	Total Cost
	B1	B2	B3	B4		B1	B2	B3	B4		
0.00	8.61	16.00	7.71	16.00	48.32	8.07	0.00	8.19	0.00	16.26	510.38
0.10	8.64	17.14	7.83	17.14	50.74	8.10	0.00	8.32	0.00	16.42	445.35
0.20	8.67	18.27	7.94	18.27	53.16	8.13	0.00	8.44	0.00	16.57	392.07
0.30	8.70	19.41	8.06	19.41	55.58	8.16	0.00	8.56	0.00	16.72	350.54
0.40	8.74	20.54	8.17	20.54	58.00	8.19	0.00	8.68	0.00	16.87	320.76
0.50	8.77	21.68	8.29	21.68	60.42	8.22	0.00	8.81	0.00	17.03	302.73
0.60	***8.80***	***22.82***	***8.40***	***22.82***	***62.84***	***8.25***	***0.00***	***8.93***	***0.00***	***17.18***	***296.45***
0.70	8.83	23.95	8.52	23.95	65.25	8.28	0.00	9.05	0.00	17.33	301.93
0.80	8.86	25.09	8.63	25.09	67.67	8.31	0.00	9.17	0.00	17.48	319.16
0.90	8.90	26.22	8.75	26.22	70.09	8.34	0.00	9.30	0.00	17.64	348.14
1.00	8.93	27.36	8.86	27.36	72.51	8.37	0.00	9.42	0.00	17.79	388.87
1.10	8.96	28.50	8.98	28.50	74.93	8.40	0.00	9.54	0.00	17.94	441.35
1.20	8.99	29.63	9.09	29.63	77.35	8.43	0.00	9.66	0.00	18.09	505.58
1.30	9.02	30.77	9.21	30.77	79.77	8.46	0.00	9.79	0.00	18.25	581.57
1.40	9.06	31.90	9.32	31.90	82.19	8.49	0.00	9.91	0.00	18.40	669.30
1.50	9.09	33.04	9.44	33.04	84.61	8.52	0.00	10.03	0.00	18.55	768.79

Table 2-3: Line Search Values – Second Search Direction

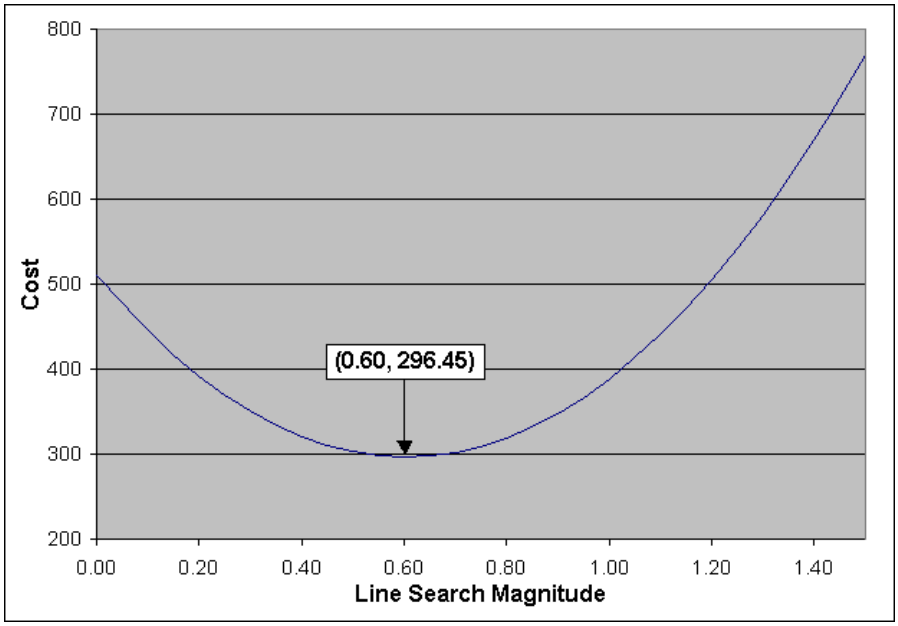


Figure 2-6: Line Search Costs – Second Search Direction

According to Table 2-3, the minimum cost for the second line search direction is 296.45. The improvement in cost from the first to second iteration is then

$$\frac{510.01-296.45}{1024}\cdot 100 = 21.86\% \qquad (2\text{-}69)$$

Third Iteration

We do one final line search. With the new weights for the beamlets, D_T is 62.84 Gy and D_N is 17.18 Gy. Thus:

$$\frac{\partial F}{\partial w} = 2\cdot(62.84-64)\cdot\frac{\partial D_T}{\partial w} + 2\cdot 17.18\cdot\frac{\partial D_N}{\partial w} \qquad \textbf{(2-70)}$$

$$\frac{\partial F}{\partial w_1} = 2\cdot(-1.16)\cdot 16.0 + 2\cdot 17.18\cdot 15.0 = 478.28 \qquad \textbf{(2-71)}$$

$$\frac{\partial F}{\partial w_2} = 2\cdot(-1.16)\cdot 16.0 = -37.12 \qquad \textbf{(2-72)}$$

$$\frac{\partial F}{\partial w_3} = 2\cdot(-1.16)\cdot 16.0 + 2\cdot 17.18\cdot 17.0 = 547.00 \qquad \textbf{(2-73)}$$

$$\frac{\partial F}{\partial w_4} = 2\cdot(-1.16)\cdot 16.0 = -37.12 \qquad \textbf{(2-74)}$$

$$\sqrt{478.28^2 + (-37.12)^2 + 547.00^2 + (-37.12)^2} = 728.50 \qquad \textbf{(2-75)}$$

$$\left|\frac{\partial F}{\partial w_1}\right| = \frac{478.28}{728.50} = 0.66 \qquad \textbf{(2-76)}$$

$$\left|\frac{\partial F}{\partial w_2}\right| = \frac{-37.12}{728.50} = -0.051 \qquad \textbf{(2-77)}$$

$$\left|\frac{\partial F}{\partial w_3}\right| = \frac{547.00}{728.50} = 0.75 \qquad \textbf{(2-78)}$$

$$\left|\frac{\partial F}{\partial w_4}\right| = \frac{-37.12}{728.50} = -0.051 \qquad \textbf{(2-79)}$$

Thus,

$$-\nabla F = -0.66\hat{w}_1 + 0.051\hat{w}_2 - 0.75\hat{w}_3 + 0.0.51\hat{w}_4 \quad \textbf{(2-80)}$$

Line Search Magnitude	Dose Per Beamlet (Target)				Target Dose	Dose Per Beamlet (OAR)				OAR Dose	Total Cost
	B1	B2	B3	B4		B1	B2	B3	B4		
0.00	8.80	22.82	8.40	22.82	62.84	8.25	0.00	8.93	0.00	17.18	296.45
0.05	8.27	22.86	7.80	22.86	61.79	7.76	0.00	8.29	0.00	16.05	262.36
0.10	7.74	22.90	7.20	22.90	60.74	7.26	0.00	7.65	0.00	14.91	233.02
0.15	7.22	22.94	6.60	22.94	59.70	6.77	0.00	7.02	0.00	13.78	208.44
0.20	6.69	22.98	6.00	22.98	58.65	6.27	0.00	6.38	0.00	12.65	188.61
0.25	6.16	23.02	5.40	23.02	57.60	5.78	0.00	5.74	0.00	11.52	173.54
0.30	5.63	23.06	4.80	23.06	56.56	5.28	0.00	5.10	0.00	10.38	163.22
0.35	5.10	23.10	4.20	23.10	55.51	4.79	0.00	4.47	0.00	9.25	157.65
0.40	***4.58***	***23.14***	***3.60***	***23.14***	***54.46***	***4.29***	***0.00***	***3.83***	***0.00***	***8.12***	***156.84***
0.45	4.05	23.18	3.00	23.18	53.42	3.80	0.00	3.19	0.00	6.99	160.79
0.50	3.52	23.22	2.40	23.22	52.37	3.30	0.00	2.55	0.00	5.85	169.49
0.55	2.99	23.26	1.80	23.26	51.32	2.81	0.00	1.92	0.00	4.72	182.95
0.60	2.46	23.31	1.20	23.31	50.28	2.31	0.00	1.28	0.00	3.59	201.16
0.65	1.94	23.35	0.60	23.35	49.23	1.82	0.00	0.64	0.00	2.46	224.13
0.70	1.41	23.39	0.00	23.39	48.19	1.32	0.00	0.00	0.00	1.32	251.85
0.75	0.88	23.43	0.00	23.43	47.74	0.82	0.00	0.00	0.00	0.82	265.20

Table 2-4: Line Search Values - Third Search Direction

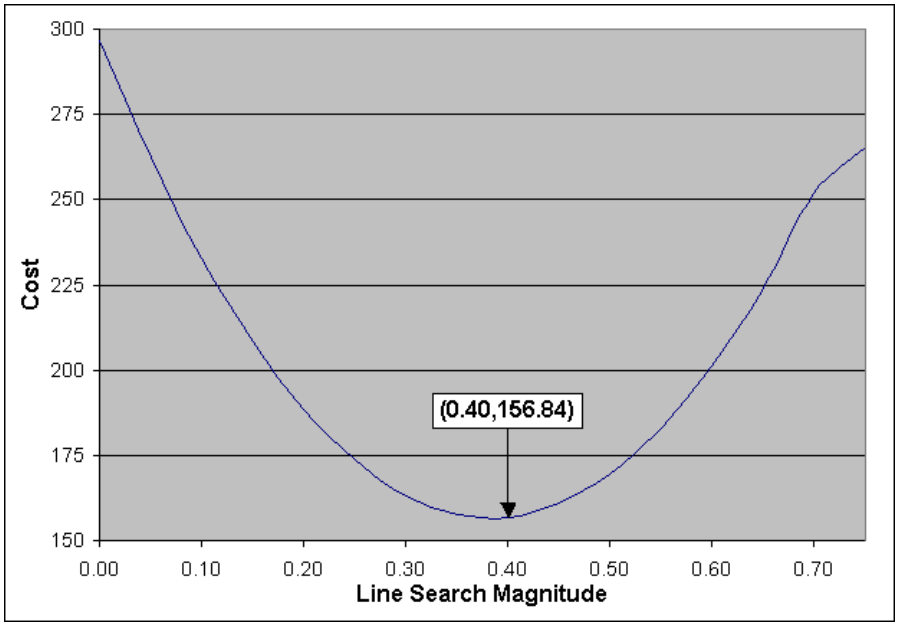


Figure 2-7: Line Search Costs – Third Search Direction

According to Table 2-4, the minimum cost for the third line search direction is 156.84. The improvement in cost from the second to third iteration is then

$$\frac{296.45-156.84}{1024}\cdot 100 = 13.63\% \qquad \textit{(2-81)}$$

Dose Convergence

From the earlier analysis of the cost function, we would expect D_T to converge to 64, and D_N to converge to 0 Gy. With enough iterations, the doses will certainly do just that. However, as shown in Table 2-5 and Figure 2-8, the progress towards these goals is somewhat circuitous.

D_T starts out at 64 Gy and due to pressure from the OAR to lower its dose, D_T immediately moves away from its goal in iteration 1. On iteration 2, much of D_T is recouped, but again on iteration 3 some of the regained dose is lost. All the while, D_N is making somewhat divergent steps towards its goal of 0 Gy. Such back and forth oscillation of doses is commonly found with gradient descent optimizers, but found less often with conjugate gradient optimizers. Hence, the observation that conjugate gradient optimizers are more efficient than gradient descent optimizers.

Iteration	D_T	D_N
0	64.00	32.00
1	48.32	16.26
2	62.84	17.18
3	54.46	8.12

Table 2-5: Dose Convergence Values

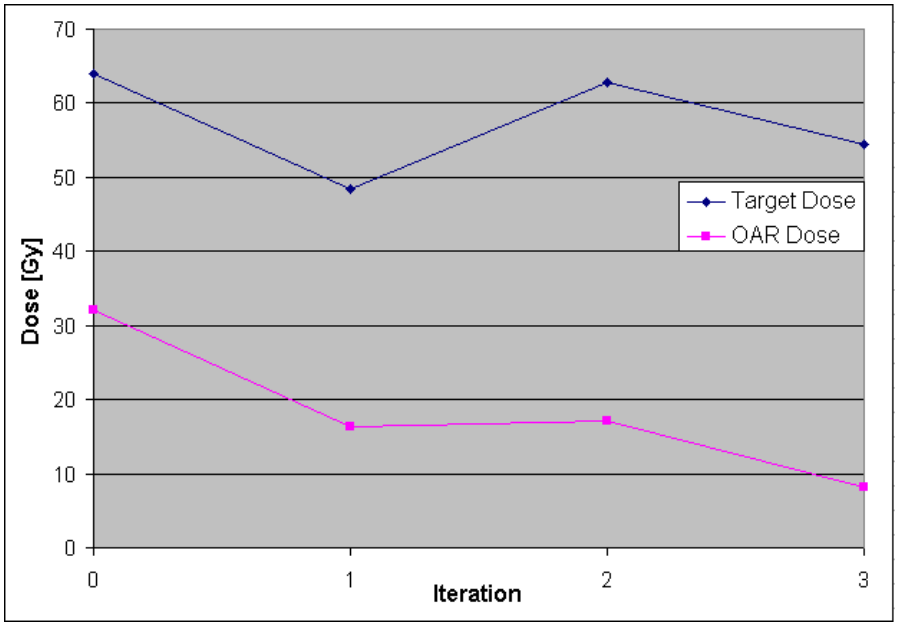


Figure 2-8: Dose Convergence

Cost Convergence

It is not surprising that the evaluation of total cost should decrease with every iteration. In fact, if costs do not decrease, the optimizer will stop. Table 2-6 and Figure 2-9 show that although the total cost decreases, the costs of the individual objectives may oscillate; this corresponds to the dose oscillation described earlier. From Figure 2-9, the trend of the cost is seen to begin an asymptotic approach to its limit, which is known from the previous algebraic calculation to be 0.0.

Table 2-6: Cost Convergence

Iteration	O_T	O_N	Total Cost
0	0.00	1024.00	1024.00
1	245.86	264.52	510.38
2	1.35	295.10	296.45
3	90.94	65.90	156.84

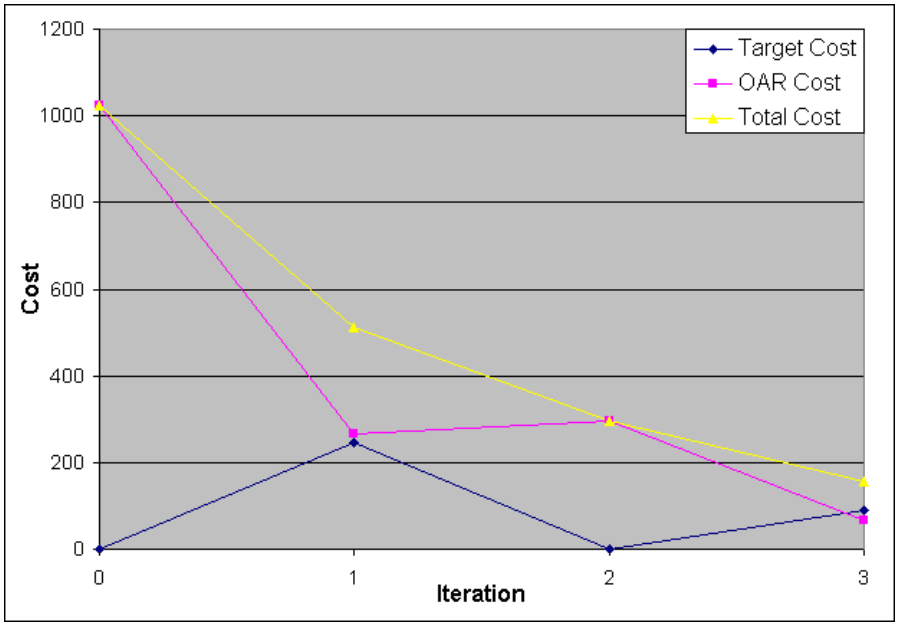


Figure 2-9: Cost Convergence

Summary

XiO uses a conjugate gradient optimizer and a cost function that is composed of the sum of dose and smoothing objective functions. During the optimization process, the optimizer finds the best combination of beamlet intensities, or beamlet weights that can deliver tumorcidal dose to targets while limiting dose to organs at risk and maintaining an slowly modulating intensity map. This combination of beamlet weights leads to a minimum evaluation of the cost function.

Beamlets are subdivisions of beams that represent an ideal relationship between fluence cross sections and dose to voxels in the patient. The legal range of beamlet weights defines a search space. This legal range is governed by what are called hard constraints. For instance, in XiO, the hard constraints are that the optimizer does not allow negative beamlet weights and also does not allow beamlets to be weighted zero if a non-zero minimum transmission for a beam has been specified.

The gradient optimization process consists of a series of iterations. Each iteration requires a gradient calculation followed by a line search for a minimum of the cost function in the direction of the negative gradient. The optimizer iterates for a specific number of cycles or until the cost function converges. Convergence occurs when the current evaluation of the cost function is not significantly better than the last evaluation. In most optimizers, including that of XiO, the user controls the maximum number of iterations and the convergence criterion. Conjugate gradient optimization, a modified form of gradient descent optimization, uses line search directions that are influenced by the local gradient but, for reasons of efficiency, do not exactly follow the negative gradient.

3. How to Use XiO IMRT

Successful inverse planning requires an understanding of the trade offs and limitations inherent in the inverse planning process. Foremost, it is important to understand the interplay between the specification of anatomy contours, beams, and objectives. This should not be surprising as anatomy contours define the structures whose dose we are interested in; beam specifications define how dosage is to occur; objectives set up dose goals and evaluate how well we have achieved them. In addition, the method of IMRT delivery, whether by multileaf collimator (MLC) or compensating filter, can affect the dosimetric results. Successful inverse planning also requires an understanding of how to evaluate results, whether it be through dose statistics, isodose displays, dose-volume histograms, or beamlet intensity maps. This chapter provides an introduction to the specification of inverse plans and the analysis of inverse planning results. It is meant to orient the first-time user or hone the skills of the more experienced.

One thing to remember as you engage in inverse planning is that optimizers can do amazing things, but they cannot accomplish the physically impossible. what is the physically impossible?

Inverse Planning Flow

This section is an introduction to XiO IMRT inverse planning. It centers on the "big picture", introducing inverse planning in XiO with an annotated flowchart. In subsequent sections, we will steadily expand on this foundation.

Figure 3-1 shows that the inverse planning process can be divided into two stages. During the first stage, the optimizer uses inverse planning objectives, anatomy contours, and beamlets to produce what are called ideal intensity maps and ideal doses. By "ideal", it is meant beamlet intensities and doses that might be delivered if the realities of treatment delivery could be ignored. 第一步就是优化出理想通量和剂量

In the second stage, ideal beamlet intensities are converted to a deliverable form; that is, either field segments for multileaf collimator (MLC) delivery or a form deliverable by a specific type of compensator. New beam deliverable doses are recalculated and the dose is then reoptimized using the same objectives as in the first stage, but instead of changing individual beamlet intensities, the beam weights of deliverable beam doses are modulated. This second round of optimization, global beam weight optimization, allows for remediation of some of the effects of delivery, such as those introduced by level quantization during field segmentation or introduced by compensator smoothing. The outputs of the second stage are field segments or compensator definitions and an estimate of the deliverable dose. (See Figure 3-1: XiO IMRT Inverse Planning.)

不改变单个beamlet的强度

In the third, optional stage, the optimizer is used to reweight individual segment weights using the current IMRT Rx and user defined segment parameters to remove extraneous segments and improve the dose distribution. The output of the third state is a decrease in the number of MLC segments and a deliverable dose distribution which better meets the desired optimized dose distribution.

Plan Setup (1)

In order to begin inverse planning, you must 1) select a patient; 2) contour anatomy structures requiring inverse planning objectives; 3) specify a number of usually non-opposed, isocentric beams that intersect

the target; 4) create objectives that represent your prescription and; 5) specify a treatment delivery mode (i.e., MLC or compensating filter).

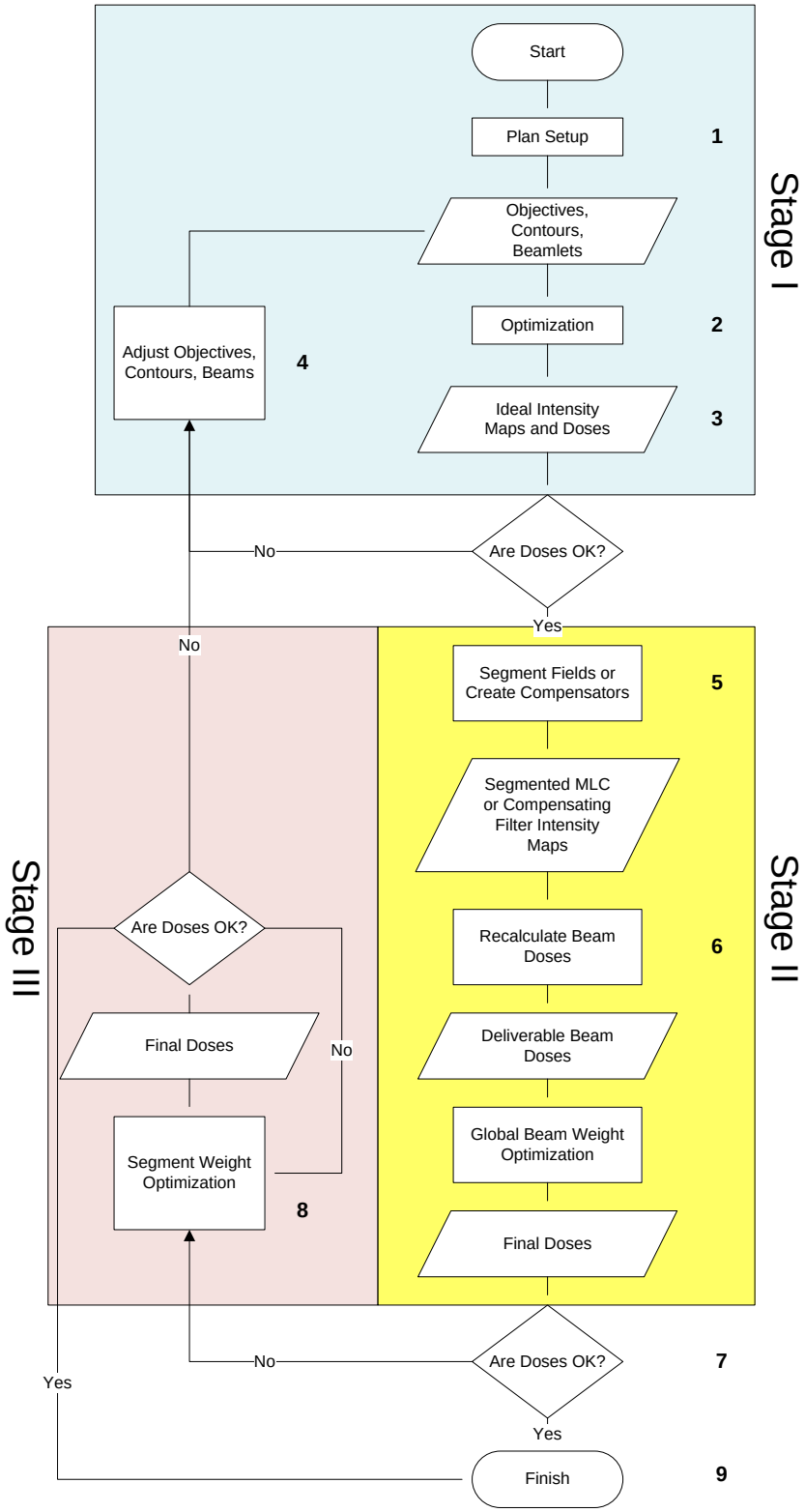


Figure 3-1: XiO IMRT Inverse Planning

Stage I Optimization (2)

The optimizer uses knowledge of the location of anatomy structures and the dose from the collection of weighted beamlets to minimize a cost function defined as the sum of inverse planning objectives. The output of Stage I is beamlet ideal intensity maps and ideal doses.

Ideal Dose Checking (3)

A basic rule of thumb is that the ideal doses from Stage I optimization must be acceptable in order to continue on to Stage II. Isodose displays as well as dose-volume histograms (DVH) and their accompanying dose statistics should be used to determine if doses are satisfactory.

Adjusting Objectives, Contours, and Beams (4)

If doses are unsatisfactory, you will need to change, in order of preference, objectives, beams, or anatomy contours. In addition, bolus can be helpful in situations where targets are drawn very near to the patient's skin. It may be beneficial, using isodose displays, to find locations in the patient where your dosimetric objectives are not being satisfied. Competition between organ at risk (OAR) objectives and target minimum dose objectives is inevitable when using unconstrained optimization methods and much of the effort that you will expend in inverse planning will be in determining which objectives are competing with each other and resolving these conflicts.

Segment Fields or Create Compensators (5)

This is the point

Ideal beamlet maps must be converted to a form deliverable by either MLC or compensators. During this conversion process, not all aspects of the ideal dose obtained during Stage I can be retained.

With MLC delivery, specifying more discrete levels of segmentation can lead to more faithful reproduction of ideal doses, but this also leads to longer treatment times on certain linear accelerators.

When the segmentation page is accessed the ideal maps will be discretized using the default number of intensity levels. The user should re-assess isodoses and DVHs at this point to see the effects of discretizing the map on the solution. The user may vary the number of intensity levels and discern what is the best tradeoff between the number of intensity levels and preserving the "ideal" optimized solution. Once the optimum number of intensity levels has been decided, the user should then proceed to produce segments.

With compensator delivery, limiting the effects of the smoothing parameter can also help to maintain ideal dose characteristics, but this may not be always possible due to compensator milling restrictions.

Recalculate Beam Doses (6)

After dose delivery issues have been resolved, new beam doses are calculated using either FFT Convolution, Superposition, or Fast Superposition. These dose calculations are highly accurate, accounting for fluence quantization effects for MLC delivery and fluence smoothing effects for compensators. For MLC delivery, the knowledge of leaf positions improves estimates of fluence for beam penumbrae, individual field segment penumbrae, and head scatter. Thus, many of the limiting approximations of the beamlet representation of the relationship between fluence and dose representation are eliminated.

Stage II Optimization (7)

第二部的优化是对射野剂量的优化，不是beamlets,最终的剂量是可以被执行的，而不是理想剂量。

The deliverable beam doses are then reoptimized in a process called global beam weight optimization, which is computed exactly like Stage I optimization, except that optimization is over beam doses rather than beamlets, smoothing is not applied, and the final result of the optimization is deliverable dose rather than ideal dose. Internally, the dose for each beam is converted to a single beamlet. If in Stage I optimization, there were perhaps five beams and more than 1000 beamlets, in Stage II optimization there would only be five beamlets, one for each beam (assuming there were no fields that required splitting). Each of these beam-sized beamlets include detailed modeling of the dose artifacts that are introduced by field segmentation of ideal beamlets or conversion of beamlets to a form deliverable by compensators.

Segment Weight Optimization (8)

子野权重优化

The deliverable segment weights can be optimized in a process called segment weight optimization, which is computed exactly like Stage II optimization, except that optimization is over segment weights rather than beamlets. Internally, the dose for each segment is calculated using the algorithm used for the full dose calculation, or if the Superposition algorithm was used, the Fast Superposition algorithm can be used for the segment dose calculations. It may not be possible to calculate dose for all segments in the plan if a weight point for a segment can't be located (Flash segments for example). The IMRT Rx can be edited and segment weight optimization can be repeated.

Final Dose Checking (9)

Final dose checking is much like ideal dose checking and uses the same dose analysis and visualization tools.

Anatomy Structures

All anatomy structures of interest used during inverse planning should be contoured. This includes structures that will have objectives as well as structures for which dose-volume histograms (DVH) will be calculated. Anatomy contours and anatomy rank are used to identify volume elements (voxels) claimed by different objectives. It is very important to contour anatomy carefully, because the XiO optimizer uses contour information exactly as it is specified, and achieving prescription goals may be difficult or impossible with inexact or inappropriate contours.

The next subsections describe two potentially problematic contouring issues for inverse planning. The nomenclature of the International Commission on Radiation Units and Measurements (ICRU) reports 50 and 62 is used in further discussions of patient anatomy.

Infinite Gradient Problems

Infinite gradient problems occur when contoured anatomy structures abut, or are in contact with each other, and have different dose objectives, as shown in Figure 11. Suppose that the mandible maximum dose objective is 54 Gy and that the planning target volume (PTV) minimum dose objective is 70 Gy. Because the mandible and PTV are in contact with each other, neither structure will completely achieve its dose objective, since that would require an instantaneous change in dose; that is, an infinite gradient.

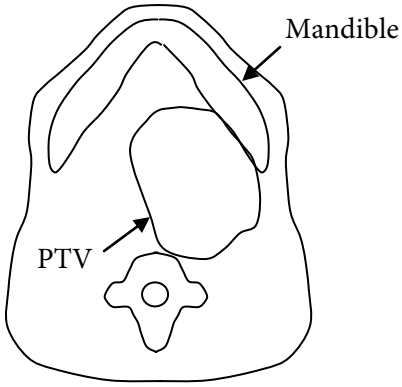


Figure 3-2: Anatomy Structure with Conflicting Dose Goals.

Figure 3-2 shows a schematic of structures and doses that further illustrates the abutment/infinite gradient problem. When structures are too close to each other, instead of a perfect step in dose, the optimizer will produce a physically-realizable gradient (as shown by the thick line in the figure) in which the normal tissue is minimally overdosed and the target is minimally underdosed. This behavior is not due to an idiosyncrasy of optimizers; rather, it is due to the continuous nature of dose distributions in tissue and the physical limits of dose modulation.

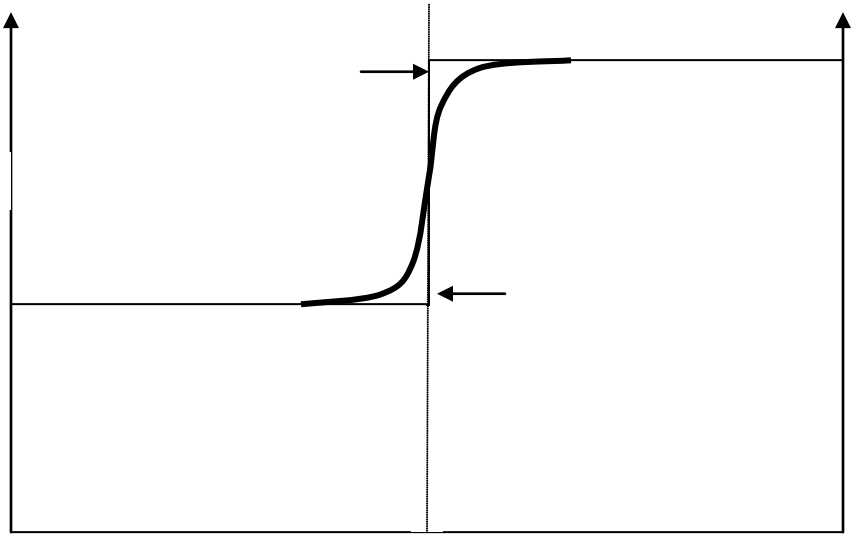


Figure 3-3: Infinite gradient problems

Figure 3-4 shows "improperly" and "properly" drawn contours of a gross tumor volume (GTV) and a PTV. Typically, in setting minimum dose objectives for a GTV and PTV combination, the minimum GTV dose is somewhat higher than the minimum PTV dose (*e.g.,* 66 Gy vs. 60 Gy), and 100% of the GTV is expected to get its specified minimum dose.

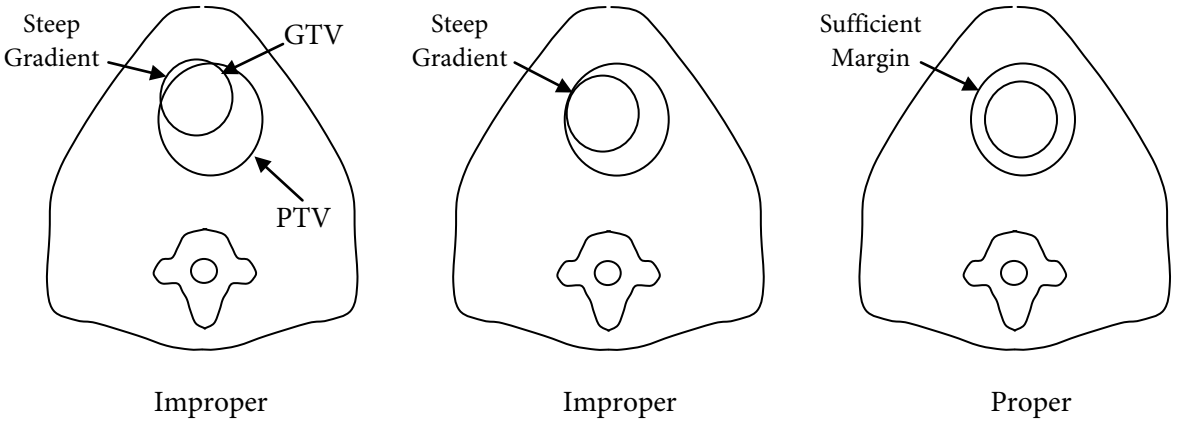


Figure 3-4: Overlapping contours

When there is a maximum dose objective for unspecified tissue, it may be set to the minimum dose objective for the PTV. This practice tends to alleviate dose gradient problems between unspecified tissue and PTV doses. When drawn properly, there is sufficient margin between the patient and GTV for dose to build up from the patient tissues through the PTV and finally into the GTV.

Dose Transition Problems

Figure 3-5A shows a contour of a target that is very close to the skin surface. The result of such a contour is that a portion of the target will be in the buildup region of any beam that directly enters the patient near the target (*e.g.,* beams 2 & 5). When a target is in the build up region of a beam, it is often difficult to achieve minimum dose objectives. Dose to normal tissues may be unnecessarily increased due to increased dose to the target that is required to make up for low dose at the skin surface. Alternatives are either to redraw the target away from the skin surface (which may not be clinically feasible) or to add bolus to the patient at problematic beam entry points.

当靶区贴近体表，靶区最小剂量目标很难满足

为了补偿在体表的靶区低剂量，在靠近正常组织的深度靶区剂量会更高。

shows a schematic of two idealized beam doses and their dose sum; it is meant to represent the situation in Figure 3-5A. The stippled region represents the target and the solid white region on the right represents a bolus.

The left of
shows that the beam enters the "patient" directly and that the build up region of the beam is in the target. On the right of the figure, the build up region is in the bolus.

Figure 3-5B shows another contouring problem – having low density holes or air gaps in a target outline. Ironically, this is only a problem when using homogeneity corrections with photon dose calculation algorithms that *correctly* estimate dose in low density regions (*e.g.,* superposition, Monte Carlo). Figure 3-7 represents such a dose decrease in a low density region.

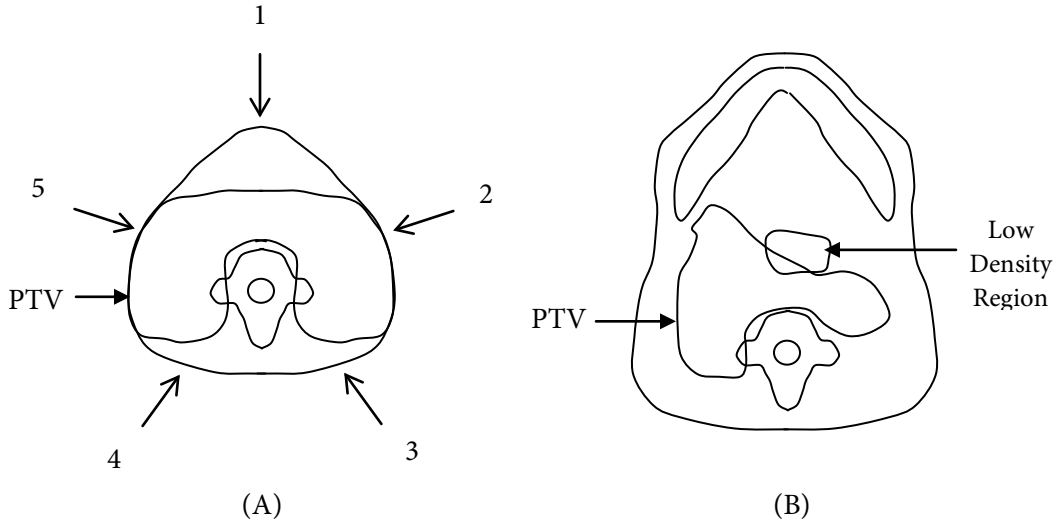


Figure 3-5: Dose build up and build down problems

A depression in the dose in a low density region may be interpreted by the optimizer as a failure to achieve a minimum dose to the target. Even though, in reality, the air gap is not part of the target, according to the contour, it is The low dose depression may again, as with targets in the build up region, make it difficult to achieve minimum dose objectives and indirectly increase dose to normal tissues. At this time, the only remedy for this problem is to be careful not to contour low density tissues that do not belong in tumors or to add special structures with higher rank that will cut "donut holes" out of larger target structures. These problems may be especially difficult to avoid when using the automargin expansion option[4].

4 In the future, an option to ignore dose to voxels that are less than some minimum Housfield number may be offered.

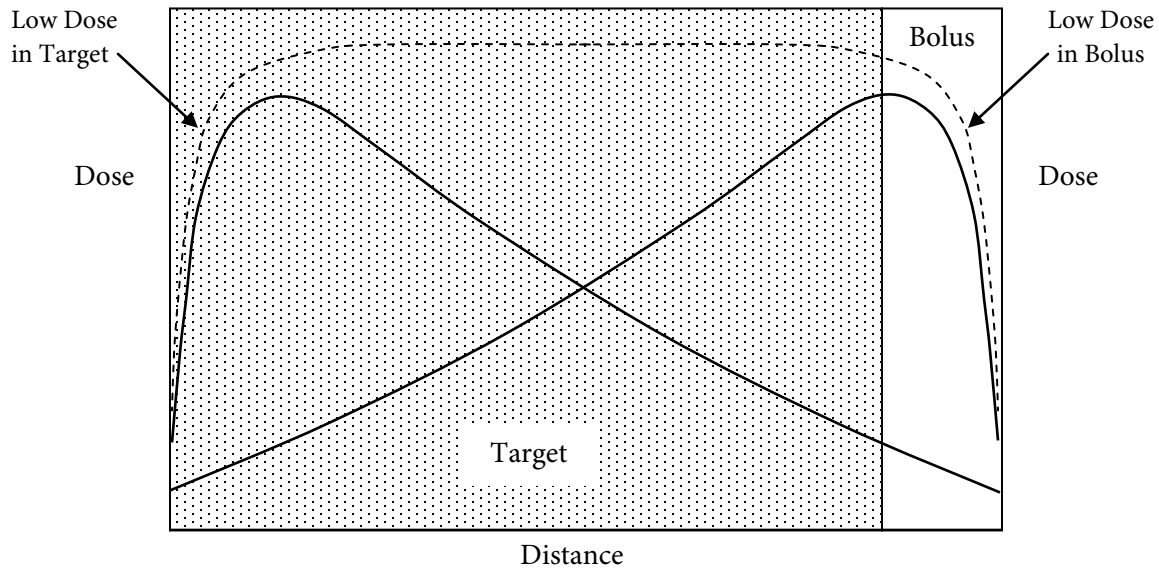


Figure 3-6: Build up at skin and in bolus

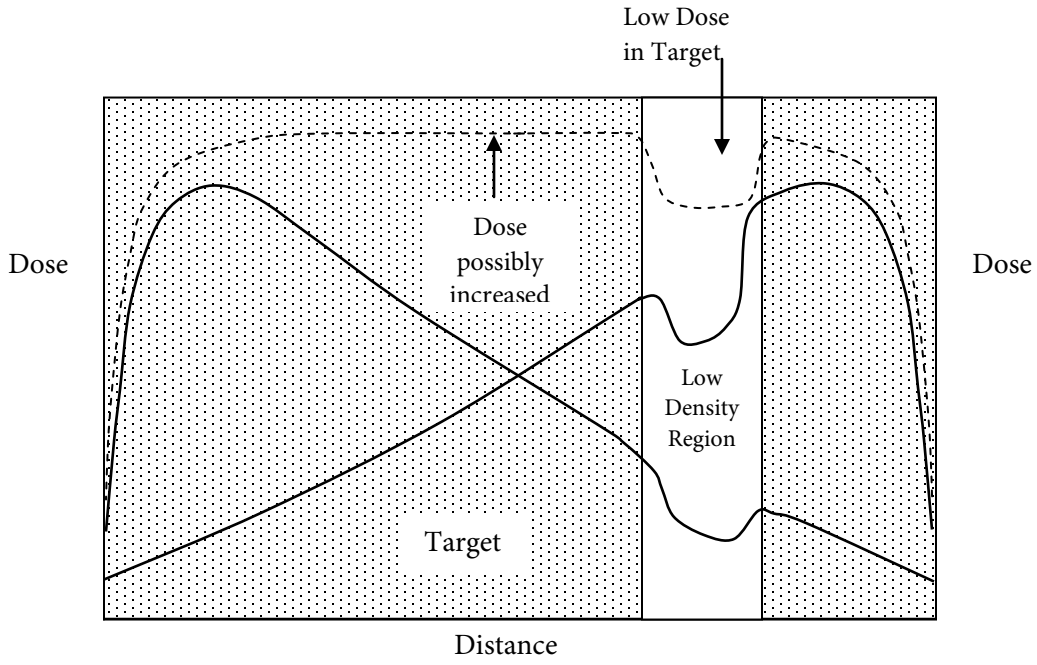


Figure 3-7: Low density region doses with superposition photon dose calculation

Beams

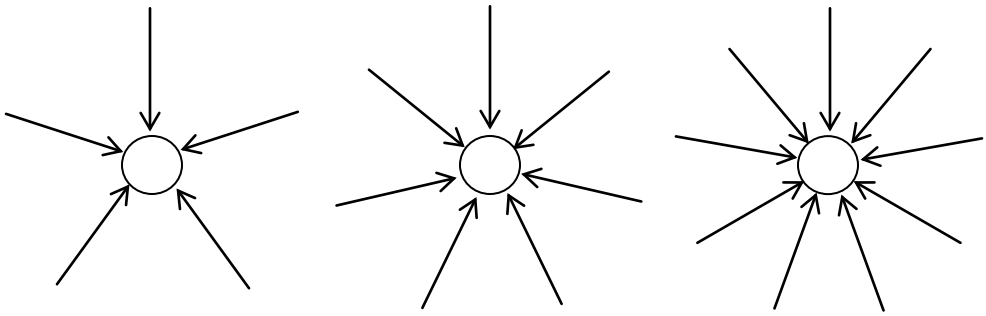


***Figure 3-8**: Beam arrangements for 5, 7, and 9 evenly-spaced beams*

In many cases, a small, odd number of evenly-angled coplanar beams will yield satisfactory results in inverse planning. At the low end, five beam plans have the advantage of few dose calculations, fast optimization[5] and relatively fast treatment times. As more and more beams are used, the possibility of finer dose nuances increases as well as dose calculation time, optimization time, and treatment time. Figure 3-8 shows evenly-angled beam arrangements for 5, 7 and 9 beams. There is little benefit in using more than nine beams in any IMRT plan [Pirzkall *et al.*, 2002].

Beam数量多，调制能力强，肿瘤适形度好，但治疗时间长；
Beam数量少，调制能力弱，适形度不够好，但治疗时间短。

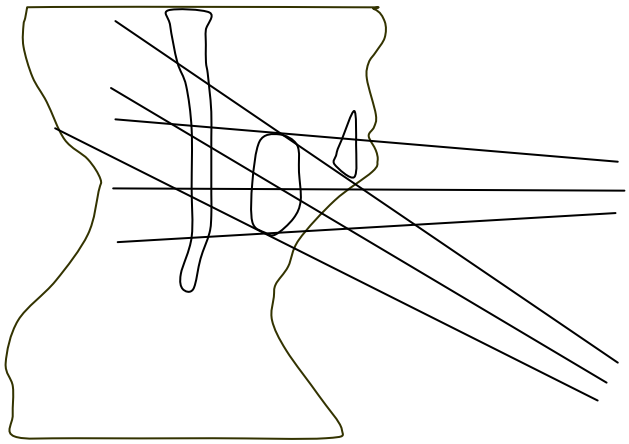


Figure 3-9: Shortening a beam path using a non-coplanar beam

5 The speed of optimization is linearly related to the number of beamlets and voxels represented by beamlets.

Non-coplanar beams may improve results in some instances. One obvious possibility is to use non-coplanar beams to entirely remove an OAR from a beam path. In the case of follow-up treatments this may be very useful if the physician's intention is to eliminate all new dose to a previously-treated OAR. However, in the general case, optimizers can achieve good results limiting dose to OARs, even if they are directly in the path of beams, and it is not necessary to intentionally place beams according to anatomy. Figure 3-9 shows a case in which using a non-coplanar beam may improve results. It shortens the beam path from the skin to a target, lengthens the path of the exit dose to the cord, and avoids the mandible, all of which lower dose to the cord.

Independent Beam Dose

Odd numbers of evenly-angled beams provide the optimizer with beamlets that have independent doses. By independent doses, we mean that by changing a specific beamlet's weight, the dose is changed in the patient in a way that is unique and perhaps not possible by any other combination of beamlet changes. This independence allows the optimizer many degrees of freedom in which to affect doses to targets and normal tissues when iteratively changing beamlet weights. The more that pairs of beams approach being parallel or directly opposed, the fewer the degrees of freedom for the optimizer to find a good solution to an inverse planning problem. A good rule of thumb is that beam angles should be at least 15 degrees apart for independence. Figure 3-10A**Error! Reference source not found.** shows an example of poor beam placement for inverse planning that is very acceptable and, in fact, the norm in many clinics for conventional prostate treatments. The problem with this beam placement for IMRT is that the parallel opposed beams (1 & 3 and 2 & 4) affect doses in very much the same way. The five beam arrangement in **Error! Reference source not found.**B is more typical in inverse planning for IMRT.

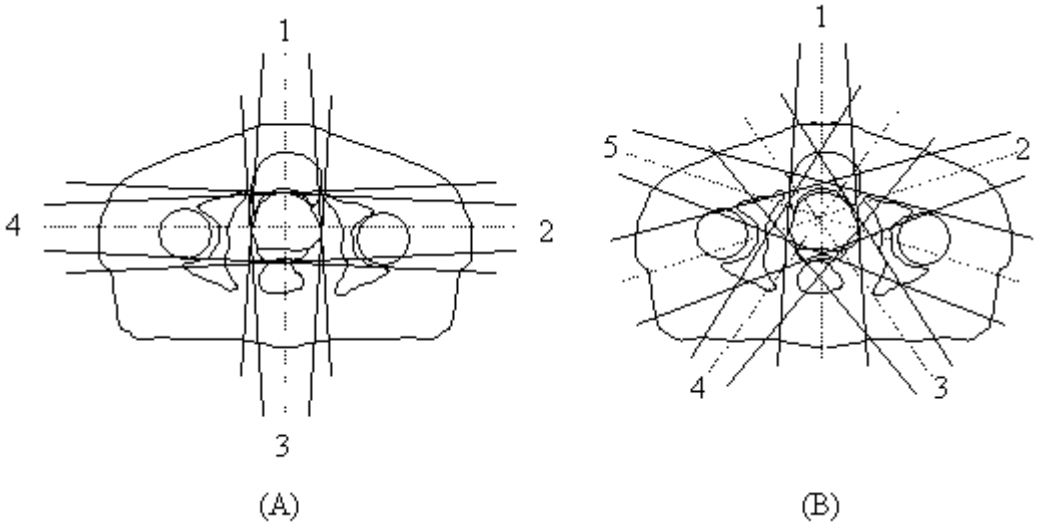


Figure 3-10: Poor beam angles and good beam angles for inverse planning

Figure 3-11 shows two beam arrangements that may make the concept of beamlet dose dependence and independence clearer. Suppose that the beams shown in the figure are subdivided into beamlets. On the left of the figure, it is clear that the doses at the top and bottom of the target are tightly coupled to doses within the target. On the right of the figure, doses at the top and bottom of the target are decoupled from doses on the left and right of the target by virtue of the 90^0 separation of beam angles. From Figure 3-11B

it is obvious that adding more non-parallel beams will have the effect of creating more ways for the optimizer to modulate dose.

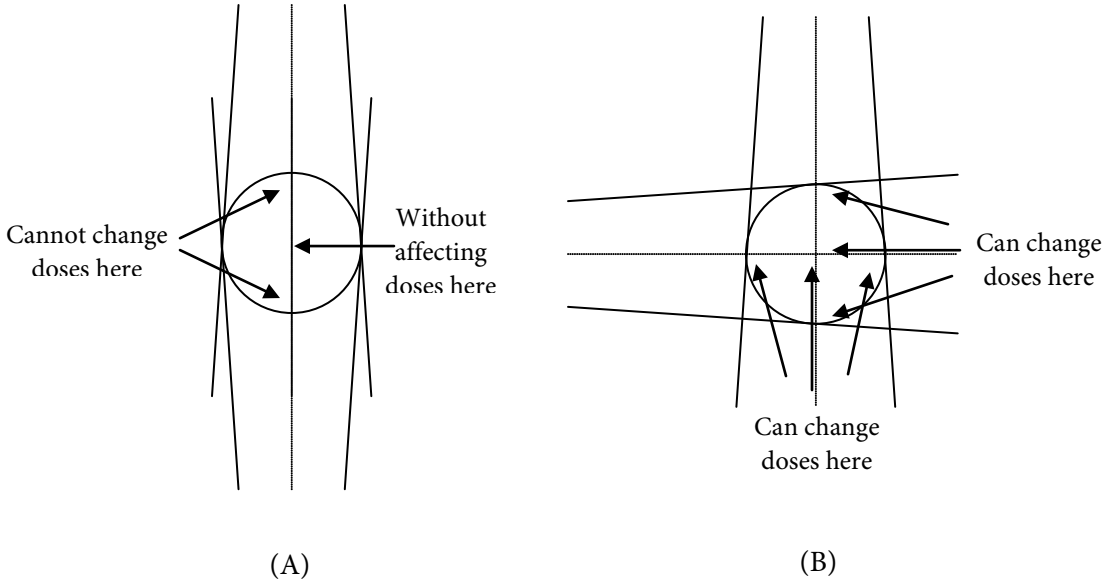


Figure 3-11: Target doses for dependent and independent beamlet doses

Excessive Beam Paths

Even though evenly-angled beam arrangements are often preferred, as long as beams are separated by at least 15 degrees, good results are possible. One situation where non-evenly-angled beams may be useful is when a beam in a standard configuration traverses an excess amount of normal tissue before it encounters the target. Figure 3-12 shows a seven beam arrangement in which the two anterior oblique beams are traversing too much normal tissue. Without a maximum dose objective to keep dose to patient unspecified tissues under control, hot spots may occur in the shoulders.

没有偶数个野是有效的，当一个射野穿过很多的正常组织在到达靶区之前

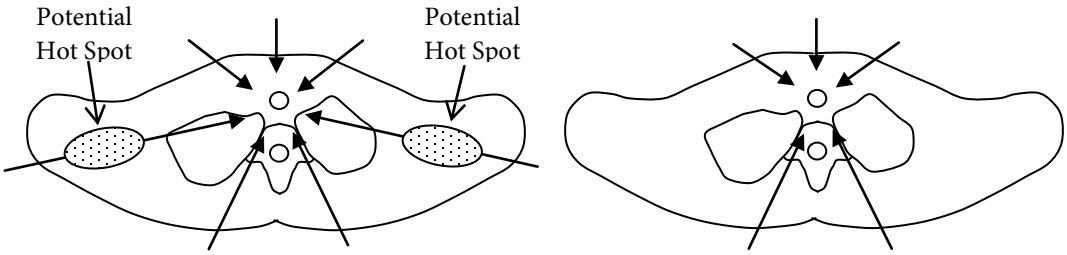


Figure 3-12: Seven field evenly-angled plan with beams traversing too much tissue and five field improved beam angle selection (0, 55, 154, 206, 305 degrees).

Dose Calculations and Beamlets

XiO IMRT requires the use of the FFT Convolution, Superposition or Fast Superposition photon dose calculation algorithms. This is due to two factors: 1) Measurement-based algorithms like "Clarkson" are not accurate enough for IMRT and therefore not supported for use by XiO IMRT and 2) The photon pencil beam dose calculation algorithm used by XiO IMRT to create "beamlets" uses any one of the above-mentioned three-dimensional photon dose algorithms to "calibrate" the beamlet doses – the choice of algorithm is up to the user. Calibration of beamlet doses with three-dimensional photon algorithm doses makes XiO beamlet doses much more accurate than beamlets used in other commercial treatment planning systems. This is especially true when a superposition algorithm is used in heterogeneous tissues. In this case, beamlet doses mimic build up and build down associated with changing patient densities. See *Pencil Beam Algorithm*.

In XiO IMRT, dose calculations are done a number of times during inverse planning:

When a new beam is specified, the dose for that beam is calculated based on the beam parameters and calculation volume specification.

When the Stage I optimization begins, beamlets are calculated using the pencil beam photon dose calculation algorithm.

During Stage I optimization, beamlet weights are iteratively changed representing changing dose in the patient and resulting in ideal doses.

When Stage II optimization begins, ideal beamlet intensities are segmented for delivery by multileaf collimator (MLC) *or* converted to a form deliverable by compensators. A final deliverable dose is calculated for each beam.

Concluding Stage II optimization, beam weights are reoptimized using the deliverable beam dose for each beam and the same optimization method and objectives as in Stage I.

Doses are calculated for individual segments during segment weight optimization.

XiO IMRT has two different types of beamlets: beamlets whose weights are actively changed by the optimizer during line searches and beamlets whose weights are passively changed after each line search. Active beamlets generally contribute dose to targets and are not blocked by MLC leaves or other treatment aids. Passive beamlets tend to represent leakage dose for regions covered by MLC leaves or blocks and are also used to model wedge doses.

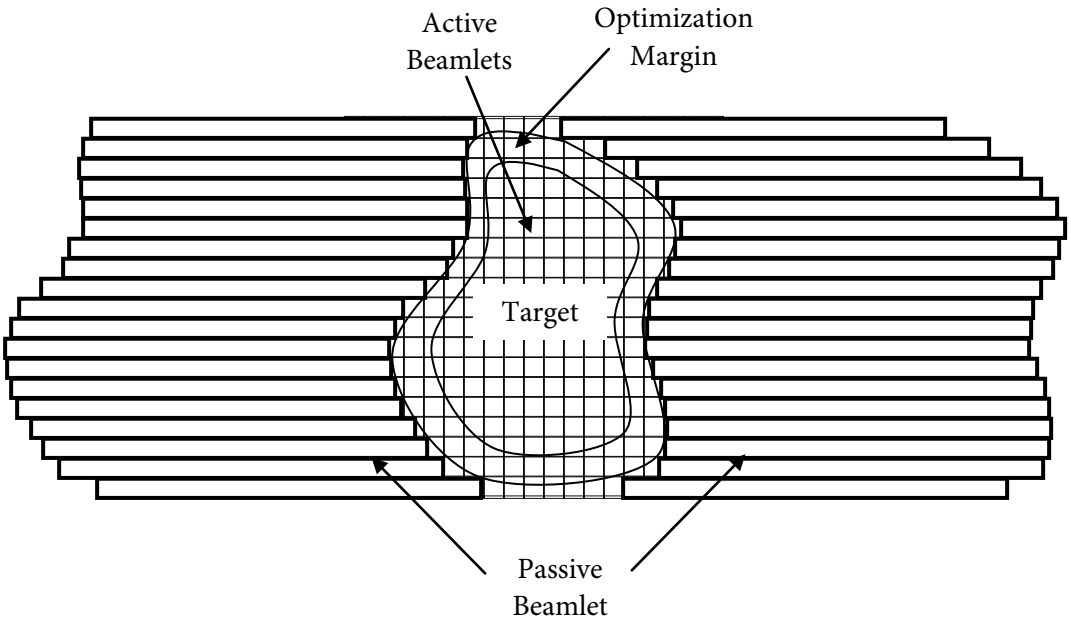


Figure 3-13: Beamlets with zero insertion showing an optimization margin

Figure 3-13 shows that XiO allows for specification of an "optimization margin" which is an additional region around a target that will not be covered by MLCs – this is to allow a rim of active beamlets outside of a target without the need to specify a separate target expansion structure.

Objectives

In the following discussion, an intuitive description of XiO's inverse planning objectives is developed. The goal is to be able to apply this understanding logically to the practice of inverse planning. For a mathematical description of the XiO objectives, refer to *Cost Function Mathematics*.

Specifying Objectives

Table 3-1 represents the XiO IMRT prescription page, which is where inverse planning objectives are specified. Imagine a prescription which requires the dose to a target to be between 65 and 75 Gy and the dose to a single organ at risk (OAR) to be less than 45 Gy[6]. This prescription can be realized in terms of the objectives in Table 3-1.

Structure	Type	Rank	Objective	Dose (Gy)	Volume (%)	Weight	Power	Status
PTV	Target	1	Minimum	65	100	100	2.0	On
			Goal	70	100	---	1.0	On
			Maximum	75	0	100	2.0	On

6 We chose this prescription for instructional purposes, because it is relatively easy to show graphically. In routine practice, acceptable target dose limits will be much closer typically resulting in target dose homogeneity in the 5% range or better.

Cord	OAR	2	Threshold	0	100	---	1.0	On
			Maximum	45	0	100	2.0	On
Patient	OAR	3						

Table 3-1: The XiO IMRT prescription page for a simplified head and neck case

Briefly,

- The Structure column lists all anatomy structures that have been contoured, including the patient exterior. Any contoured anatomy structure can have objectives.
- The Type column identifies a structure as either a target or an OAR. Targets can have minimum, maximum, and goal dose objectives, while OARs can have threshold, maximum and dose-volume (DVH) objectives.
- The Rank column specifies how voxels are shared when anatomy contours overlap. Typically, targets have a lower numerical rank than OARs. The patient exterior should have the highest numerical rank, thus excluding all voxels from the patient that belong to other anatomic structures with active objectives. Note that only active objectives are considered when rank is concerned; thus, what is considered unspecified tissue will be a function of the contoured anatomy as well as the active objectives for the contoured anatomy.
    - A structure with a lower numerical rank than another structure will "own" dose voxels in regions where the two structures overlap. Therefore, any objective for the lower numerically-ranked structure will be affected by doses in the overlap region. The higher numerically-ranked structure will not be affected by doses in the overlap region.
    - When two overlapping structures have the same rank, dose voxels in regions where the two structures overlap are shared. As a result, objectives for both structures are affected by the doses in the overlap region.
- The Objective column contains two different types of information: the type of objective and its goal or threshold dose.
- The Dose column specifies doses associated with the objectives.
- The Weight column specifies the relative importance weight for the objective; that is, its relative importance with respect to other objectives.
- The Power column specifies the magnitude of the penalty that will be applied to dose outside acceptable extreme dose ranges for objectives.
- The Status column indicates whether an objective is "on" or "off." This convenient feature allows you test to the effect of an objective, turning it on and

off without deleting it or having to re-enter it. This feature may be especially helpful when you have are using IMRT templates with objectives.

We now discuss objectives and their dose, volume, weight, and power parameters in more detail.

Objective Types

XiO IMRT supports three different types of inverse planning objectives:

1. minimum dose objectives
2. maximum dose objectives
3. dose-volume objectives (called DVH objectives on the IMRT prescription page).

Goal doses pertain to target objectives and threshold doses pertain to OAR objectives[7].

Dose Objectives

We discuss minimum and maximum dose objectives first, because they are fairly simple, and because dose-volume objectives build on concepts in maximum dose objectives.

Figure 3-14 shows cost per voxel as a function of dose for the objectives in Table 3-1. This graph is similar to the dose voxel cost graphs shown in *Optimization – Basic Concepts*. The difference is that the cost per voxel functions in Figure 3-14 are the actual ones that XiO would use given the specified objectives. The objective functions in Chapter 0 were simply introduced to illustrate optimization concepts. The graph shows for the OAR maximum dose objective that between 0 Gy (the threshold dose) and 45 Gy (the maximum dose) the cost per voxel increases slowly and linearly. Past 45 Gy, the increase in cost is much faster, and in fact is "quadratic", as indicated by the power 2.0 in the table. Similarly, the cost per voxel for the target maximum dose objective increases linearly between the goal dose (70 Gy) and its maximum dose (75 Gy); the increase is again quadratic past 75 Gy. For the target minimum dose objective there is a linear section between the goal dose (70 Gy) and its minimum dose (65 Gy); any outlying dose less than 65 Gy is again penalized quadratically.

All XiO objectives, including dose-volume objectives, treat dose values in different dose regions differently. Dose outside of extreme acceptable dose limits (*i.e.,* minimum and maximum doses) is penalized more strongly than dose that is within these limits. Thus, during optimization, the optimizer will concentrate on dose points whose values are in higher penalty regions and try to change their values until the dose at these points are in lower penalty regions. Once in lower penalty regions, dose values can still improve but the impetus to improve is much smaller. Thus, getting all dose values for a structure inside of its acceptable extreme dose limits is primary, yet residual improvement is possible once all dose is within acceptable dose limits.

7 Goal doses and threshold doses objectives actually help to define objectives but are not by themselves objectives. See Cost Function Mathematics.

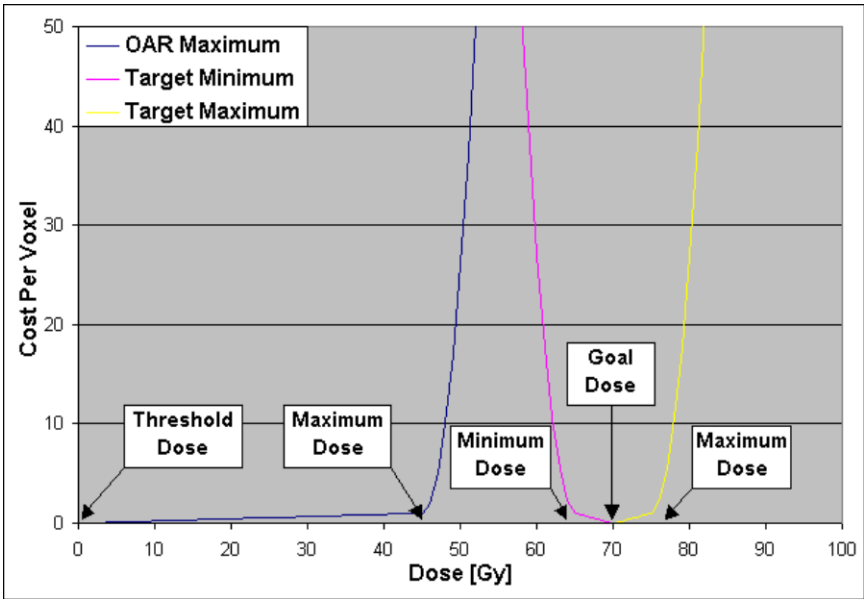


Figure 3-14: Minimum and maximum dose objectives from Table 7

XiO objectives are calculated as the average of the summed costs for all dose voxels claimed by an anatomic structure. Because average cost is used, objectives are not favored by the number of voxels that their structures claim. That is, the relative size of anatomic structures is factored out of the objective cost calculation. This is appropriate, because minimum target dose objectives and maximum dose OAR objectives will tend to compete against each other, as shown in Figure 3-15. It would not necessarily be advantageous for a big structure to be favored over a small structure in the competition to raise target doses and lower normal tissue doses. Note that dose-volume objectives may also compete against target minimum dose objectives, since their goal is also to lower dose in normal tissues.

Also shown in Figure 3-15, the target minimum and maximum dose objectives may, in some sense, cooperate with each other. The upward pull on target doses from a minimum dose objective that is balanced by a downward pull from a maximum dose objective can lead to relatively homogeneous target dose. It is good practice to specify appropriate target minimum and maximum dose objectives first before specifying any normal tissue objectives. This will let you determine the "best" homogeneous dose to target that can be obtained given the patient geometry and the beams specified. As you add normal tissue objectives, target dose homogeneity will degrade. This is what is supposed to happen, because the normal tissue objectives will be competing with minimum target objectives. Note that in cases where the acceptable range of target doses is too small, target minimum and maximum dose objectives will also compete against each other, as will later be shown.

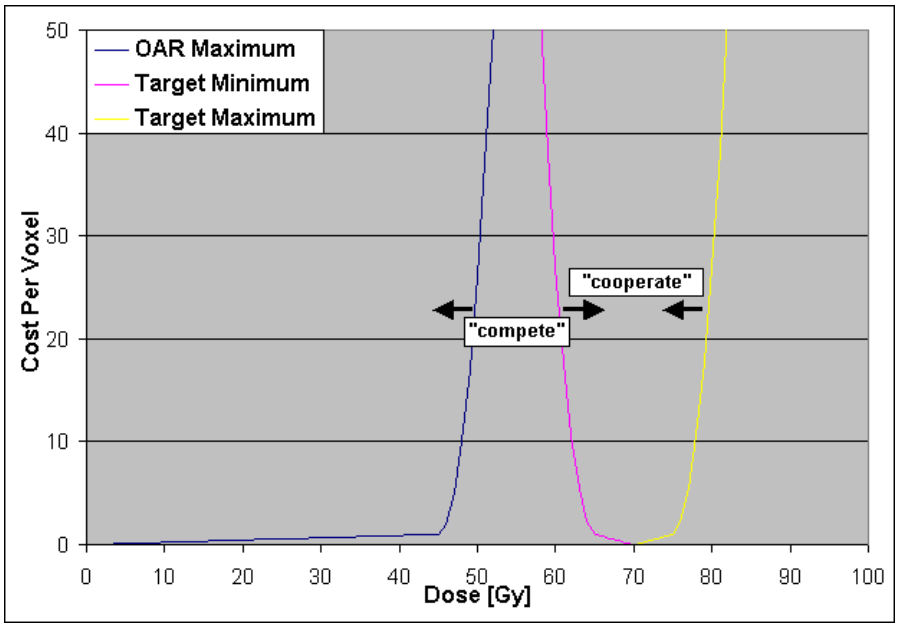


Figure 3-15: Competing and cooperating objectives

Adjusting Objectives

In XiO IMRT, there are several ways to "fix" unsatisfactory inverse planning dose distributions by adjusting objective parameters[8]. These include changing objective importance weights, powers, and dose specifications. Experience will guide the user to determine situations in which changing importance weights, penalty powers, or reference doses might be preferable.

Adjusting Importance Weights

Changing importance weights of objectives is the simplest way to affect the distribution of optimized doses. The next few figures are used to facilitate a discussion of the effect of changing importance weights and the dose distributions that might result from specific changes. The figures are meant to *qualitatively* represent dose distributions after convergence of inverse planning optimization. For the purpose of enhancing the discussion of the mechanics of the objectives, the examples are made to appear more problematic than might normally occur in daily clinical practice.

Figure 3-16 shows what a final dose distribution for the objectives in Table 3-1 might look like if the target minimum and target maximum dose objectives were activated. The results show that the target dose is, for the most, part bounded by the extreme acceptable objective values of 65 and 75 Gy. The tails marked

8 Before adjusting objective parameters, it is helpful to locate the problematic doses in the patient. This may reveal problems in how you may have contoured anatomy structures or specified photon beams, which may need to be fixed first.

"Dose Low" and "Dose High" are meant to indicate that, for some physical reason it was not possible to get all of the target dose inside the acceptable dose region[9].

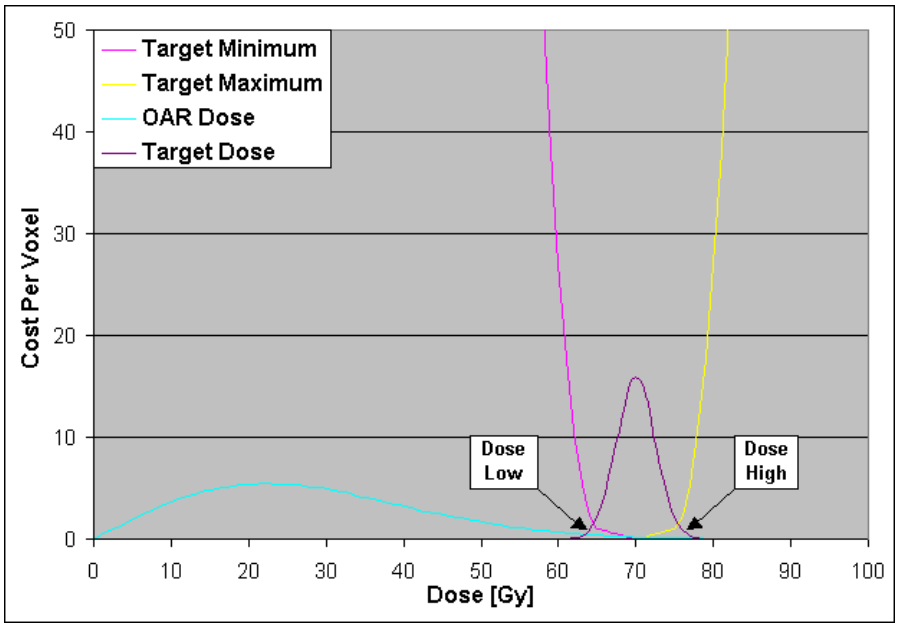


Figure 3-16: Target objectives only with dose distributions

Notice that the length of the tails outside of the acceptable region on both the low and the high side is just about the same. If you were analyzing these results, you would correctly conclude that this is because the minimum and maximum target dose objectives are competing against each other. That is, the minimum dose objective is trying to raise all of the target dose above 65 Gy and the maximum dose objective is trying to lower all of the target dose below 75 Gy. The reason for such a conflict can be a combination of:

1. An unreasonably restrictive target dose range;
2. One or more especially long beamlet paths through the target;
3. The use of low-energy beams.

This sort of even-handed compromise between competing objectives is typical of unconstrained conjugate gradient optimization, the method of optimization that the XiO optimizer uses.

Figure 3-17 shows what the dose distribution might look like if both target and OAR objectives were activated. Not surprisingly, the OAR dose is improved somewhat and the target dose is degraded. The OAR maximum dose objective is competing with the target minimum dose objective. None of the objectives is satisfied.

---

9 In reality, for a 10 Gy acceptable target dose range it would be unlikely that all optimized target dose would not end up in the acceptable range. The inability to get target dose within an acceptable range only occurs in practice with much smaller acceptable target dose ranges.

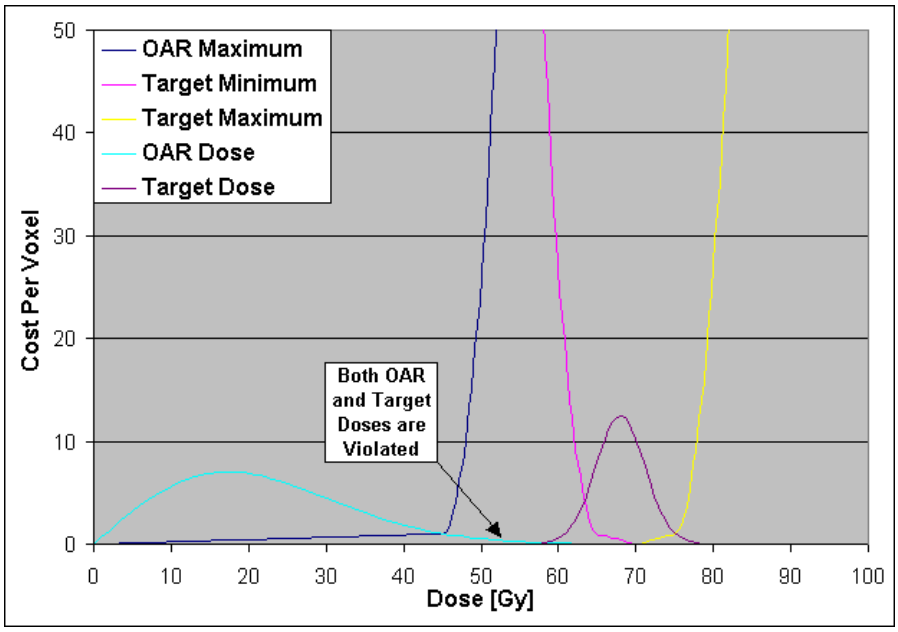


Figure 3-17: Target and OAR objectives with dose distributions

Suppose that we wanted to have a better chance at satisfying the OAR maximum dose objective. Figure 3-18 shows changes in cost per voxel and the dose distributions that might occur if the OAR objective's importance weight was raised from 100 to 500. Notice that the rate of increase for both the linear and the higher power sections of the OAR objective is affected by a factor of 5. As expected, the dose distributions would be pulled lower in order to better satisfy the OAR maximum dose objective. And, since the OAR maximum dose objective is competing with the target maximum dose objective, some fraction of the target dose must be also be lowered.

Figure 3-18 shows that the OAR maximum dose objective is nearly satisfied and the target dose minimum dose objective is much more violated. A good question to ask is how high the OAR objective's importance weight must be adjusted before the OAR objective is completely satisfied. There is actually no way of definitively saying – the user must experiment. This will always be the case with optimizers that use unconstrained optimization methods.

In Figure 3-18, the objectives are most strongly competing for the dose points somewhere between 45 and 65 Gy. If we were planning a head and neck case, Figure 3-19A might show a transverse slice that could represent where dosage compromises might be taking place. Notice that each beam has the cord directly in its path. That is not necessarily a problem, because the optimizer will tend to block dose to the cord by lowering the weights of beamlets that intersect the cord. This will result in the appearance of a trough in beamlet intensity maps for each beam. *See* Section Evaluating Results."

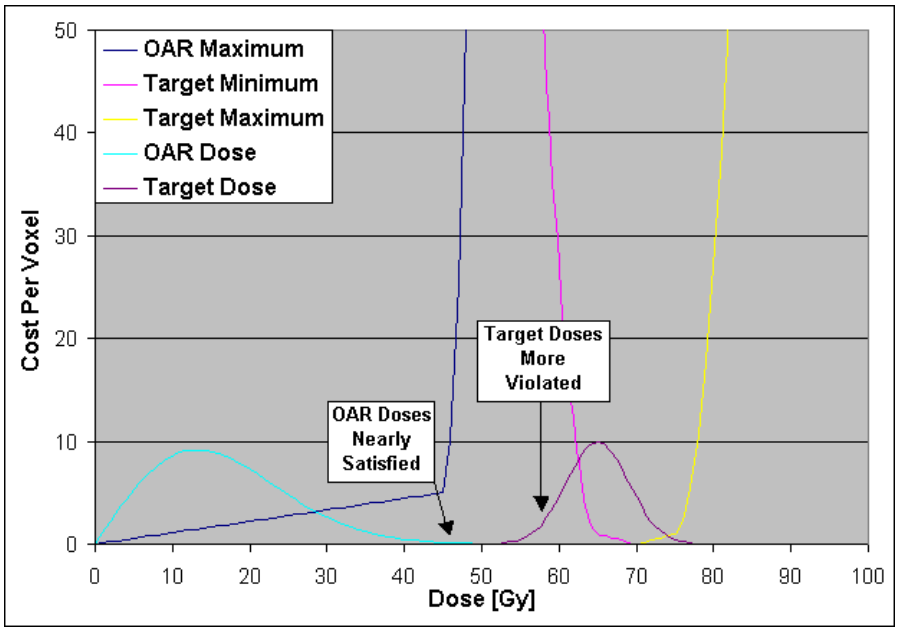


Figure 3-18: Organ at risk maximum dose objective, importance weight = 5, with dose distributions

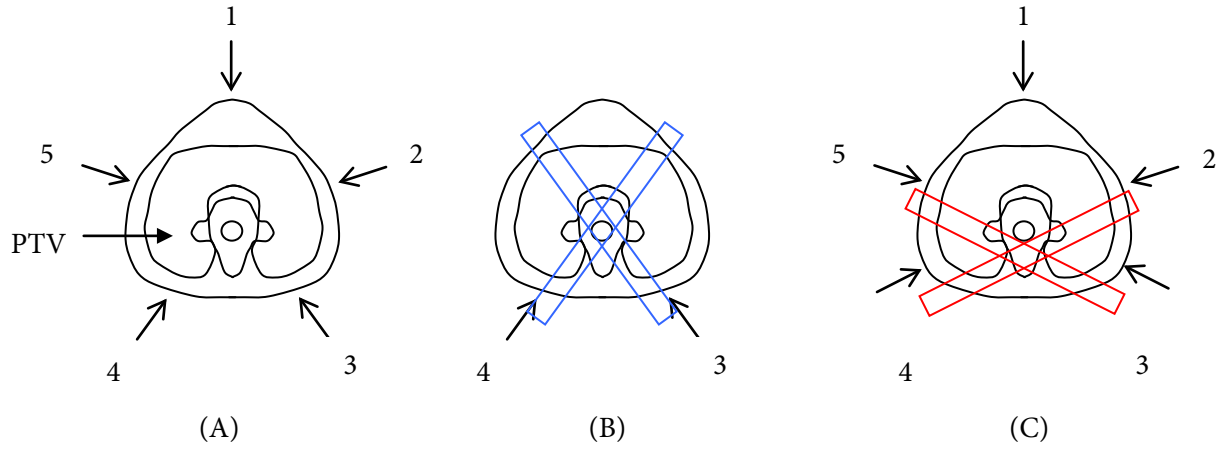


Figure 3-19: Beam angles and competing objectives

The actual problem could be in depth of dose trough required for beams 3 and 4 to (nearly) satisfy the OAR maximum dose objective. An overly-deep trough might lead to a lower than acceptable dose to the posterior aspect of the PTV. A reason to suspect this is that the length of the beam paths for beams 1, 2 and 5 to the posterior aspect of the PTV is quite long. For beams 1, 2 and 5 to deliver dose to the posterior aspect of the PTV sufficiently might require dose to the target to increase above its maximum target dose objective. Thus, the maximum target dose objective could make it the task of beams 3 and 4 to deliver dose to the posterior aspect of the PTV.

The blue bars shown in Figure 3-19B represent the effect of potential troughs in the beamlet intensities for beams 3 and 4. Beams 1, 2 and 5 should be able to make up the dose deficit towards the anterior of the PTV. Figure 3-19C shows an alternative beam arrangement that might help to improve the dose distribution. The red bars show how the beamlet intensities and dose to the posterior aspect of the PTV can be increased in the areas that were previously underdosed. Notice that with the change, doses from beams 3 and 5 are less independent, as well as doses between beams 2 and 4. This might cause other problems. Another possible fix might be to increase the number of beams from 5 to 7 beams, or to increase the dose for the target's maximum dose objective. This discussion underscores the iterative and experimental nature of inverse planning.

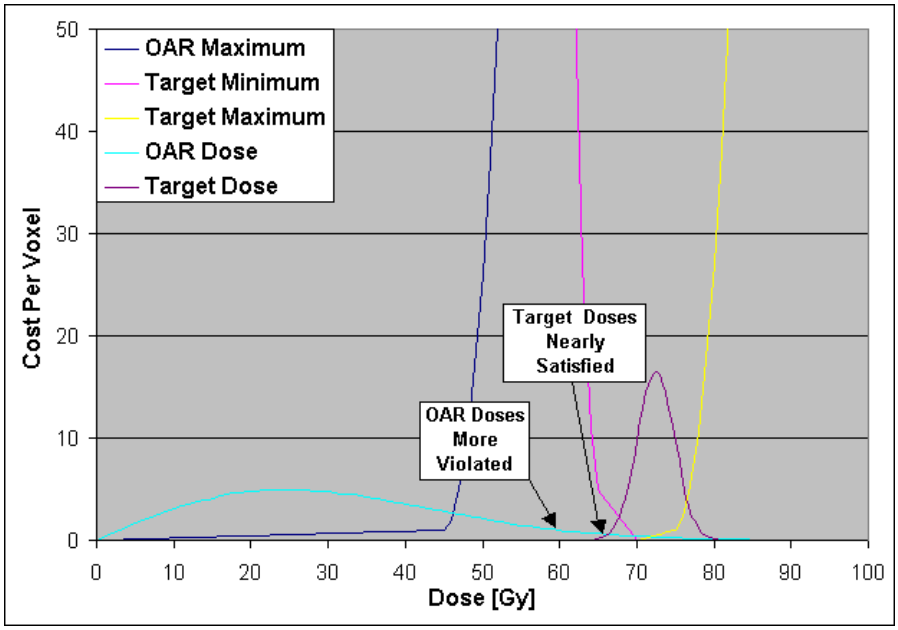


Figure 3-20: Target minimum dose objective importance weight = 5 with dose distributions

Getting back to our example, suppose that we wanted to satisfy the target minimum dose objective. Figure 3-20 shows changes in cost per voxel and the dose distributions that might occur if the OAR maximum dose objective's importance weight was returned to 100 and the target minimum dose objective's importance weight was increased to 500. Notice again that the rate of increase for both the linear and the higher power sections of the target objective are affected by a factor of 5. As expected, the dose distributions would be pulled higher in order to better satisfy the target minimum dose objective.

Adjusting Powers

XiO IMRT supports penalty powers between 2.0 and 5.0. Figure 3-21 shows that adjusting the power of an objective affects how it penalizes dose in the higher order penalty region of the cost per voxel function. Penalty powers do not affect cost per voxel in the linear penalty region. This may be helpful when it is more important to get doses into acceptable bounds than to improve doses once they are inside the bounds.

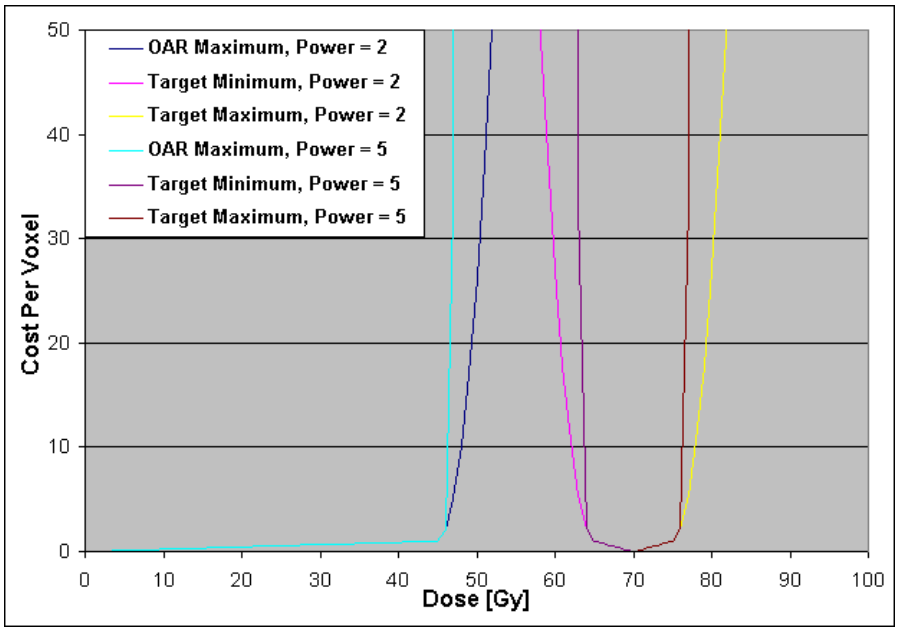


Figure 3-21: Comparison of objectives with power = 2.0 and power = 5.0

One of the best uses of higher-order penalties is to create objectives that effectively behave as barrier functions. However, caution is required because the use of higher-order penalties can cause individual objective evaluations to oscillate somewhat during optimization, which will result in needing more iterations to converge to a result. Also, since initial cost function evaluations may be much higher than with quadratic penalties, the convergence criterion may need to be reduced.

Adjusting Doses

Of the three alternatives for tailoring dose distributions, changing the dose specification is the least pleasing, because it will require you to use different doses in your objectives than the prescription specifies. Figure 3-22 shows the effect on cost per voxel of setting the maximum dose objective acceptable dose range to be no greater than 40 Gy.

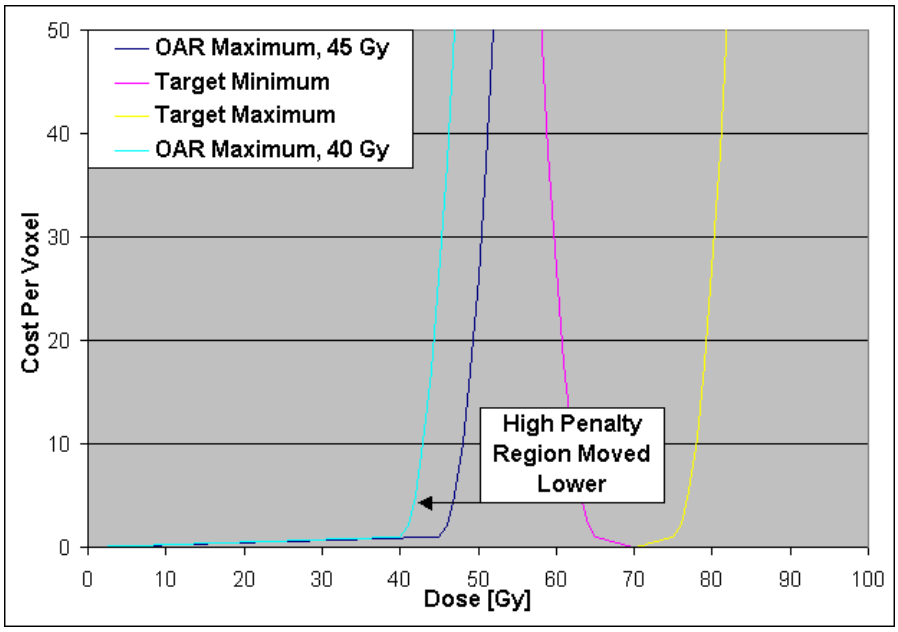


Figure 3-22: Organ at risk maximum dose objectives with maximum doses of 40 and 45 Gy

Dose-Volume Objectives

The XiO IMRT dose-volume objectives were adapted from a well-known formulation [Bortfeld *et al.*, 1997]. In XiO IMRT, dose-volume objectives specify that no more than a certain percentage of an OAR should get more than a certain dose. Table 3-2 shows actual objectives for a prostate case that achieved good results that includes a dose-volume objective for the rectum. Notice that there are no active objectives for many of the other structures. Differences between prescriptions and objectives are discussed in Section 3.7, "Prescriptions and Objectives."

Structure	Type	Rank	Objective	Dose (Gy)	Volume (%)	Weight	Power	Status
PTV2	Target	1	Minimum	64.2	100	100	2.0	On
			Maximum	78.0	0	100	2.0	On
GTV	Target	1						
PTV1	Target	2	Minimum	78.0	100	100	2.0	On
			Maximum	79.0	0	100	2.0	On
SemVes	Target	2						
Rectum	OAR	3	DVH	40	35	100	2.0	On
Bladder	OAR	3						
RtFemur	OAR	3						
LtFemur	OAR	3						
RtPelvis	OAR	3						
LtPelvis	OAR	3						
Patient	OAR	4						

Table 3-2: The XiO IMRT prescription page for a prostate case

As previously mentioned, dose-volume objectives have a lot in common with maximum dose objectives. Figure 3-23 shows a dose-volume histogram (DVH) of the initial open field rectal dose before optimization for a patient. The labels identify important regions related to the rectal dose-volume objective in Table 3-2. The rectal dose-volume objective specifies that no more than 35% (reference volume) of the rectum should get more than 40 Gy (reference dose).

The difference between dose-volume objectives and maximum dose objectives is that the penalty functions for dose-volume objectives are separated into three regions and that the second two regions change as the optimizer minimizes the cost function. Figure 3-24 shows the cost per voxel for the dose-volume objective for the initial rectal dose. The first linear region, from 0 Gy to 40 Gy, does not change, but the higher order penalty region and second linear penalty region does change. The first linear region and the higher order penalty region operate basically as in maximum dose objectives. Because the major penalty for the objective is between the reference dose and the dose where the reference volume is satisfied, the dose where the reference volume is satisfied will approach the reference dose during the course of optimization. Figure 3-25 shows what the cost per voxel function would be, if 35% of the rectal volume exceeded 50 Gy, as opposed to approximately 65 Gy originally.

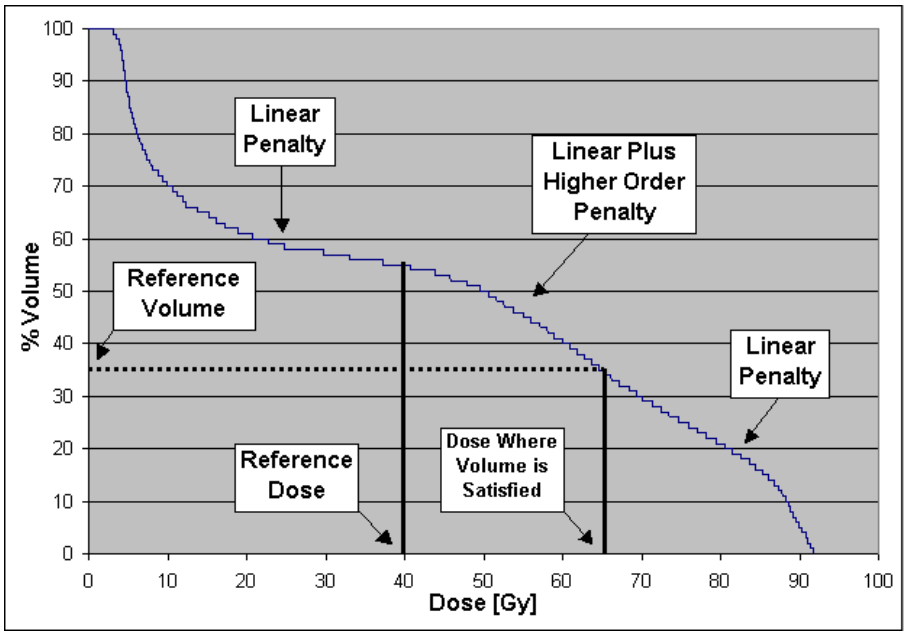


Figure 3-23: Dose penalty regions for a dose-volume objective

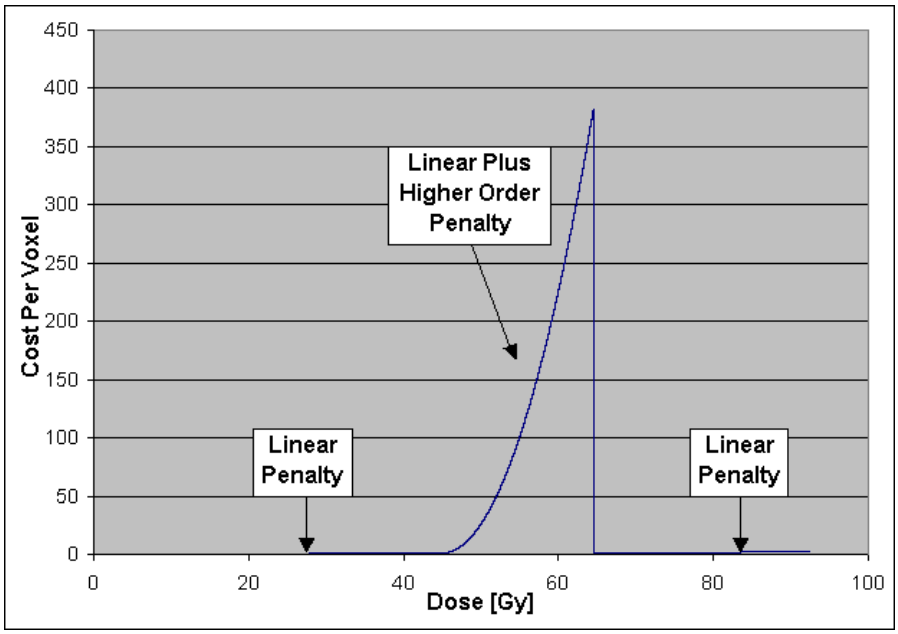


Figure 3-24: Dose-volume objective cost per voxel vs. dose for Figure 3-23

Figure 3-25: Dose-volume objective cost per voxel vs. dose for 35% of the volume exceeding 50 Gy

The previous discussions about minimum and maximum dose objective adjustment apply to dose-volume objectives. Additionally, dose-volume objectives, by definition, have a volume parameter. The volume parameter of dose-volume objectives should be changed as a last resort, when trying to achieve some dose objective.

Effective Cost Functions

The success of an inverse planning system depends to a large extent on offering a cost function that effectively represents clinical concerns and that a user can intuitively regulate [Ezell *et al.,* 2003]. Practically, in unconstrained optimization, intuitive regulation of cost functions requires that importance weights be meaningful; that is, they should only control dose distribution tradeoffs between objectives. It will be shown that importance weights are not meaningful when they are required to account for objective function shape problems, such as costs and cost derivatives that are ill-suited for representing prescription requirements.

Ideally, the role of importance weights should be minimal, given uncomplicated patient geometry and reasonable prescriptions. In XiO IMRT, good results may often be obtained with importance weights that are identical for all objectives [Wiesmeyer, 2004]. If you have used inverse planning systems before, this will probably seem impossible – good results without importance weights playing much of a role! But as you use XiO IMRT, you will see that this can be the case.

Cost Function Formulation

Importance weights can be "ineffective" in the context of unconstrained optimization and ill-suited cost function objectives. Recall the prescription and objectives used in *Optimization – Basic Concepts*.

The example had a target and an organ at risk (OAR). The goal dose to target, D_{GT}, was 64 Gy and the goal dose to the OAR, D_{GN}, was 0 Gy. The cost function was

$$F(D_T, D_N) = (D_T - 64)^2 + (D_N - 0)^2 = (D_T - 64)^2 + D_N{}^2 \quad \textbf{(3-1)}$$

A general cost function that applies to any number of objectives and any number of voxels in an anatomy structure can be expressed as

$$F = \sum_{\sigma=1}^{m} f_\sigma \quad \textbf{(3-2)}$$

where f_σ is the subcost of objective σ, and the sum is over *m* objectives. All objective subcosts can be calculated in a similar manner

$$f_\sigma = \frac{w_\sigma}{n} \sum_{i=1}^{n} g_\sigma(D_i) \quad \textbf{(3-3)}$$

where w_σ is an objective-specific importance weighting; n is the number of voxels in the structure for objective σ; and $g_\sigma(D_i)$ is the cost per voxel at dose voxel D_i for objective σ.

The general form of the cost function, Equations 3-2 and 3-3, is satisfactory and even has the useful property that it is volume normalized; that is, the sum of the costs per voxel for each objective is divided by *n*, the number of voxels in the structure. This eliminates any tendency for an optimization process to favor large structure objectives over small structure objectives. Another way of looking at volume normalization is that the normalized objective really expresses the weighted average cost per voxel for the objective.

Voxel Cost Problems

It is the cost per voxel for the above objectives that is problematic

$$g_\sigma(D_i) = (D_i - D_o)^2 \quad \textbf{(3-4)}$$

or, equivalently

$$g_\sigma(D_i) = w_\sigma(D_i - D_o)^2 \quad \textbf{(3-5)}$$

Suppose that we want some fraction (*e.g.*, 95%) of the target volume to get 95% of the goal dose and only a fraction (*e.g.,* 10%) of the target to get more than 110% of the goal dose. In this example, 95% of 64 Gy is 60.8 Gy and 110% of 64Gy is 70.4 Gy. Therefore, it is likely that we would want target dose voxels with doses outside of the range 60.8 – 70.4 Gy to be penalized similarly to OAR dose voxels whose dose is in excess of 45 Gy.

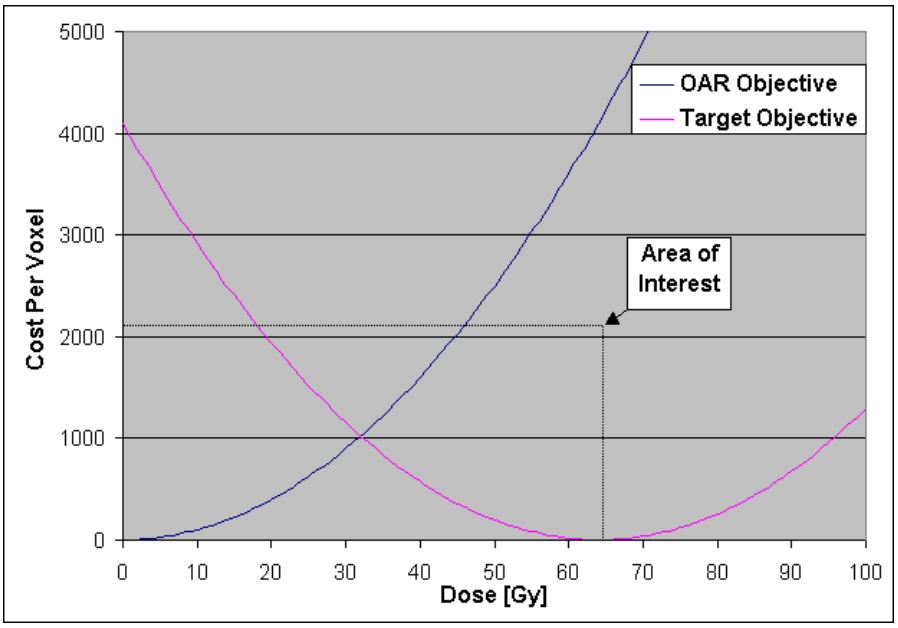


Figure 3-26: Ineffective quadratic objectives, $w_\sigma = 1$ for both objectives

The second form of the cost per voxel, Equation 3-5, can help us understand the problem. Figure 3-26 shows the cost per voxel for the objectives in Equation 3-1 with unity weights. Figure 3-27 is a detail of the area of interest shown in Figure 3-26. Recall that in Prescriptions and Objectives the notion that objectives compete with each other was presented. From Figure 3-27, it is clear that if OAR voxels with 45 Gy were competing with target voxels with 60.8 Gy, with no weight changes, that an optimizer would heavily favor decreasing OAR dose as well as target dose. This could have the effect of improving OAR dose past where it really necessary and decreasing target dose homogeneity.

The problem is that OAR voxels with 45 Gy and target voxels with 60.8 Gy really should be penalized similarly by their respective objectives. Given the objective formulation as in Equations 3-5, the only way that the objectives would penalize doses similarly is to alter the objective importance weights in an attempt to compensate for the ill-suited objective function shapes.

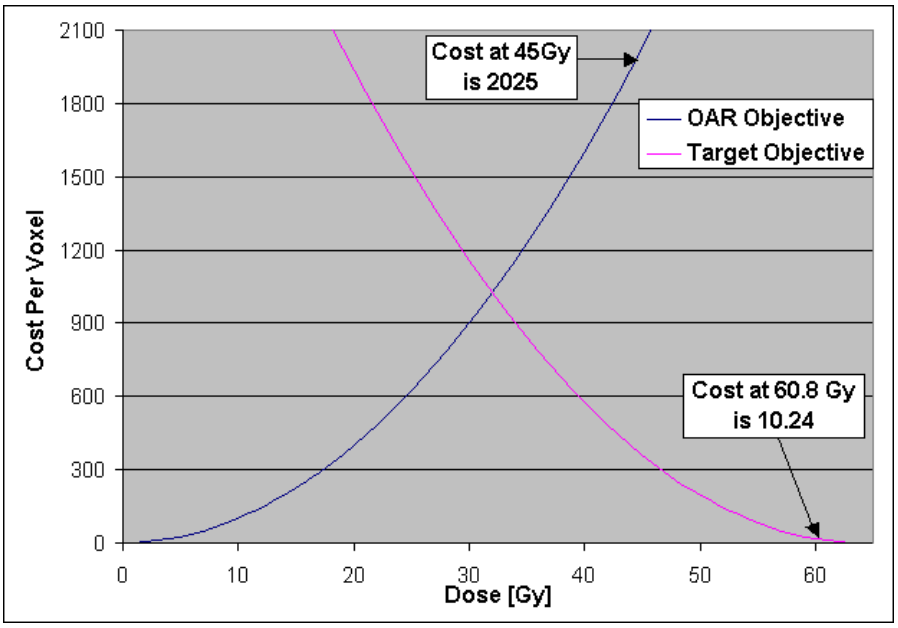


Figure 3-27: Ineffective quadratic objectives detail, w_σ = 1 for both objectives

Deriving Compensating Weights

Imagine that the dose for only two dose voxels needs to be optimized (one for the target and one for the OAR) and that after several iterations of an optimizer, the target voxel dose is 60.8 Gy and the OAR voxel dose is 45 Gy, as shown in Figure 3-27. If it is assumed that the target dose and the OAR dose are correlated, any decrease in OAR dose will cause a proportional decrease in target dose, and any increase in target dose will cause a proportional increase in OAR dose. Any dose change *should*, according to the intent of the prescription, increase the overall evaluation of the cost function. This is because the dose values are at the extreme limit of their acceptable ranges. Any small dose change will cause one dose voxel to move further into its acceptable range while the other moves out of its acceptable range.

Compensating Weights 是不是就是敏感度

With a dose proportionality constant of $\lambda = 1$, simultaneous equations describing a linear relation between target and OAR dose and the point at which objectives are minimally violated can be formulated:

$$D_{OAR} = \lambda \cdot D_{Target} - 15.8 \qquad \textbf{(3-6)}$$

$$w_{OAR} \cdot D_{OAR}^2 = w_{Target} \cdot \left(D_{Target} - 64\right)^2 \qquad \textbf{(3-7)}$$

The solution indicates that OAR dose could further decrease to 24.1 Gy and target could decrease to 39.9 Gy before a minimum of the cost function is found. Thus, without adjusting importance weights, OAR dose is overly restricted and target dose is sacrificed.

Figure 3-28 shows that the adjustment that will satisfy the prescription is the ratio of the OAR objective cost at 45 Gy (2025) to the target objective cost at 60.8 Gy (10.24) which is 197.75:1. That is, a ratio of target to OAR importance weights of almost 200 to 1 is needed to control shape problems of the objective functions. Note that this is a single calculation, and additional importance weight adjustments may be required to get acceptable results.

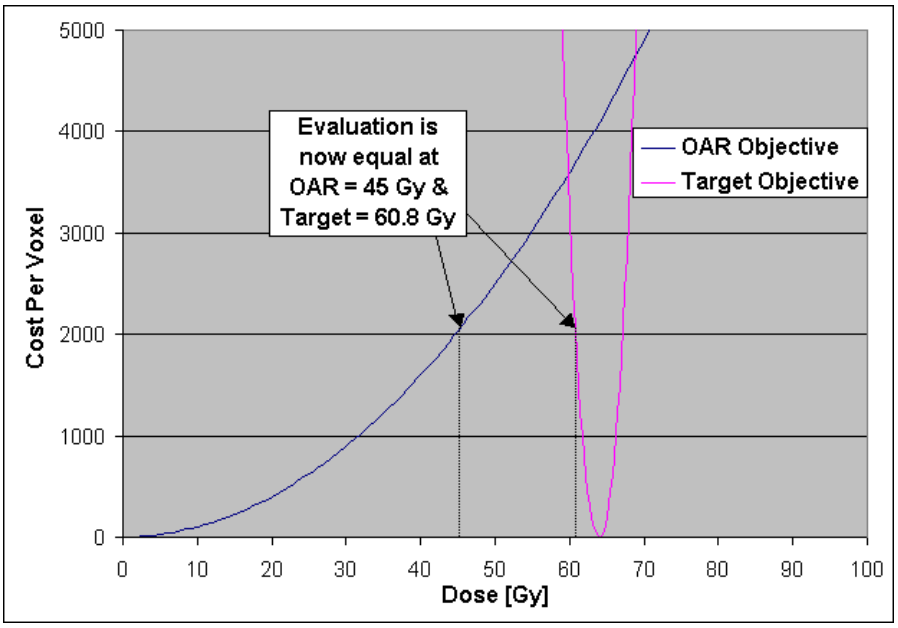


Figure 3-28: OAR objective $w_\sigma = 1$ and target objective $w_\sigma = 197.75$

In contrast to these ineffective objectives, XiO objectives have function shapes that naturally penalize doses outside of reference dose values in a way that does not require importance weights to "fix" shape problems. Figures from Figure 3-14 to Figure 3-25 show that large penalties only apply after reference doses are exceeded.

Prescriptions and Objectives

The most important goal in inverse planning is to satisfy the treatment prescription. In our experience, it is usually best to not directly translate an entire prescription into a collection of corresponding objectives. It is better to start the inverse planning process with active objectives for targets and then add objectives for OARs as needed. There are several reasons for this:

先优化target，再优化危及器官

- By starting with only active target objectives, the achievable limits of target dose homogeneity can be determined.
- As OAR objectives are added, they will tend to compete with target dose minimum dose objectives and degrade target dose homogeneity.
- By limiting the number of OAR objectives, which tend to compete with target minimum dose objectives, it is easier to determine which, if any, of the OAR objectives may compete with target objectives.
- With fewer objectives, better solutions can often be found faster.

Optimization is faster with fewer objectives because of two factors: 1) If there is no active objective for an anatomic structure, the doses for voxels in that structure are updated less often during optimization and 2) a solution can often be found faster because there may be more unconstrained "paths" to an acceptable minimum cost solution.

Additionally, the proper selection of target dose objectives often obviates the need for a maximum dose objective for the patient. It may also obviate the need for other objectives, so the best strategy to get good results may often be an incremental one. incremental one ! ! !

Class Solutions

Class solutions are configurations of beams and collections of objectives that seem to work well in solving inverse planning problems for different classes of cancers and disease sites. In XiO, class solutions, when created, can be saved as XiO templates. The XiO IMRT, implemented in release 4.2, yields good class solutions because its objectives are volume normalized and its cost functions effectively set reference doses and reference volumes.

Anatomy Rank

During inverse planning optimization it is important for the dose voxels inside contoured anatomy structures to be assigned properly to objectives. When anatomy structures do not overlap, this is straightforward. When anatomy structures overlap, the rank of the overlapping anatomy structures is used to resolve voxel ownership. *Anatomy rank only affects overlapping anatomy structures that have active objectives.*

Higher ranked structures have priority over dose voxels in overlapping volumes, while structures at the same rank share dose voxels in overlapping volumes. In XiO IMRT, a structure with a rank of 1 has priority over a structure with rank 2, a structure with rank 2 has priority over one with rank 3, etc. Typically, targets should have the highest rank among anatomy structures, very important OARs should have the next lower ranks, etc. Note that earlier versions of IMRT (*e.g.,* XiO Release 4.1.1) did not allow structure ranks, then called "priorities", to be the same, so volume "sharing" was not supported.

Figure 3-29A shows two structures of different rank and how the voxels in the intersecting region between the structures are owned by the structure of higher rank. Figure 3-29B shows two structures of the same rank and how they share intersecting voxels.

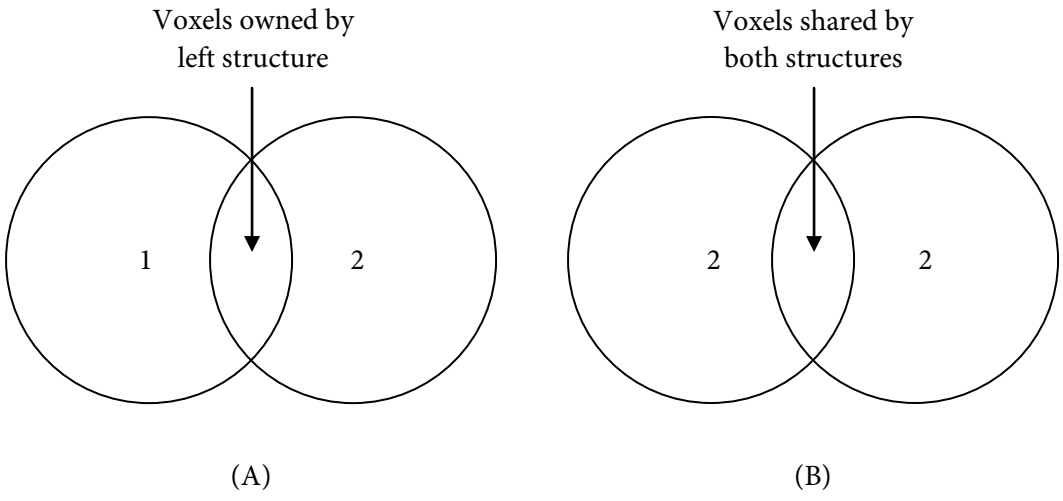


Figure 3-29 Structure rank and voxel ownership

The determination of "unspecified tissue", which is defined in XiO as two slightly different concepts, is related to anatomy rank. For the purposes of calculating dose-volume histograms, unspecified tissue is defined as the tissue volume inside the patient exterior or "skin" that is not inside any other contoured anatomy structure. In XiO IMRT, it is defined, when there is a patient dose objective, as the tissue volume inside the patient exterior that is not inside any contoured anatomy structure for which there is an active objective. Thus, when active objectives change, as is normally the case during inverse planning, the voxels defined to belong to unspecified tissue will change as well. To make these two quantities equal during inverse planning, there must be an active objective for every contoured anatomy structure and the rank of each of these anatomy structures must be less than that of the patient exterior.

Dose-volume objectives often complicate issues related to anatomy rank, because they imply voxel sharing relationships among anatomy structures. For overlapping anatomy, a dose-volume objective specifies that it is all right for the dose to some portion of the anatomy structure with the dose-volume objective to be ignored or given over to dosing requirements of the other structure. For instance, as shown in Figure 3-30, it is not uncommon for a target in a head and neck case to overlap a parotid gland. If the target has higher rank than the parotid gland, a dose-volume objective for the parotid will only claim the non-overlapping voxels. Thus, the tissue volume that the dose-volume objective controls will be less than the actual volume of the parotid and the dose to the parotid may be greater than it should be. Dose-volume histograms and isodose displays will show correct doses.

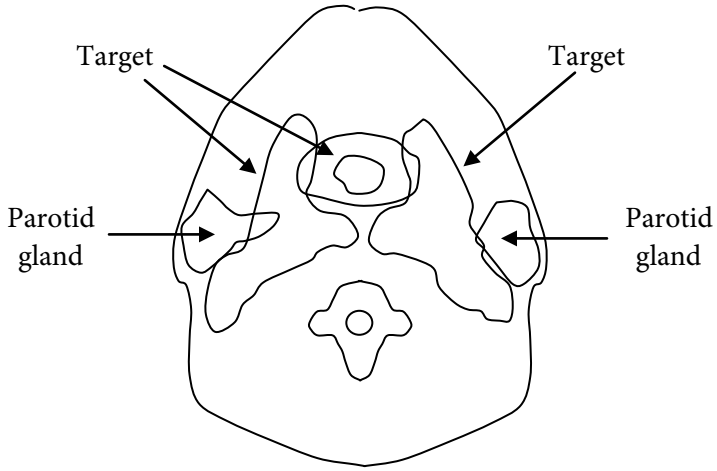


Figure 3-30: Target volumes overlapping parotid glands

Evaluating Results

Show head and neck case with dose trough for cord beamlets.

Summary

Inverse planning in XiO IMRT is accomplished in three stages. In the first stage, optimization is over beamlets, and in the second stage, optimization is over beam doses deliverable by either multileaf collimator (MLC) or compensating filter. In the third stage the segment weights are adjusted so the dose distribution better meets the IMRT Rx.

In contouring anatomy structures for inverse planning, it is important to anticipate the potential interaction between anatomy definitions and dose objectives. Two special cases should be considered: 1) abutting anatomy structures with different dose objectives and 2) dose build-up or build-down in a target. Structure abutment issues lead to situations in which only unattainably high dose gradients can satisfy dose objectives. The solution is to redraw contours, change dose objectives, or to carefully check dose in dose transition regions between the abutting structures. Unwelcome dose build-up in a target, is caused by targets being drawn too close to the patient exterior. Dose build-down in a target can be caused by air cavities drawn as part of the target. One solution to dose build-up/build-down problems is to redraw the target away from the patient exterior or to redraw the target without air cavities. Bolus can be used for some dose build-up problems, moving the build-up region out of the target and into the bolus. Air cavities can be removed from targets by contouring them and setting them to a higher rank than the surrounding target. The air cavity problem in targets is only seen when using algorithms like Superposition that estimate dose depressions in low density tissue properly. This sort of dose depression is largely nonexistent when using FFT Convolution. This is not to say that FFT Convolution should be preferred when dealing with air cavities in targets. On the contrary, if there are air cavities that affect dose in a target to any appreciable degree, their affect on overall dose is probably significant and noteworthy.

The key to specifying effective beams for inverse planning is to maintain beam dose independence, which requires beams that are maximally free to modulate dose to the patient anatomy. In other words, we would like beams in which changing a specific beamlet's weight, changes dose in the patient in a way that is unique and perhaps not possible, by any other combination of beamlet changes. Typically, this means choosing 5-9 coplanar beams that cover the target. Non-coplanar beams may improve results when they decrease the radiological path to a target or remove an organ at risk from a beam path. In some cases, unevenly spaced beams can reduce long radiation path lengths to targets, which will tend to decrease dose to normal tissues. A minimum beam angle separation of about 15° seems to work for many users.

The XiO IMRT pencil beam dose calculation algorithm uses the FFT Convolution, Superposition or Fast Superposition photon dose calculation algorithm to calibrate beamlet doses. Calibration of beamlet doses with a three-dimensional photon algorithm dose makes XiO beamlet doses much more accurate than beamlets used in other commercial treatment planning systems. This is especially true when a superposition algorithm is used in heterogeneous tissues. In this case, beamlet doses mimic build up and build down associated with changing patient densities.

XiO IMRT has two different types of beamlets: beamlets whose weights are actively changed by the optimizer during "line searches" and beamlets whose weights are passively changed after each line search. Active beamlets generally contribute dose to targets and are not blocked by MLC leaves or other treatment aids. Passive beamlets tend to represent leakage dose for regions covered by MLC leaves or blocks.

Minimum and maximum dose objectives, as well as dose-volume objectives, are supported by XiO IMRT. Any contoured anatomy structure can have objectives. Targets can have minimum and maximum dose objectives, while organs at risk can have maximum and dose-volume objectives. Anatomy rank specifies how voxels are shared by objectives when anatomy contours overlap. Typically, targets have lower numerical rank than organs at risk. The patient exterior should have the highest numerical rank; thus excluding all voxels from the patient that belong to other anatomic structures that have active objectives.

Because XiO IMRT uses unconstrained optimization, objective parameters and the definition of patient geometry may need to be altered to obtain satisfactory dosimetric results.

It is usually best not to directly translate an entire prescription into a collection of corresponding objectives. It is better to start the inverse planning process with active objectives for targets and then add objectives for organs at risk as needed. By starting with only active target objectives, the achievable limits of target dose homogeneity can be determined and as organ at risk objectives are added, the ones that compete with target dose minimum dose objectives can be identified. Additionally, with fewer objectives requiring evaluation, better solutions can often be found faster. Class solutions in XiO can be saved as plan templates. XiO Release 4.2 yields good class solutions because its objectives are volume normalized and its cost functions effectively set reference doses and reference volumes.

4. Beamlets

The XiO IMRT optimizer, as in almost every other inverse planning system, uses "beamlets" to test the effect of different fluence intensity profiles on cost function evaluation. As shown in Figure 4-1, beamlets, sometimes referred to as "bixels", are subdivisions of photon beams meant to idealize the way in which photon beam fluence subdivisions behave physically in the patient. Beamlets are divergent and typically incorporate both primary and scatter dose. The dimensions of beamlets are dependent on the physical characteristics of the beam modulating delivery device. As shown in Figure 4-1B, beamlets using multileaf collimator (MLC) delivery could not be narrower in one dimension than the MLC leaf width. In the other dimension, as shown in Figure 4-1C, the beamlet width could be smaller, but it would be limited by the real resolution of the MLC, practical considerations of the beamlet dose representation, and the optimization calculation[10].

Figure 4-1D shows an individual beamlet. It is essentially a rectangular pyramid that can be imagined to emanate from the radiation source and project to its defined dimensions at the isocenter plane. Beamlet dimensions for compensator delivery are somewhat independent of milling considerations. After optimization for compensators, beamlet intensities are smoothed and converted into fluence intensities that will be used to perform a final dose calculation.

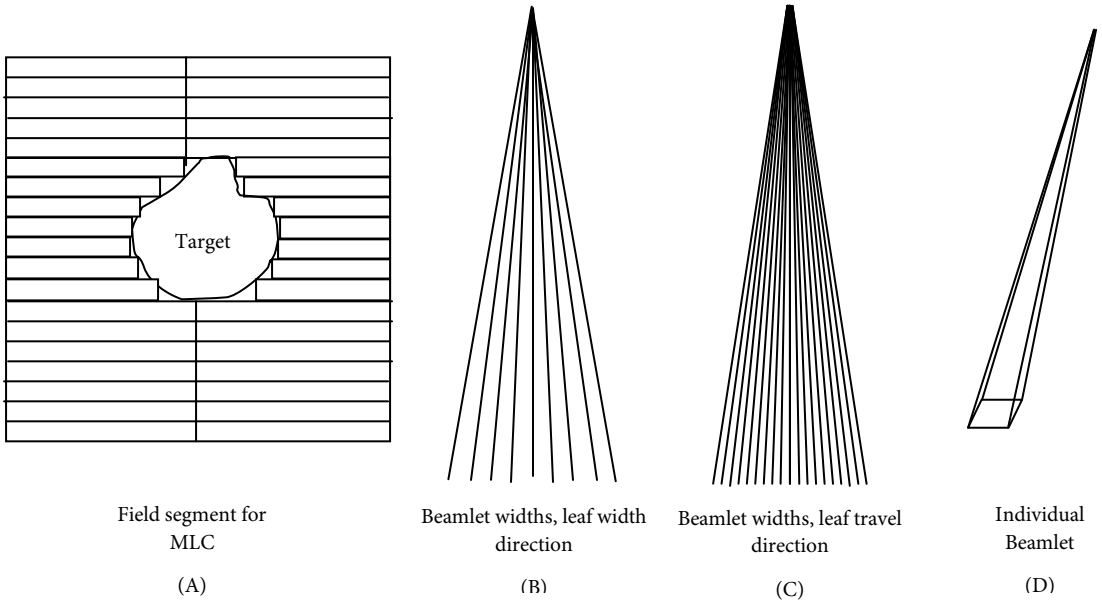


Figure 4-1: Beams divided into beamlets

10 These practical considerations are due indirectly to computer memory limitations. Memory limitations lead to a lower bound on fluence fan spacing for the three-dimensional dose calculations. These fan spacing limitations in turn lead to dose sampling limitations that preclude the use of beamlets smaller than a certain dimension.

Figure 4-2 shows beamlet profiles as calculated by the pencil beam algorithm that is used by XiO IMRT (Release 4.2 and later IMRT releases). This pencil beam algorithm uses either the FFT convolution or the superposition algorithms to calibrate pencil beam profiles similar to those used by Hyperion [Alber, 2000]. ”

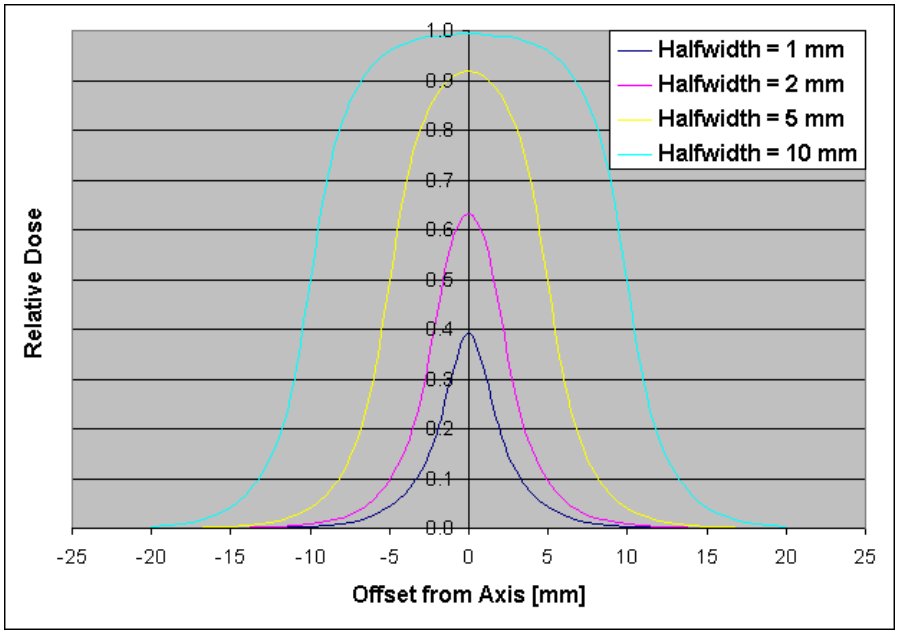


Figure 4-2: Beamlet profiles for different half widths (as in Hyperion)

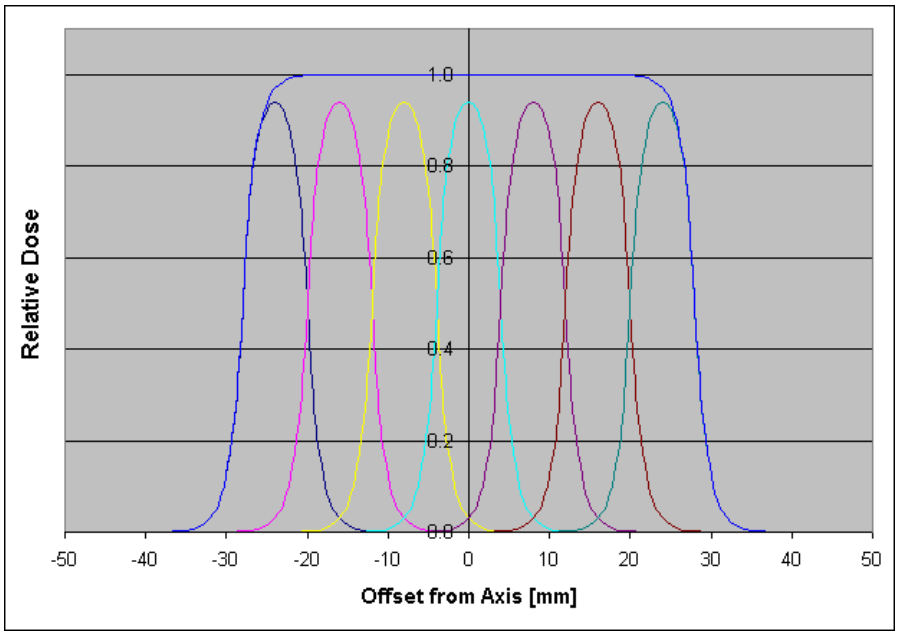


Figure 4-3: Side-by-side beamlets illustrating self-consistency

One of the most important characteristics of beamlets is that a number of beamlets packed side-by-side must be able to reproduce open field doses. Packing can be rectangular, hexagonal, etc. This self-consistency property is illustrated in Figure 4-3, which shows primary and scatter dose in profile overlapping to sum into a smooth open field dose. In XiO, beamlets are rectangular.

The way in which primary and scatter dose is defined in XiO IMRT is different from how these quantities are typically defined. These dose quantities correspond neither to the primary and scatter doses of algorithms like Clarkson, nor do they correspond to primary and scatter doses of more advanced algorithms, like superposition, FFT convolution, or Monte Carlo.

Figure 4-4 shows a relative dose profile at the isocenter plane for a rectangular pyramidal beamlet as in Figure 4-1D. The vertical lines in the figure are meant to correspond to the edges of the pyramid. In XiO IMRT, a beamlet's primary component is defined to be inside of the pyramid's sides and the scatter component is outside of its sides. Figure 4-4 shows this concept for a beamlet that has 1 cm "scatter tails." The length of the scatter tails is controllable with a parameter called the "scatter extent." The exact shape of beamlet profiles is controlled by a mathematical formula that can be seen Pencil Beam Algorithm.

This purely spatial definition of primary and scatter dose works well if the beamlet profiles match the lateral dose distributions that would be predicted for a rectangular unit of fluence. Of course, lateral fluence distributions are likely to be a function of beam energy, field size, depth of measurement, and the off center location of the beamlet. In reality, a perfect match of beamlet profiles and real dose distributions will be rare.

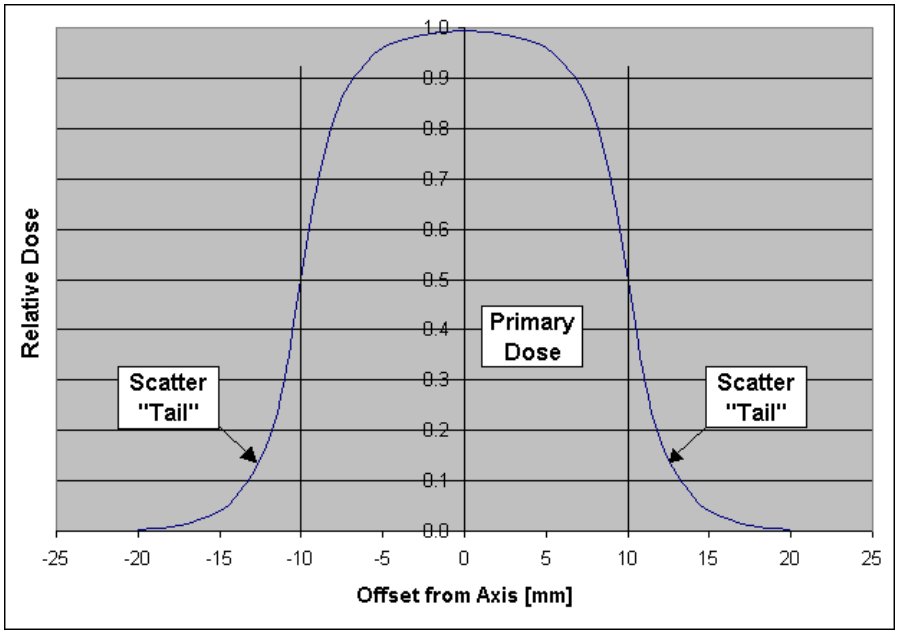


Figure 4-4: Beamlet profile with primary dose and scatter tails

Beamlet Memory Usage

In software, beamlets can be represented as lists of voxels in the dose volume; that is as collections of four-tuples:

{(x1, y1, z1, dose1), (x2, y2, z2, dose2), (x3, y3, z3, dose3)... (xN, yN, zN, doseN)}

In the above representation, each beamlet dose voxel requires exactly 16 bytes of memory (3 integers and 1 floating point number). Perhaps surprisingly, one of the biggest challenges in developing good IMRT software is that the representation of beamlets with overlapping "scatter tails" can require very large amounts of memory.

To get an idea of the amount of memory required for IMRT, think about memory usage in a very simple way: Imagine beamlets that project 5 mm "square" primary dose profiles at the isocenter plane and 1 cm scatter tails, as shown in Figure 4-5. Assume that any increase in the projected area at the isocenter plane from primary voxels only to primary and scatter voxels is proportional to the required increase of memory. In this case, memory requirements for beamlets with primary and scatter dose voxels will be 25 times that for primary beamlet voxels alone (5 * 5 = 25). Using the same logic, 2 cm scatter tails would require 81 times as much storage (9 * 9 = 81). For high resolution (2 mm), multibeam (7-9 beam), large anatomy cases (breast), memory requirements in the range of several gigabytes may be required for beamlet representations alone.

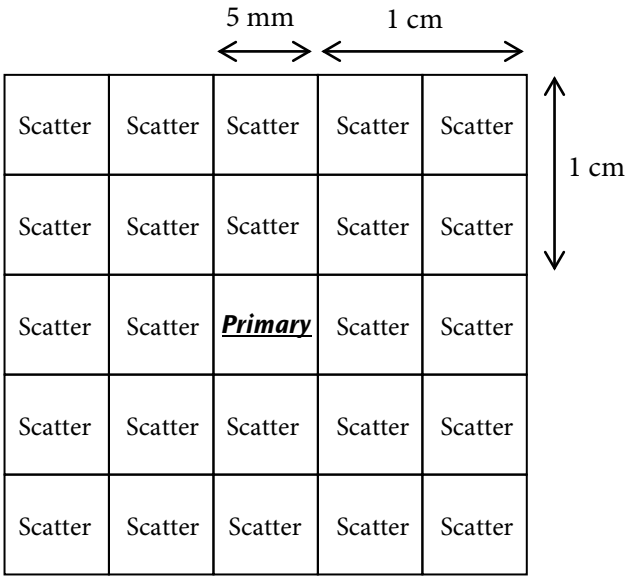


Figure 4-5: Projection of 5mm "square" beamlets at the isocenter plane

This is problematic for several reasons:

- Memory usage for XiO itself is substantial, and the overall addressable memory space for processes is only about 1 gigabyte.
- Disk storage requirements for beamlet doses will increase substantially from the current storage for volumetric dose arrays.
- Reading and writing beamlet dose files to and from disk storage may prove time consuming.

- The speed of inverse plan calculation is roughly proportional to the number of beamlet voxels represented, so 25 times as many voxels means an IMRT calculation that may be 25 times slower.

Some memory can be saved (as it is in XiO Release 4.2) by taking advantage of the fact that beamlet voxels can be represented as a pointer offset from the beginning of allocated memory and a dose.

{(offset1, dose1), (offset2, dose2), (offset3, dose3)... (offsetN, doseN)}

Each beamlet dose voxel using this representation requires exactly 8 bytes of memory (1 integer and 1 floating point number); that is, it requires half as much beamlet voxel memory as the earlier, more obvious representation. This memory savings is very desirable, and may make some plan setups and calculation resolutions feasible that were not feasible with the other representation. It may also improve performance somewhat due to less paging for virtual memory, but it does not solve the issue of calculation time being proportional to the number of beamlet voxels represented.

Another beamlet representation, wavelets, has memory saving potential [Zakarian and Deasy, 2004]. The problem, however, with this representations is that it is "indirect"; that is, to access beamlet doses, the wavelet representation must first be decompressed. The decompressed representation of the beamlet is a volume, much of which has zero dose. Since beamlet voxels with zero (or very low) dose are not used in the optimization process, searching for non-zero dose while ignoring zero dose could be a very inefficient process.

Access to more memory through the use of 64-bit architectures and different operating systems, faster machines, multiple processors, and so forth, may provide some relief for the potential data explosion required of IMRT. However, in the present, the best solution to the beamlet data explosion is to limit the number of beamlet voxels in an intelligent way by either limiting the extents of beamlet scatter tails or limiting the resolution of the calculations without sacrificing the requisite accuracy required. Future releases may have multi-resolution voxel representations in which the voxel size varies according to the size or importance of the structure, the distance from structure boundaries, or gradients in the dose.

Beamlet Types

As mentioned earlier, in Beamlets and Dose, XiO IMRT has two different types of beamlets: beamlets whose weights are actively changed by the optimizer during line searches and beamlets whose dose contributions are passively changed after each line search. Active beamlets generally contribute dose to targets and are not blocked by MLC leaves or other treatment aids. Passive beamlets tend to represent leakage dose for regions covered by MLC leaves or blocks. Passive beamlets cover large areas of the field inside the collimators that have similar transmissions or whose dose can be estimated as a percentage of the maximum active beamlet fluence. When compensators are the delivery method of choice, all beamlets are active and wedges are not supported.

Figure 4-6 shows that XiO allows for specification of an "optimization margin" which is an additional region around a target that will not be covered by MLCs – this is to allow a rim of active beamlets outside of a target without the need to specify a separate target expansion structure. Note that another beamlet parameter, the "scatter extent" is independent of the optimization margin.

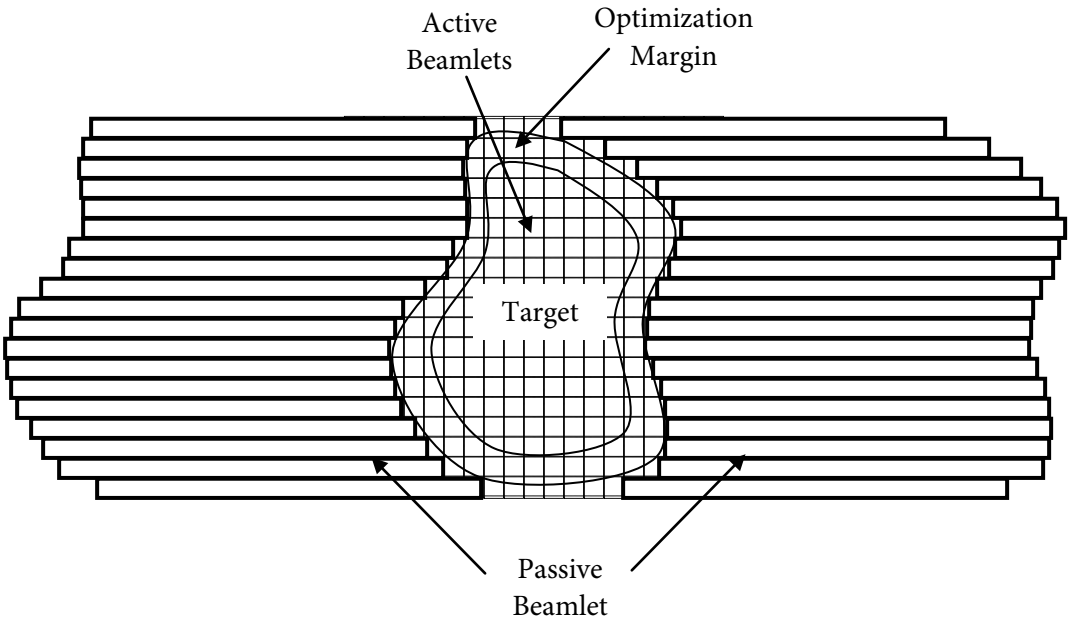


Figure 4-6: Beamlets with zero insertion showing an optimization margin

Summary

Beamlets, sometimes referred to as "bixels", idealize the relationship between photon beam fluence subdivisions and dose in the patient. They are divergent and typically incorporate both "primary" and "scatter" dose. The dimensions (i.e., projection at the isocenter plane) of beamlets may be dependent on the physical characteristics of the beam modulating delivery device. For instance, beamlets for multileaf collimator (MLC) delivery could not be narrower in one dimension than the MLC leaf width. In the other dimension, the beamlet width could be smaller, but would be limited by the resolution of the MLC and practical considerations of the beamlet dose representation and the optimization calculation.

XiO IMRT has two different types of beamlets: beamlets whose weights are actively changed by the optimizer during line searches and beamlets whose dose contributions are passively changed after each line search. Active beamlets generally contribute dose to targets and are not blocked by MLC leaves or other treatment aids. Passive beamlets tend to represent leakage dose for regions covered by MLC leaves or blocks.

One of the biggest issues in using beamlets in inverse planning is the extraordinary amount of memory needed to represent scatter dose; that is, essentially, the dose that overlaps from one beamlet to its neighbor. The user can limit scatter dose memory requirements by increasing voxel size or limiting the scatter extent of beamlets, or both. 在逆向计划中，使用beamlet最大的问题是散射剂量存储的问题，非常消耗内存。

5. Cost Function Mathematics

A cost function for XiO has the form:

$$F = \sum_{\sigma=1}^{m} f_\sigma \qquad \textbf{(5-1)}$$

where f_σ is the subcost of objective σ, and the sum is over *m* objectives.

Objective Subcosts

All dose objective subcosts in XiO IMRT are calculated in a similar manner. Each dose objective has the form:

$$f_\sigma = \frac{w_\sigma}{n} \sum_{i=1}^{n} g_\sigma(D_i) \qquad \textbf{(5-2)}$$

Where

w_σ	is an objective-specific importance weighting;
n	is the number of voxels in the structure for objective σ
$g_\sigma(D_i)$	is the voxel cost at dose voxel D_i for objective σ

Equation 5-2 says that the subcost of any objective is calculated by determining the dose at each voxel in the anatomy structure pertaining to the objective and summing the corresponding voxel subcost. That summation is then divided by the number of voxels in the structure and multiplied by an importance weight. Dividing through by the number of voxels eliminates potential bias towards objectives for large structures, so a cost function is essentially penalized for the fraction of points exceeding the reference dose. The word "essentially" is used because the distribution of dose values also affects the evaluation of the objectives. A few far outlying dose points might contribute to the cost just as much as a larger number of violating dose points nearer to the reference dose. That is the nature of these cost functions, and yet another reason why objective parameters may require some adjustment during the user's interaction with any IMRT implementation.

Dose Objectives

XiO IMRT supports both minimum and maximum dose objectives. Targets require a combination of a minimum dose objective to achieve the nominal target dose and a maximum dose objective to enforce target dose homogeneity. Organs at risk (OAR) typically use a single maximum dose objective or, as will be discussed in Dose-Volume Objectives, one or more dose-volume objectives. Maximum and minimum dose objective functions are composed of a linear section followed by a one-sided, higher-order (2-5) penalty for overdose and underdoses in excess of an extreme reference dose value.

Recall the first example in Chapter 0 shown in Table 3-1. Figure 3-14, the cost per voxel graph for the example dose objectives, is reproduced in Figure 5-1. Figure 5-2 shows the same objectives using mathematical symbology that will be used in equations describing the objectives. In Figure 5-2, we can see that for the maximum dose objective, D_0 corresponds to the threshold dose and D_{0+} corresponds to the maximum dose. For the target, which has both a minimum dose objective and a maximum dose objective, D_0 is the goal dose. For the target maximum dose objectives, again, D_{0+} corresponds to the maximum dose and for the target minimum objective, D_{0-} corresponds to the minimum dose.

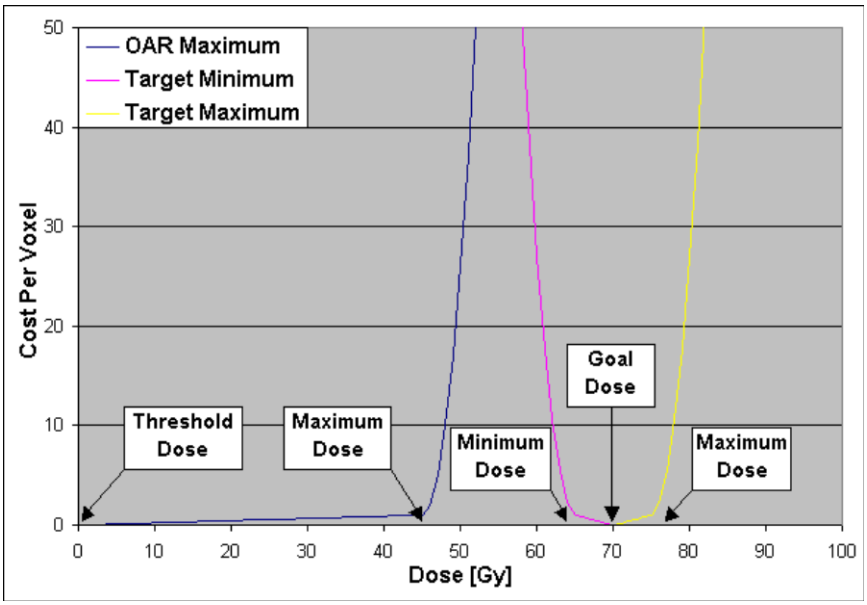


Figure 5-1: Cost per voxel functions from simple target and OAR example

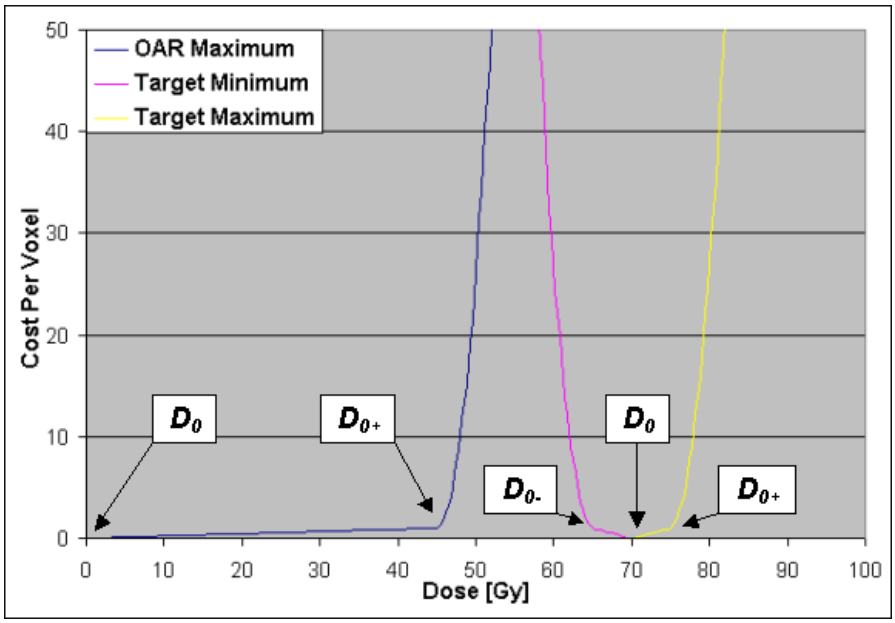


Figure 5-2: Cost functions per voxel using mathematical symbols

Cost per voxel for maximum dose objective

$$0 \le D_i \le D_0 \quad : g_\sigma(D_i) = 0 \qquad \textit{(5-3)}$$

$$D_0 < D_i \le D_{o+} \quad : g_\sigma(D_i) = m_+ \cdot (D_i - D_0) \ \textit{(5-4)}$$

$$D_{0+} < D_i : g_\sigma(D_i) = (D_i - D_{0+})^k + m_+ \cdot (D_i - D_0) \qquad \textit{(5-5)}$$

where

D_0	is a goal dose
D_{0+}	is a maximum desired dose
m_+	is a linear penalty for doses greater than the goal dose
k	is a penalty power.

Cost per voxel for minimum dose objective

$$0 \le D_i < D_{o-} \quad : g_\sigma(D_i) = (D_{0-} - D_i)^k + m_- \cdot (D_0 - D_i) \ \textit{(5-6)}$$

$$D_{0-} \le D_i < D_o \quad : g_\sigma(D_i) = m_- \cdot (D_0 - D_i) \qquad \textit{(5-7)}$$

$$D_0 \le D_i \quad : g_\sigma(D_i) = 0 \qquad \textit{(5-8)}$$

Where

D_0 is a minimum desired dose

m_- is a linear penalty for doses less than the goal dose.

$$m_+ = \max\left(1, \frac{1}{D_{0+} - D_0}\right) \quad m_- = \max\left(1, \frac{1}{D_0 - D_{0-}}\right) \quad \textbf{(5-9,5-10)}$$

Typically, the threshold dose, D_0, is 0 cGy for OARs. For targets, the goal dose is, typically, the middle value between the minimum and maximum desired target doses; that is, ½(D_{0-} + D_{0+}).

The partial derivatives of objectives, with respect to beamlet weights, are used to calculate gradient directions along which to search during the line search portions of the optimization process.

Maximum dose objectives, cost per voxel derivatives:

$$0 \le D_i \le D_0 \quad : \frac{\partial g_\sigma}{\partial D_i} = 0 \quad \textbf{(5-11)}$$

$$D_0 < D_i \le D_{o+} \quad : \frac{\partial g_\sigma}{\partial D_i} = m_+ \quad \textbf{(5-12)}$$

$$D_{0+} < D_i \quad : \frac{\partial g_\sigma}{\partial D_i} = k \cdot (D_i - D_{0+})^{k-1} + m_+ \quad \textbf{(5-13)}$$

Minimum dose objectives, cost per voxel derivatives:

$$0 \le D_i < D_{o-} \quad : \frac{\partial g_\sigma}{\partial D_i} = k \cdot (D_i - D_{0-})^{k-1} + m_- \quad \textbf{(5-14)}$$

$$D_{0-} \le D_i < D_o \quad : \frac{\partial g_\sigma}{\partial D_i} = m_- \quad \textbf{(5-15)}$$

$$D_0 \le D_i \quad : \frac{\partial g_\sigma}{\partial D_i} = 0 \quad \textbf{(5-16)}$$

Partial derivative of objective with respect to beamlet weight:

$$\frac{\partial f_\sigma}{\partial b_j} = \frac{w_\sigma}{n} \sum_{i=1}^{n} \frac{\partial g_\sigma}{\partial D_i} \cdot \frac{\partial D_i}{\partial b_j} \quad \textbf{(5-17)}$$

$\frac{\partial D_i}{\partial b_j}$ turns out to simply be the voxel dose for the beamlet. In other words, if the beamlet weight is 1, this is the contribution to the dose that beamlet ***j*** makes at dose volume location ***i***.

Before each line search, Equation 5-17 is calculated for all beamlets and active objectives, including dose-volume objectives. The gradient is calculated by summing the contribution of each these partial derivatives with respect to beamlet weights.

$$\nabla F = \sum_{\sigma=1}^{m} \frac{\partial f_\sigma}{\partial b_1} \cdot \hat{b}_1 + \sum_{\sigma=1}^{m} \frac{\partial f_\sigma}{\partial b_2} \cdot \hat{b}_2 + \sum_{\sigma=1}^{m} \frac{\partial f_\sigma}{\partial b_3} \cdot \hat{b}_3 + \ldots \sum_{\sigma=1}^{m} \frac{\partial f_\sigma}{\partial b_k} \cdot \hat{b}_k \qquad \textbf{(5-18)}$$

where *k* is the number of beamlets. The gradient is then normalized (and multiplied by negative 1) and used to direct the line search.

Dose-Volume Objectives

Dose-volume objectives were adapted from a well-known formulation [Bortfeld *et al.*, 1997] and are similar to maximum dose objectives except that only dose voxels greater than some specified dose equivalent to D_{0+} and less than some dose, D_v, are penalized using a higher power penalty. Doses greater than D_v are only penalized linearly. D_v is the dose at which the volume requirement would be satisfied given the current dose distribution. Figure 5-3 illustrates this concept using an arbitrary dose distribution for the objective that no more than 35% of a volume should receive more than 40 Gy. Figure 5-4, which is the same as Figure 3-24, graphically shows the penalties that would be applied in Figure 5-3, according to Equations 5-19 – 5-22.

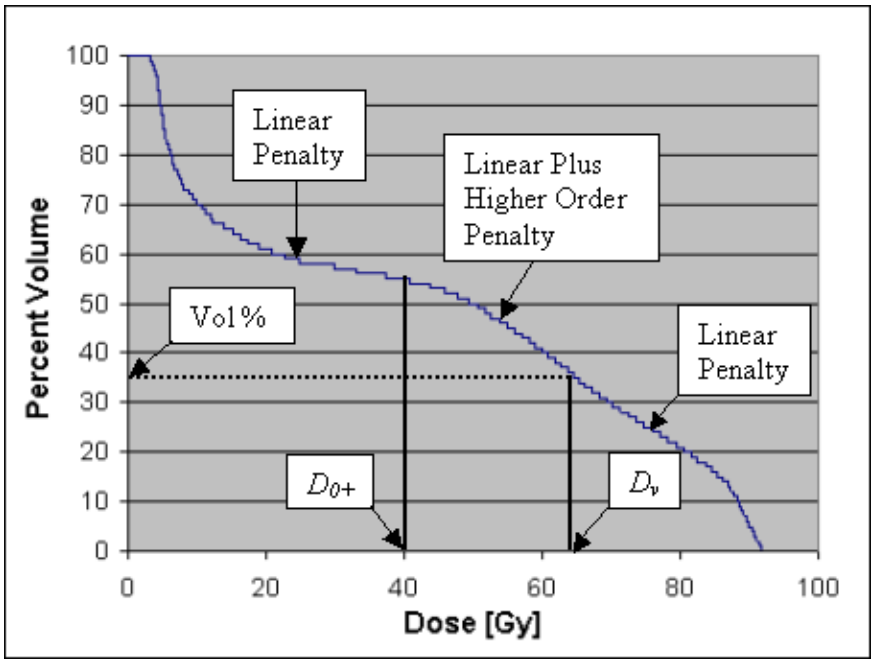


Figure 5-3: Dose-volume penalties and D_v

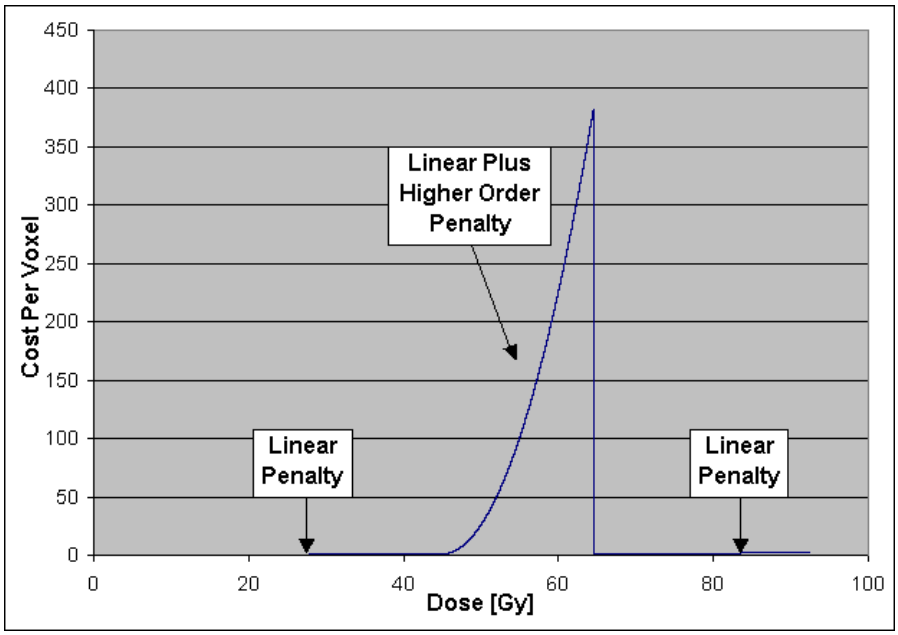


Figure 5-4: Dose-volume objective cost per voxel vs. dose

Cost per voxel derivatives for dose-volume objectives:

$$0 \le D_i \le D_0 \quad : g_\sigma(D_i) = 0 \qquad \textbf{(5-19)}$$

$$D_0 < D_i \le D_{o+} \quad : g_\sigma(D_i) = m_+ \cdot (D_i - D_0) \qquad \textbf{(5-20)}$$

$$D_{0+} < D_i \le D_v \quad : g_\sigma(D_i) = (D_i - D_{0+})^n + m_+ \cdot (D_i - D_0) \qquad \textbf{(5-21)}$$

$$D_v < D_i \quad : g_\sigma(D_i) = m_+ \cdot (D_i - D_0) \qquad \textbf{(5-22)}$$

Dose-volume objectives, cost per voxel derivatives:

$$0 \le D_i \le D_0 : \frac{\partial g_\sigma}{\partial D_i} = 0 \qquad \textbf{(5-23)}$$

$$D_0 < D_i \le D_{o+} \quad : \frac{\partial g_\sigma}{\partial D_i} = m_+ \qquad \textbf{(5-24)}$$

$$D_{0+} < D_i \le D_v \quad : \frac{\partial g_\sigma}{\partial D_i} = k \cdot (D_i - D_{0+})^{k-1} + m_+ \qquad \textbf{(5-25)}$$

$$D_v < D_i : \frac{\partial g_\sigma}{\partial D_i} = m_+ \qquad \textbf{(5-26)}$$

The key to appreciating how dose-volume objectives work is to realize that, since the dose region between D_{0+} and D_v is heavily penalized it will tend to diminish in size as optimization progresses. When D_v is equal to D_{0+}, the dose-volume objective will be satisfied.

Summary

All XiO dose objectives can be described using cost per voxel functions that have linear and higher order penalty regions. For maximum dose objectives, the linear penalty region starts at some threshold dose and increases linearly as dose increases until a maximum reference dose is reached. At this point, higher doses are penalized using higher order penalties. Minimum dose objectives are the mirror image of maximum dose objectives. Their linear penalty region starts at some goal dose and increases linearly as dose decreases until a minimum reference dose is reached. At this point, lower doses are penalized using higher order penalties. Dose-volume objectives are like maximum dose objectives, except that their higher order penalty region is between a maximum reference dose and the dose at which their volume objective is satisfied. During optimization the dose at which the volume objective is satisfied moves nearer to the reference dose. When it is equal to the reference dose, the objective is satisfied.

The combination of a linear and higher order penalty in the XiO objectives has the property 1) that the higher order penalties tend to first force dose distributions inside of acceptable limits and 2) that the linear penalties allow residual improvement in dose distributions once higher order penalties are eliminated.

6. Optimization – Advanced Concepts

In general, optimization in inverse planning is the process of finding a minimum value of a dose-based cost function in a multi-dimensional space that is defined by treatment-machine-controllable parameters. XiO IMRT has a two-stage process. In Stage I, optimization is over ideal beamlet intensities and, in Stage II, optimization is over deliverable dose beam weights. In Stage III, optimization is over individual segment weights. The optimization details covered in this chapter are independent of whether the optimization is over beamlets or beams, except when minimum transmission and passive beamlets are discussed. *See* Active and Passive Beamlets.

Gradient Methods

XiO IMRT Release 4.2 uses a "conjugate gradient" optimization algorithm, which is a special type of "gradient descent" algorithm [Nocedal and Wright, 2000; Press *et al.*, 2002]. Gradient descent methods require multiple iterations of two subprocesses. Figure 6-1 shows imaginary gradient descent optimization iterations on isocost lines for two beamlets. Each iteration starts by taking derivatives of the cost function with respect to the intensity of each beamlet. This derivative, which is actually the normalized negative of the gradient, defines a multidimensional vector in beamlet intensity space, which is called the "line search direction." The line search direction is defined to be the "best" direction in which to search for a minimum of the cost function, given the current location in the search space. The second part of the subprocess is what is called a "line search." Each line segment in the figure is meant to represent an individual line search as simulated in *Optimization – Basic Concepts*.

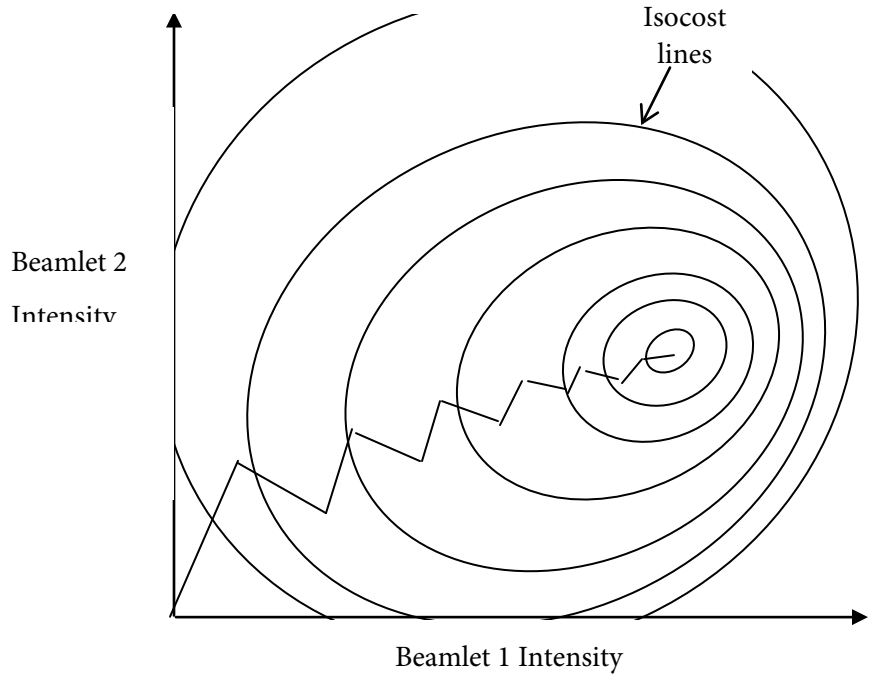


Figure 6-1: Gradient descent minimization of a cost function of two beamlets

Conjugate gradient methods differ from gradient descent methods only in the way that "best" line search directions are defined. It may be counterintuitive, but the gradient, which defines what appears to the steepest descent towards the minimized function, may actually be inefficient at finding the minimum. This is because a sequence of gradient directions can actually interfere with each other; that is, progress made during one line search can be partially undone by a subsequent line search. Compare Figure 6-1 with Figure 6-2, which represents search directions that might be selected with the conjugate gradient method. Intuitively, it can it said that conjugate gradient methods straighten out the path from start to finish. This straightening of the path leads to a good solution in fewer iterations of the line search. Mathematically, conjugate gradient search makes certain that each line search direction is *orthogonal* to the previous direction. Further discussion of conjugate gradient methods is beyond the scope of this document. The reader is encouraged to consult the previously-mentioned references for a complete mathematical and algorithmic review of these two optimization methods [Nocedal and Wright, 2000; Press *et al.*, 2002].

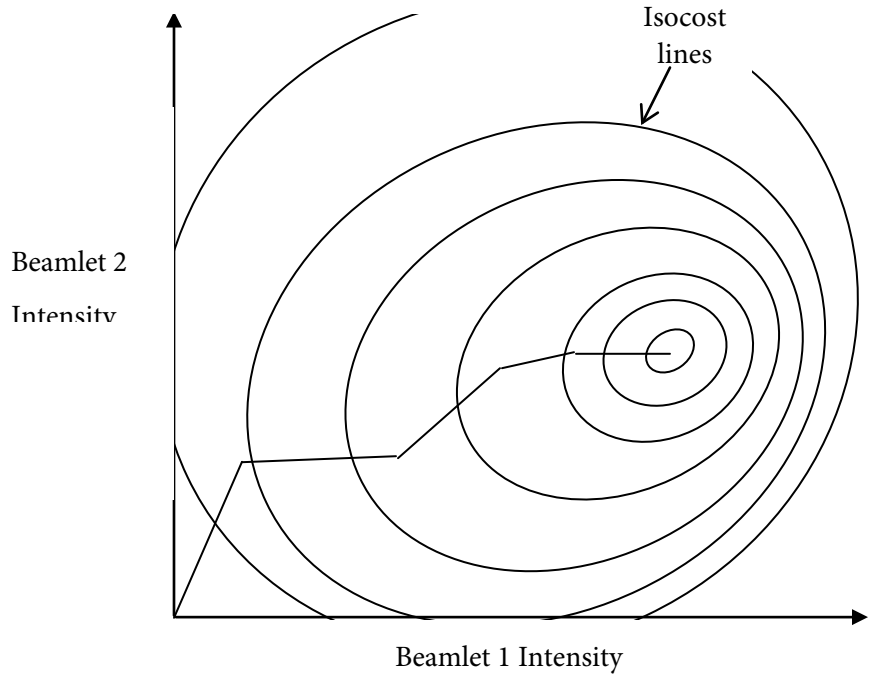


Figure 6-2: Conjugate gradient minimization of a cost function of two beamlets

Convergence

As optimization progresses, the total cost function, which is the sum of individual subcosts for each active objective, converges towards a minimum value. Figure 6-3 and Figure 6-4 show the subcost and total cost evaluations for a head and neck case for the same objectives. Figure 6-3 shows details of objective subcosts for the first thirty iterations, while Figure 6-4 shows subcosts for the first 100 iterations. The figures show that the rate of progress to minimize the total cost is asymptotic and that no significant progress is made after the thirtieth iteration. This sort of convergence behavior is typical of the XiO optimizer, but there is a possibility that significant progress can be made for higher numbers of iterations.

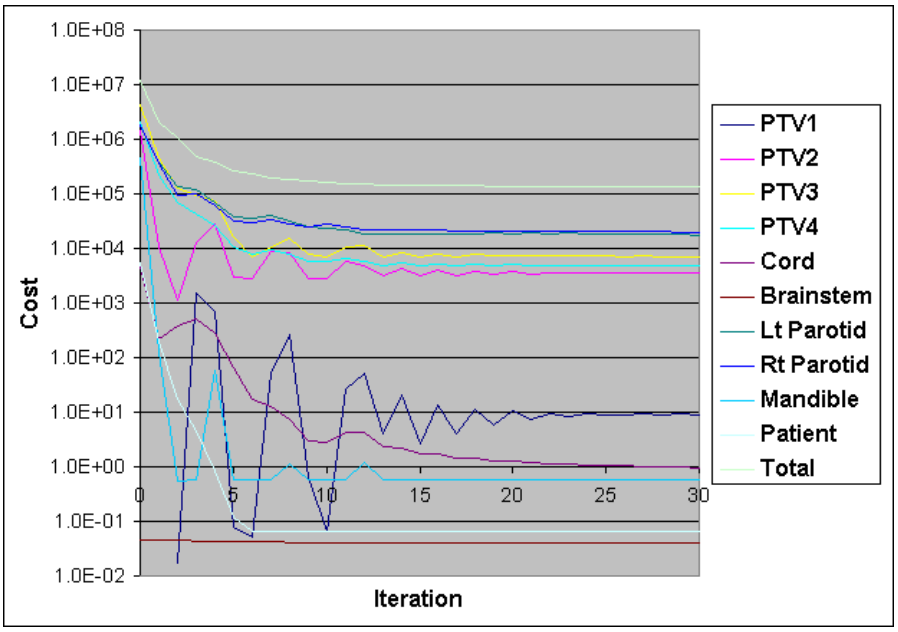


Figure 6-3: Convergence of objective subcosts and total cost function, 30 iterations

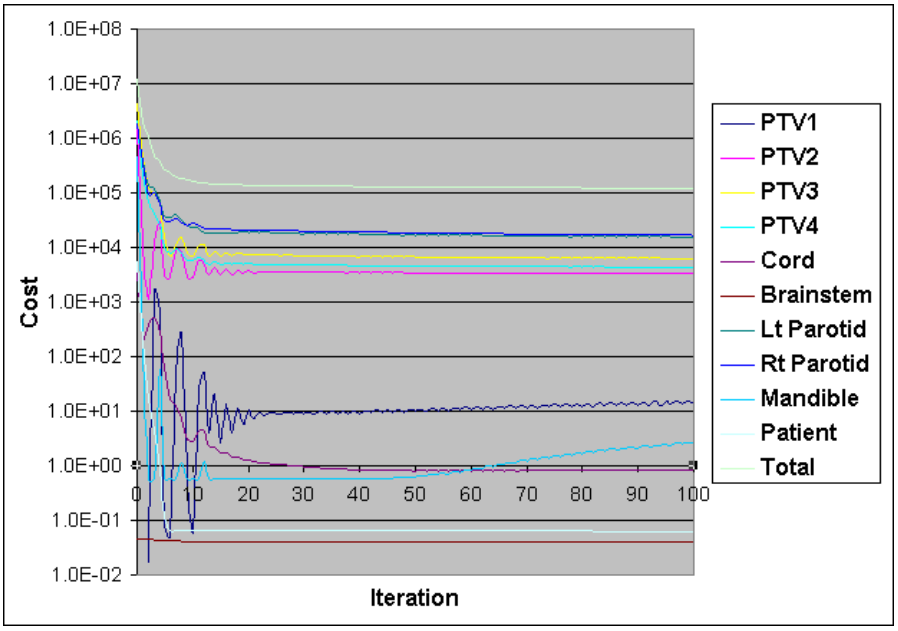


Figure 6-4: Convergence of objective subcosts and total cost function, 100 iterations

Note that the measure of convergence that the user sees in XiO outputs and the convergence thresholds that are user can set rely on the following formula:

$$PercentConvergence = \frac{PreviousCost - CurrentCost}{InitialCost} \cdot 100 \qquad \textbf{(6-1)}$$

Figure 6-5 shows a graph of Equation 6-1 iteration by iteration for the previous figures. Earlier, it was shown that the majority of cost improvement was completed by iteration thirty. From Figure 6-5, it is clear that the user must set a convergence criterion of no more than about 0.003% for the optimizer to execute thirty cycles for this case.

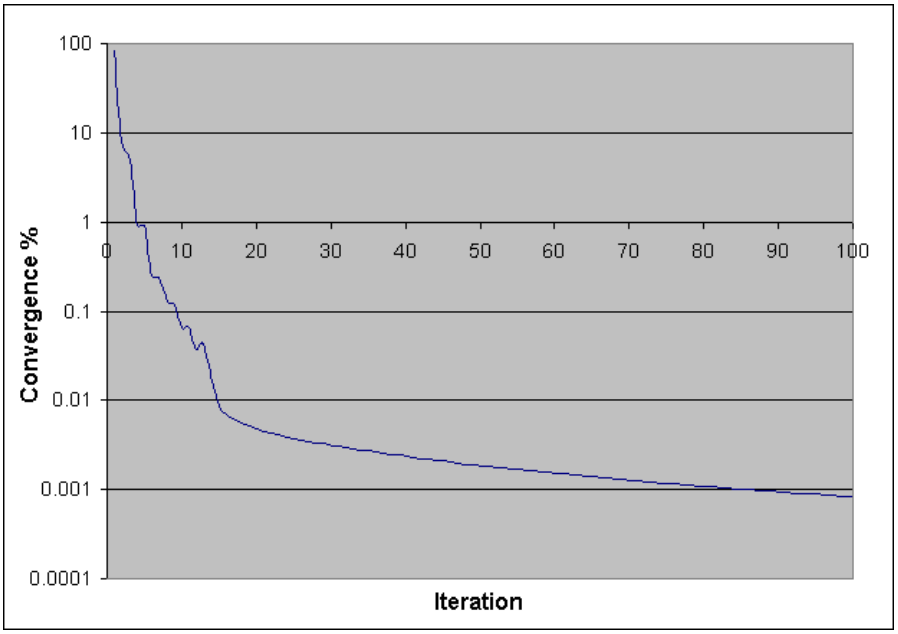


Figure 6-5: Convergence percent by iteration

Unfortunately, there are no solid rules about how many iterations to execute or how to select convergence criteria to achieve acceptable results. The only advice that can be offered is that the use of objective penalties higher than 2.0 can cause objective evaluations to oscillate somewhat during optimization. This may lead to the need for more iterations to converge to a satisfactory result. And, because initial cost function evaluations may be much higher than with quadratic penalties, the convergence criterion may need to be reduced.

Cost Function Convexity

It is well known that a linear combination of convex functions will produce a composite function that is also convex and that convex functions have a single global minimum [Nocedal and Wright, 2000]. This fact is important in inverse planning, because we would like to have confidence that a minimum cost solution that is found is optimal. That is, we would prefer that the optimizer converge at a global minimum rather than a local minimum.

Mathematically, a convex function is a continuous function whose value at the midpoint of every interval in its domain does not exceed the average of its values at the ends of the interval [MathWorld].

In other words, a function $f(x)$ is convex on an interval $[a,b]$ if for any two points x_1 and x_2 in $[a,b]$,

$$f\left[\frac{1}{2}(x_1 + x_2)\right] \leq \frac{1}{2}\left[f(x_1) + f(x_2)\right] \qquad \textit{(6-2)}$$

For minimum and maximum dose objectives, it is clear that the condition in Equation 6-1 is satisfied, so cost functions made exclusively of dose objectives should arrive at the single global minimum in the search space. On the other hand, local minima are known to exist for cost functions that include dose-volume objectives. The consensus is, however, that the impact of local minima caused by dose-volume objectives is minor. That is, very rarely do dose-volume-objective-induced local minima block better clinical solutions [Deasy, 1997;Wu and Mohan, 2002].

Line Search

The line search method employed by an optimizer is extremely important with respect to its speed of convergence. A common assumption of all line search algorithms is that in each line search direction there is a single minimum.

In the section Line Searches, "line search magnitude" was defined as a normalized "distance" in a line search direction from some starting point in the search space and the "line search minimum" as the distance at which a minimum is found during a line search. Figure 6-6 shows a theoretical cost function evaluation for a line search. We see that at a line search magnitude of nearly 1.5 that there is a minimum of the cost function in that line search direction.

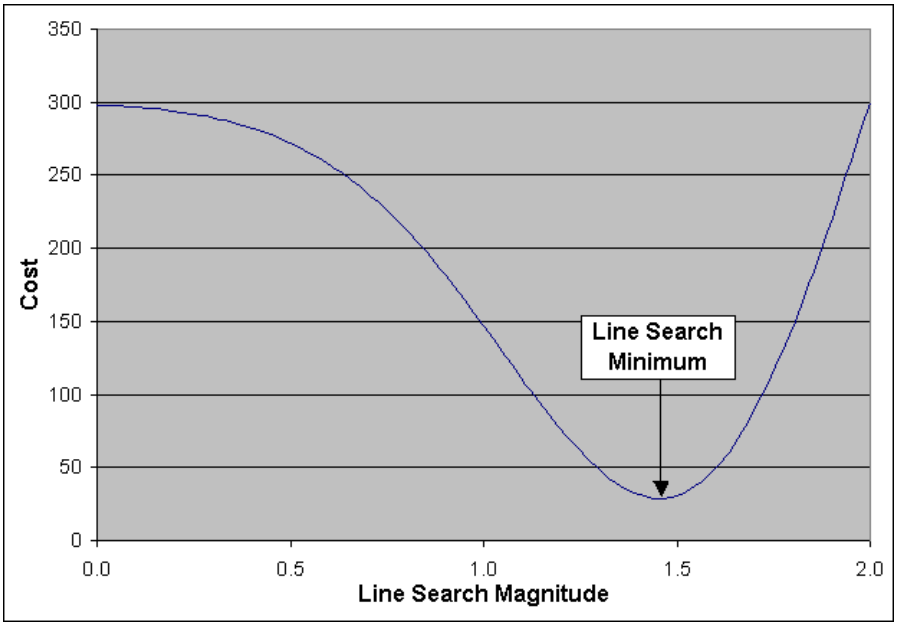


Figure 6-6: Cost as a function of line search magnitude

One of the innovative aspects of the XiO IMRT implementation is that it uses two different line search algorithms during optimization. First, it uses a successive approximation algorithm to determine the general range of line search magnitudes. For each subsequent iteration, it uses Brent's method and a

heuristically-determined line search magnitude upper bound. This hybrid search approach results in very fast optimization with the line search magnitude estimation alone yielding about a 35% gain in speed.

To make the following discussion of line search algorithms more concrete, recall that the space in which optimization is occurring is the space of legal beamlet weights; beamlet weights cannot be negative or so small that they do not represent minimum leakage dose through treatment devices. In this context, each line search magnitude represents a proportional change of the complete collection of beamlet weights with respect to some reference point in the search space and each point in the space is a vector of beamlet weights.

Initial Line Search Method

The purpose of the initial line search method, which is used only once, is to determine an upper bound on the line search magnitude that can be used by Brent's method for the second iteration and further iterations of the optimizer. Further upper bounds on the line search magnitude are determined as discussed in Line Search Estimation Heuristic. The lower bound of the line search magnitude for Brent's method is always zero, since it is likely, as the optimizer approaches convergence, that beamlet weight changes, from one iteration to the next, will become very small.

The issue with the upper bound on the line search magnitude is that it cannot be determined *a priori* and can range over several orders of magnitude. It has been shown during numerous tests that the initial line search magnitude is always the largest for the initial line search. Figures 6-7 through 6-9 graphically show how the initial line search works. The method is very simple, somewhat inefficient, but very robust, and obtains a result comparable to that obtained by Brent's method.

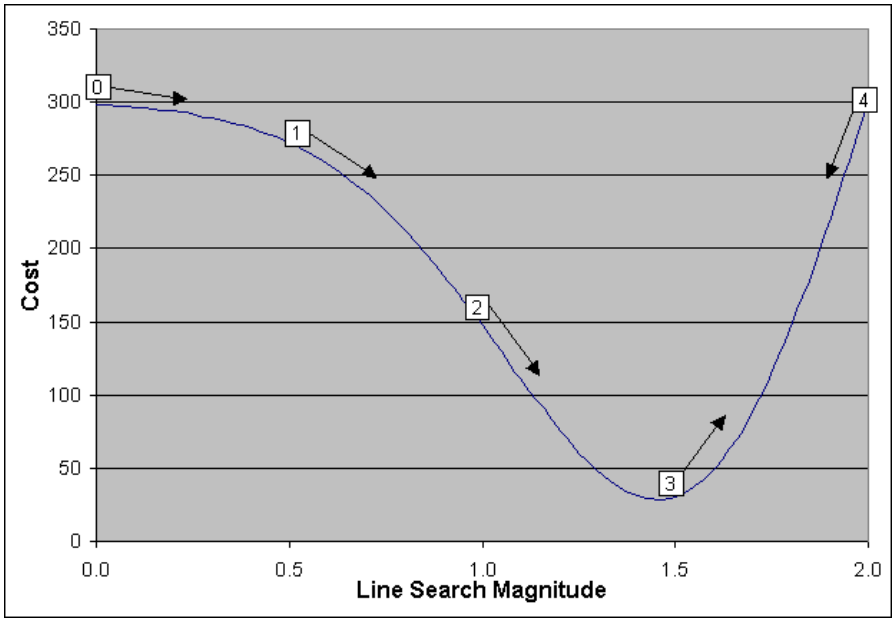


Figure 6-7: Initial Line Search First Direction

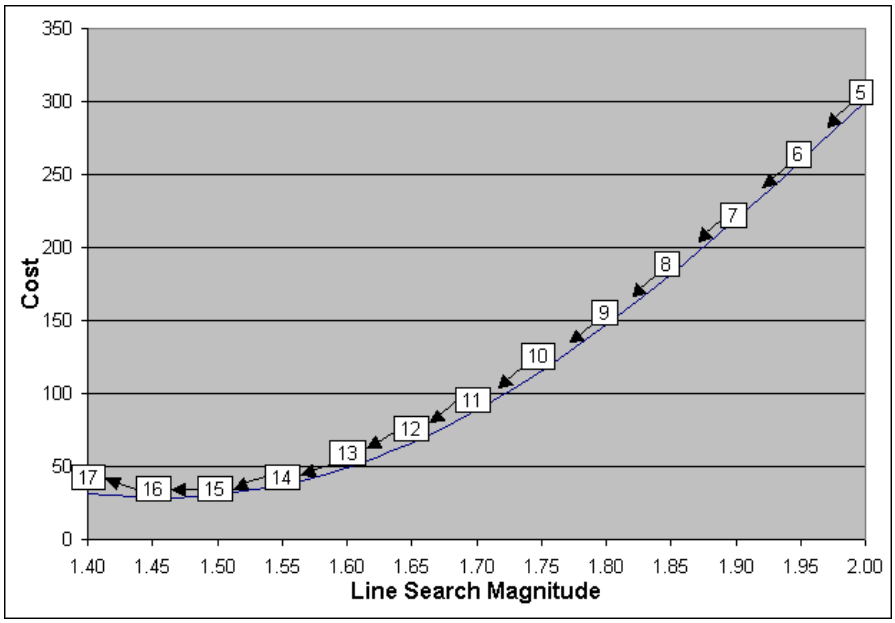


Figure 6-8: Initial Line Search Second Direction

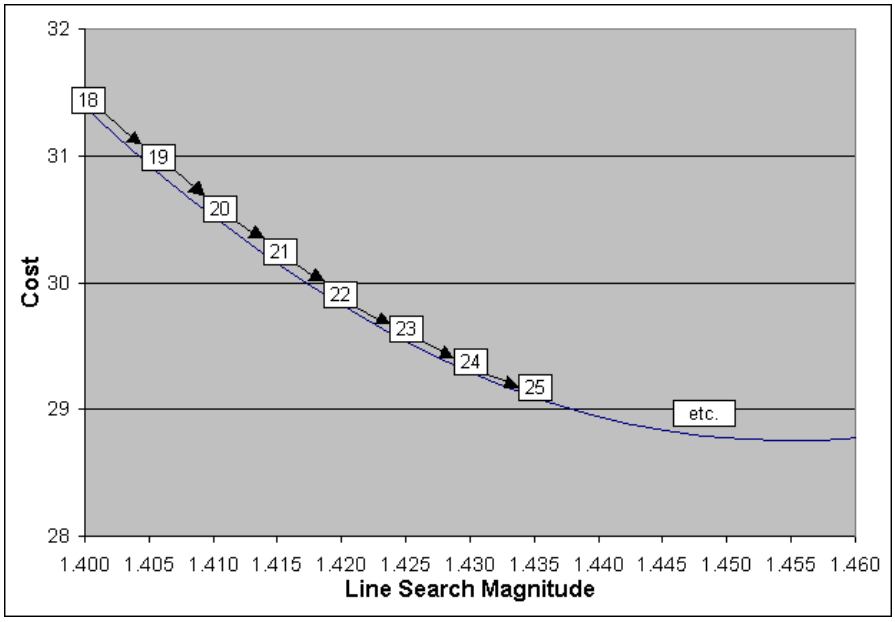


Figure 6-9: Initial Line Search Third Direction

Algorithm pseudocode:

x_0 Initial vector or beamlet weights

G Negative gradient of the cost function

S_0 Initial step size

S_{min} Minimum step size

$M(i) = 0$ Line Search Magnitude

$C(i) = F(x_0 + M(i)*G)$ Total Cost

$S(i) = S_0$ Step Size

$D = 1$ Direction of search

while($S(i) >= S_{min}$)

{

$M(i-1) = M(i)$

$C(i-1) = C(i)$

$M(i) = M(i) + S(i)* D$

$C(i) = f(x_0 + M(i)*G)$

while($C(i) < C(i–1)$)

{

$M(i-1) = M(i)$

$C(i-1) = C(i)$

$M(i) = M(i) + S(i)* D$

$C(i) = f(x_0 + M(i)*G)$

}

$C(i) = f(x_0 + M(i-1)*G)$

$M(i) = M(i-1)$

$D = D * -1$

$S(i) = S(i)/10.0$

}

$x = x_0 + M(i)*G$

In the above algorithm, each change in line search magnitude creates a new collection of beamlet weights, which in turn updates every dose point in the dose matrix that is associated with an active objective. Voxels not associated with active objectives, since they do not affect cost function evaluation, are updated after the line search has completed, that is, at the same time that passive beamlets are updated. Any cost function evaluations that can be avoided will speed the optimization process.

Brent's Method

Brent's method is a widely used and efficient line search method for convex cost functions [Press *et al.*, 2002]. As mentioned, this method requires an estimate of the lower and upper bounds of the line search magnitude. The lower bound is always set to zero and the upper bound is estimated as discussed in the next section.

Brent's method is a combination of a parabolic interpolation search and a golden section search. Parabolic interpolation search is a well-adapted and efficient method for finding minima in functions that have a strong quadratic component. Recall that XiO dose objectives are a combination of linear, quadratic, as well as higher order components. In contrast, a golden section search is an extremely robust type of search that can find a minimum for any sort of convex function. The basic idea in Brent's method is to use parabolic search when it is advantageous and to use golden mean search when the parabolic search breaks down. The logic of how Brent's method switches between parabolic and golden section search is beyond the scope of this document.

As an introduction to the notion of searching for the minimum of a function, consider the binary search method that is often effective at finding the root of an function. Imagine that the root of an equation; that is, where the evaluation of the equation is zero, is bracketed in the interval (a,b). The process of finding a root of the function is efficiently implemented by the following algorithm:

S_{min} **Minimum step size**

if(f(a) > f(b)

D = 1 Direction of search

else Flips around the logic

D = -1

```
while(b – a >= Smin)
{
      c = (a + b)/2
      y = f(c)

      if(y * D > 0)
      {
            a = c
            c = (c + b)/2
```

```
        }
        else
        {
            b = c
            c = (a + c)/2
        }
}
```

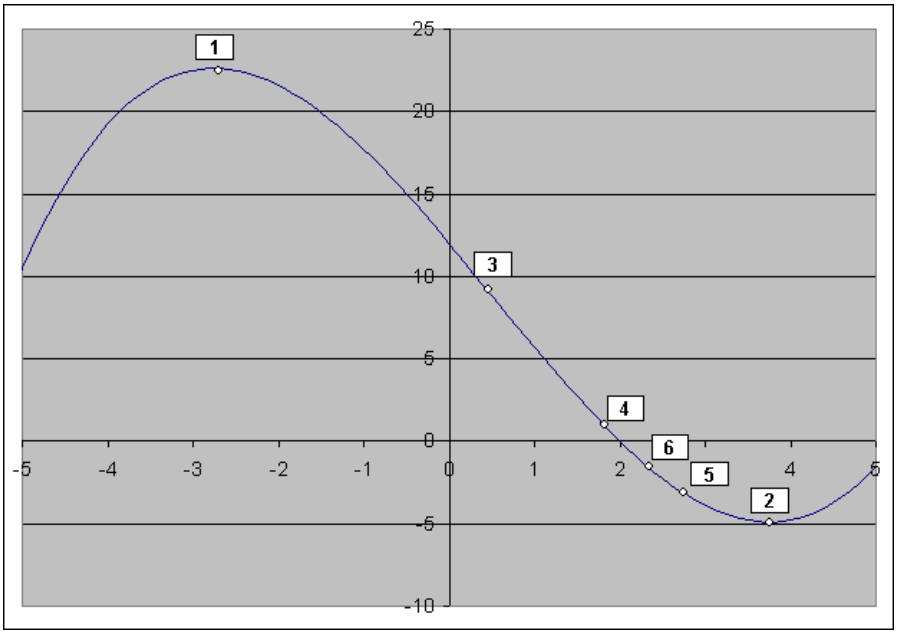


Figure 6-10 Binary function root search

Figure 6-10 shows a theoretical root search for six calculations of a function. Since calculation six is less than zero, calculation seven would be half way in between calculations four and six. The algorithm runs until the minimum step size, which controls the precision of the result, is exceeded.

The process of bracketing the minimum of a function is somewhat different, but related. A root of a function is bracketed by two points, *a* and *b* when the value of the function at the two points has different signs. By contrast, two points *(a,c)* bracket the minimum of a function when there is a triplet of points $a < b < c$ and *f(b)* is less than both *f(a)* and *f(c)*. Similar to bisection, we can take some new point *x*, either between *a* and *b,* or between *b* and *c,* and evaluate *f(x)*. If *f(x)* is between *f(a)* and *f(b)* and less than either, the bracketing points become *a* and *b*. If *f(x)* is between *f(b)* and *f(c)* and less than either, the bracketing points become *b* and *c*. The bracketing process should continue until the difference between the bracketing points is smaller than some tolerance. According to Press *et al,* this tolerance is the square root of the computer's floating point precision (e.g., 10^{-4} for floats and $3*10^{-8}$ for doubles). In practice, experimentation has shown that the precision for line searches can be limited to 10^{-5} with little effect on numerical results and totally insignificant clinical impact on final dose estimates. With this somewhat

lower accuracy, there are fewer cost function evaluations and the overall number of iterations and time to convergence is decreased.

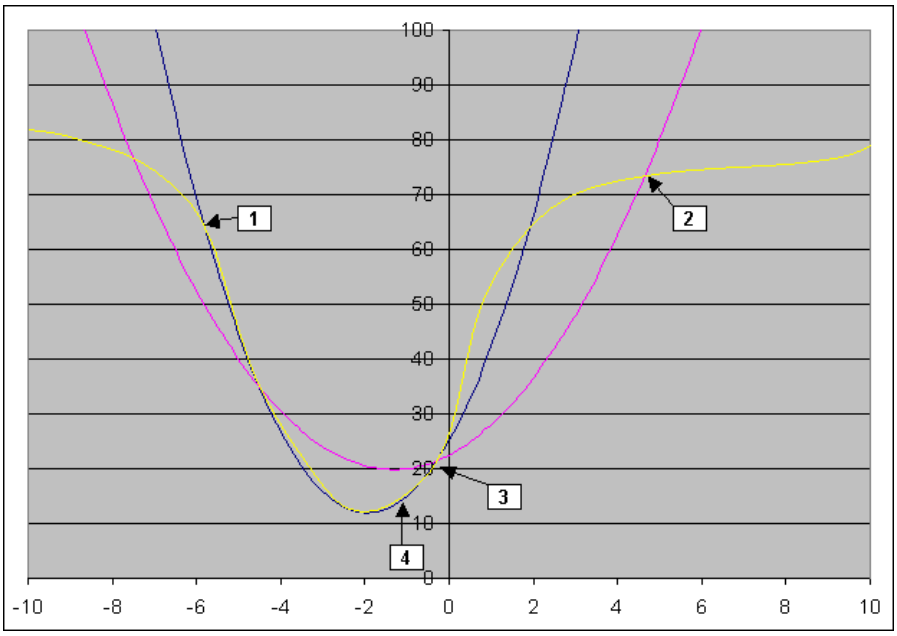


Figure 6-11: Parabolic interpolation search

Inverse parabolic search for the minimum of a cost function depends on the assumption that the function has a large parabolic component. This can be a reasonable assumption for convex search spaces. Figure 6-11 shows two steps from a parabolic interpolation search. The method is very simple: Given any three points known to exist on the function to be minimized (*i.e.,* the yellow line), find the quadratic form that goes through all three points (the pink line). Evaluate the function at minimum of the quadratic (point 4) and find a new quadratic form that passes through the converging triple of points (the blue line). The method continues until some convergence criterion is met or the golden section search is needed.

We have just seen an example of self similar search with respect to the binary method used for root finding, where each new interval was one-half the previous interval in size. The golden section search relies on the assumption that each iteration of a search for a minimum of a function is self similar [Press et al., 2002]; that is, the n + 1st iteration of the search should be in some sense proportional to the nth iteration. Suppose that *b* is a fraction of the distance between *a* and *c* and that a minimum of the function is known to exist between *a* and *c*, as shown in Figure 6-12.

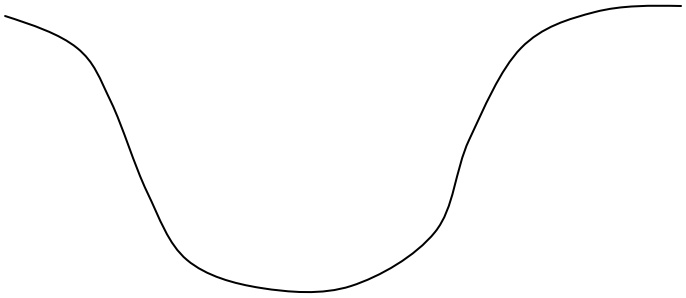


Figure 6-12: Golden section search

Suppose that the fractional distance between a and b is

$$\frac{b-a}{c-a} = w \qquad \textbf{(6-3)}$$

and between b and c is

$$\frac{c-b}{c-a} = 1-w \qquad \textbf{(6-4)}$$

If there were some new point *x* in the interval *(b,c)* the fraction of the distance between *x* and *b* would be

$$\frac{x-b}{c-a} = z \qquad \textbf{(6-5)}$$

Since we really could have chosen x first and b second, it makes sense that the fractional distance between the points should be the same. Equating the fractional distances would therefore limit the possible error in selecting either b or x as the next point to evaluate in finding a minimum of the function. Thus

$$w+z = 1-w \quad \textbf{or} \quad z = 1-2w \qquad \textbf{(6-6)}$$

In order that each iteration of the search to be self similar,

$$\frac{z}{1-w} = w \qquad \textbf{(6-7)}$$

must hold. Substituting Equations 6-6 into 6-7 yields

$$\frac{1-2w}{1-w} = w \qquad \textbf{(6-8)}$$

$$w^2 - 3w + 1 = 0 \qquad \textbf{(6-9)}$$

Solving using the quadratic formula yields

$$\frac{3-\sqrt{5}}{2} \approx 0.38197 \qquad \textbf{(6-10)}$$

This amazing number is found in art, music, and architecture and is known as the "golden ratio." At each stage of the search, given three points, the next point to select for evaluation is 0.38197 into the larger of

the intervals. This guarantees that the next interval will be no more than 1 – 0.38197 = 0.61803 times the size of the previous interval to search. Thus, the rate of convergence is not quite as fast as the binary root search, but it is never the less fast. The method continues until some convergence criterion is met or the parabolic interpolation search is preferred.

Line Search Estimation Heuristic

Figure 6-13 shows how the line search magnitude estimation heuristic works. The key to understanding the heuristic is to note that the first line search magnitude (that determined by the successive approximation algorithm) is always going to be the greatest, and that although not monotonically decreasing, line search magnitudes will tend to decrease as the optimizer converges.

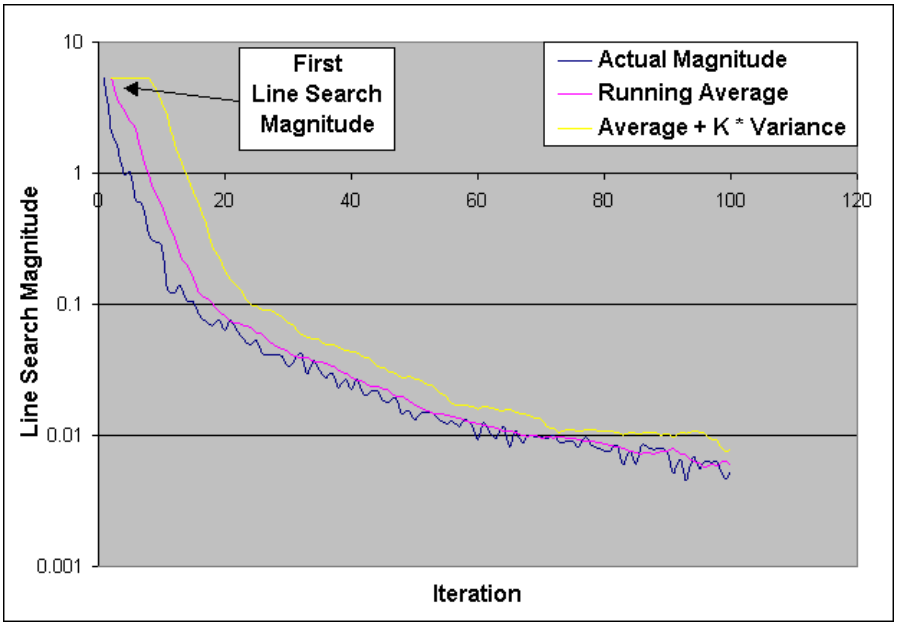


Figure 6-13: Line search magnitude estimation

The upper bound heuristic is calculated as follows:

$$U_i = \max(\overline{M}_i + K * \sigma_i, M_1),\ i > 1 \qquad \textbf{(6-11)}$$

Where:

U_i is the estimated upper bound line search magnitude for iteration *i*

M_1 is the actual line search magnitude for iteration 1

$\overline{M}_i$ is the running average of the last five line search magnitudes

σ_i is the running RMS error of the difference between the actual and

estimated line search magnitudes of the last five iterations.

K is a constant multiplier found empirically (the value "6")

Active and Passive Beamlets

As noted in earlier chapters, beamlets are of two types: active or passive. Active beamlets represent dose from photon fluence that is aimed at the target and is not blocked by multileaf collimator (MLC) leaves or any other treatment aid. Passive beamlets represent dose from photon fluence that is blocked by MLC leaves, blocks, or wedges. Active beamlet weights are updated during and between line searches, while passive beamlets are only updated between line searches. Active beamlets represent fluence and dose relationships for small units of the treatment field. Passive beamlets tend to represent fluence for much larger regions combining for efficiency large aggregations of related dose voxels. For any inverse planning problem, there will be perhaps many hundreds or thousands of active beamlets and only one or a few passive beamlets.

In section Optimization as a Search in a Space, we introduced the notion that optimization is a search in a space defined by a legal range of parameters. In inverse planning, the search space is a collection of legal beamlet weights. Obviously, final beamlet weights cannot be less than zero, which would imply the ability to deliver negative fluence, negative dose, etc. Leakage dose, related to minimum attainable MLC or compensating filter transmissions, similarly affects the legal search space. For beamlets representing fluence that can be blocked by MLC leaves, a minimum transmission of about 1-3% is often attainable. For compensating filters, a minimum transmission of only 10-20% may be attainable depending on the material of manufacture. XiO allows a minimum transmission of zero in the case that the user does not know or wish to take minimum transmissions into account during optimization. It is generally better to set these values.

Note that, during line searches in XiO IMRT, active beamlet weights are allowed to assume any value implied by a line search direction, even if that means beamlet weights are outside of the legal search space. However, dose, which affects the cost function evaluation, is never allowed to be negative. This allows the line search algorithm to fully explore a line search direction without reaping cost benefits for the illegal beamlet weights. An alternative is to penalize illegal beamlet weights during the line search process in an effort to avoid illegal beamlet weights entirely [Spirou and Chui, 1988; Alber, 2001]. The apparent disadvantage in penalization methods is that they may truncate searches in a particular line search direction. There have been no comparisons of the relative efficacy of these methods.

Between line searches, active beamlet weights that have attained illegal weights are clamped to the closest legal weight in the search space. The dose attributable to both clamped active beamlets and passive beamlets is estimated at this time. For each beam, the maximum beamlet weight is found. For clamped active beamlets with a non-zero minimum transmission, beamlet dose is proportional to the maximum beamlet weight times the minimum transmission. For passive beamlets, dose estimates are updated in a similar manner as clamped active beamlets with non-zero transmissions. For clamped active beamlets whose minimum transmission is zero, beamlet dose is set to zero

This updating of dose between line searches affects cost function evaluations, but usually in a very small way, and more often than not, the altered current cost is still better than the previous cost. If it is not better, the optimizer will reinitiate a line search using the gradient direction rather than the conjugate direction. When a new, presumably better cost has been found, the clamping process will be reinitiated. If the cost is still worse than the previous cost, the optimization process will halt.

Summary

Optimization in XiO inverse planning is the process of finding the minimum of a dose-based cost function in a multi-dimensional space that is defined by beamlet intensity weights. A conjugate gradient optimization algorithm, which is a special type of gradient descent optimizer, is used in XiO IMRT. Both conjugate gradient and gradient descent optimization methods iteratively calculate search directions in the multi-dimensional space of beamlet weights using line searches to find closer and closer approximations to a global minimum of the cost function in the search space.

The major difference between conjugate gradient optimizers and gradient descent optimizers is in the sequence of directions in the search space selected for the line searches. Gradient descent optimizers directly use the gradient of the cost function with respect to beamlet weights, which defines the steepest descent towards a minimum of the cost function. Ironically, direct use of the gradient may actually be inefficient at finding the minimum, because a sequence of gradient directions can actually interfere with each other. Conjugate gradient directions straighten out the path from start to finish. This straightening of the path leads to a good solution in fewer iterations of the line search. Mathematically, conjugate gradient search makes certain that each line search direction is *orthogonal* to the previous direction.

The convergence calculation in XiO IMRT is defined as the percent change of sequential cost function evaluations with respect to the first cost evaluation. Plotting cost function and convergence values both yield asymptotic curves. It is typical for almost all improvement in cost to be completed in the first thirty or so iterations of the optimizer. When higher order penalties are used in objectives (*i.e.,* higher than 2.0), convergence to cost function evaluations near asymptotic limits can require more iterations. When using only minimum and maximum dose objectives, the cost function is a convex function, which guarantees that there is a single minimum in the search space, and convergence to that minimum, given enough iterations. When dose-volume objectives are introduced into the cost function, it potentially becomes non-convex, which means that there can be local minima in the search space. Studies have shown that local minima introduced by dose-volume objectives do not result in clinically inferior optimized doses.

XiO IMRT uses two different line search algorithms. The first is a relatively slow but robust algorithm used to determine the proper range of line search magnitudes that will be used by the second algorithm. The second algorithm is very fast and robust; it is known as Brent's method and it is a combination of a parabolic interpolation search and a golden section search. One of the keys to the overall speed of XiO IMRT optimization is the line search magnitude heuristic that intelligently limits the range of distances in any line search direction that will be searched for a minimum of the cost function. By limiting the search, fewer calculations are required to find the minimum.

7. Pencil Beam Algorithm

XiO uses a variation on the notion of the pencil beam algorithm for calculating beamlet doses. The method allocates photon dose calculated with one of the XiO three-dimensional dose algorithms among overlapping beamlets and beamlet scatter tails. The dose algorithm used for making beamlets for a beam is the one that is selected by the user for that beam. A combination of dose calculation algorithms can be used to produce beamlets for a plan. Note that even though the base dose calculation may have been previously commissioned and used in conventional treatment planning, some extra data collection and algorithm commissioning may nevertheless be required due to issues in small field dose estimation that are unique to IMRT. For more information see the Beam Modeling Guide.

Allocation of dose among beamlets is done using an analytic formula that has been shown to produce self-consistent "realistic" pencil beam profiles [Alber, 2001]. This is the same formula that was used to produce figures in *Beamlets*. Although the dose profiles are calculated using much of the mechanics of the convolution and superposition algorithms, what the profiles yield are relative lateral dose distributions for each beamlet.

The XiO pencil beam beamlet calculation has the favorable property that it reproduces the characteristics of its base photon calculation. For instance, if the base algorithm is a superposition calculation, the general dose characteristics of the superposition calculation will be reproduced by the beamlets. That is, dose will decrease in regions of low electron density and increase in regions of high electron density. Like other pencil beam algorithms, however, it does not model the stretching out of lateral doses in low electron density regions or the contraction of lateral doses in high electron density regions.

Lateral Dose Distributions

The analytic formula that defines lateral dose distributions in the "x" direction is

$$x < -x_0 \qquad f(x,b,x_0) = \sinh(bx_0)\exp(bx) \qquad \textit{(7-1)}$$

$$-x_0 \le x \le x_0 \qquad f(x,b,x_0) = 1 - \exp(-bx_0)\cosh(bx_0) \qquad \textit{(7-2)}$$

$$x_0 < x \qquad f(x,b,x_0) = \sinh(bx_0)\exp(-bx) \qquad \textit{(7-3)}$$

The formula for lateral distributions in the "y" direction is identical and, as employed by XiO, b and half width values (*i.e.,* x_0, y_0) may be independently specified.

Figure 7-1, which duplicates Figure 4-2, shows beamlet profiles (relative dose distributions) for a fixed value of b and different half widths. Figure 7-2 shows beamlet profiles for a fixed half width of 5 mm and different values of b. Half width values in XiO are defined as the projection of the leaf width and leaf resolution at isocenter. b values are automatically estimated by XiO according to a method that seeks to minimize the error between dose profiles calculated by the FFT convolution photon dose calculation algorithm and the pencil beam photon dose calculation for different field sizes.

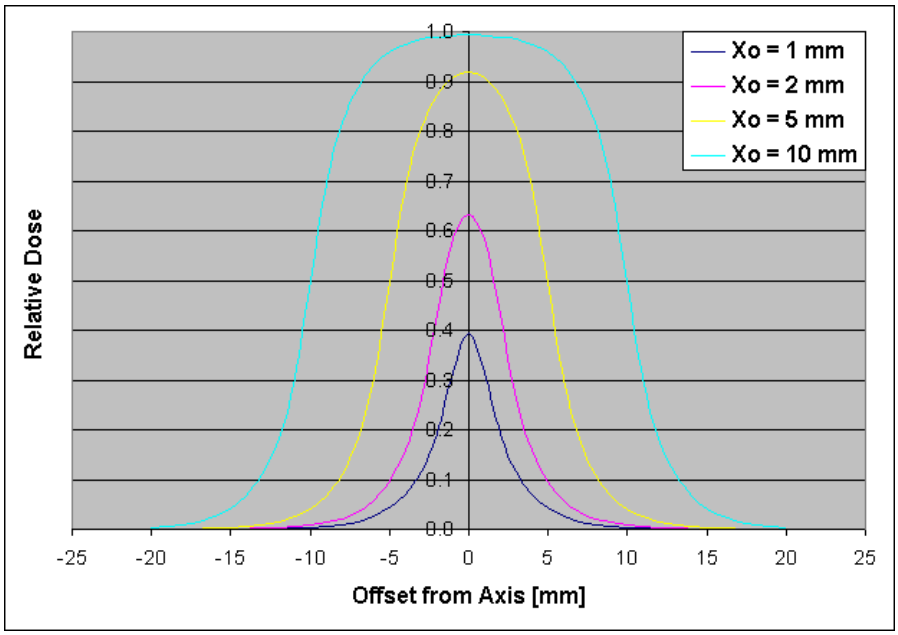


Figure 7-1: Beamlet profiles for* b *= 0.5 and different half widths

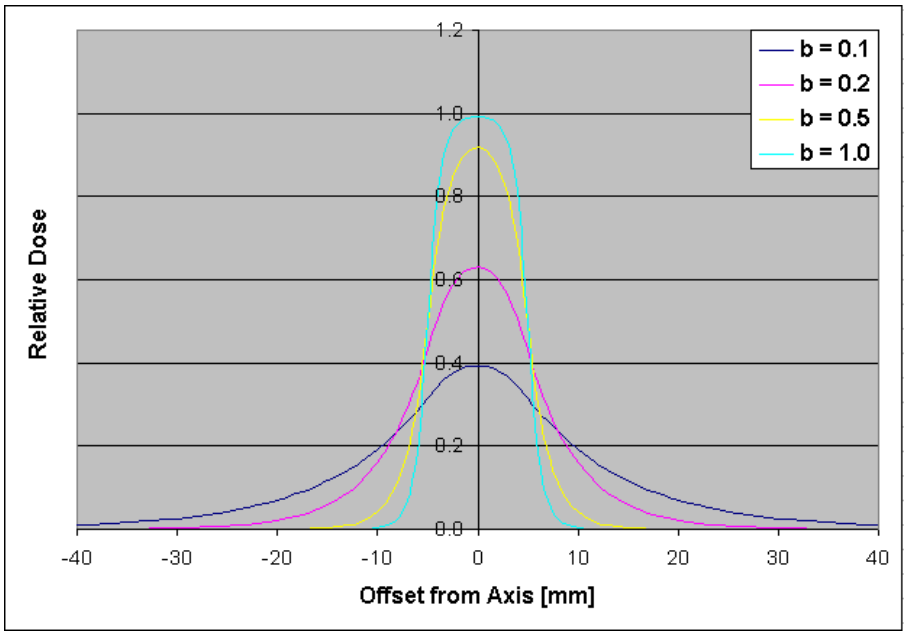


Figure 7-2: Beamlet profiles for half width = 5mm and different* b *values

Lateral dose distributions for each beamlet are calculated using Equation 7-1.

$$dose(x, y, d) = TERMA(x_p, y_p, d) \cdot f(x_r, b_x, x_0) \cdot f(y_r, b_y, y_0) \quad \textit{(7-4)}$$

Where

$TERMA(x_p, y_p, d)$	is the relative total energy released per unit mass from photon fluence at the point (x_p, y_p, d), which lies on the central axis of the beamlet at depth *d*. This energy release is calculated exactly as it would be for the convolution or superposition algorithms. A further explanation of TERMA and the related quantity KERMA can be found in Khan, 2003.
$f(x_r, b_x, x_0)$	is the cross profile formula in Equation 7-1 for the "x" direction
$f(y_r, b_y, y_0)$	is the cross profile formula in Equation 7-1 for the "y" direction

Figure 7-3 shows geometrically how lateral dose distributions are calculated. Notice that as the depth, *d*, changes, the magnification of the distributions change. Also notice that there is no tilting of the distribution for beamlets that are away from the central axis of the beam. The fact that the distribution is not tilted leads to very small errors in allocation of dose to beamlets.

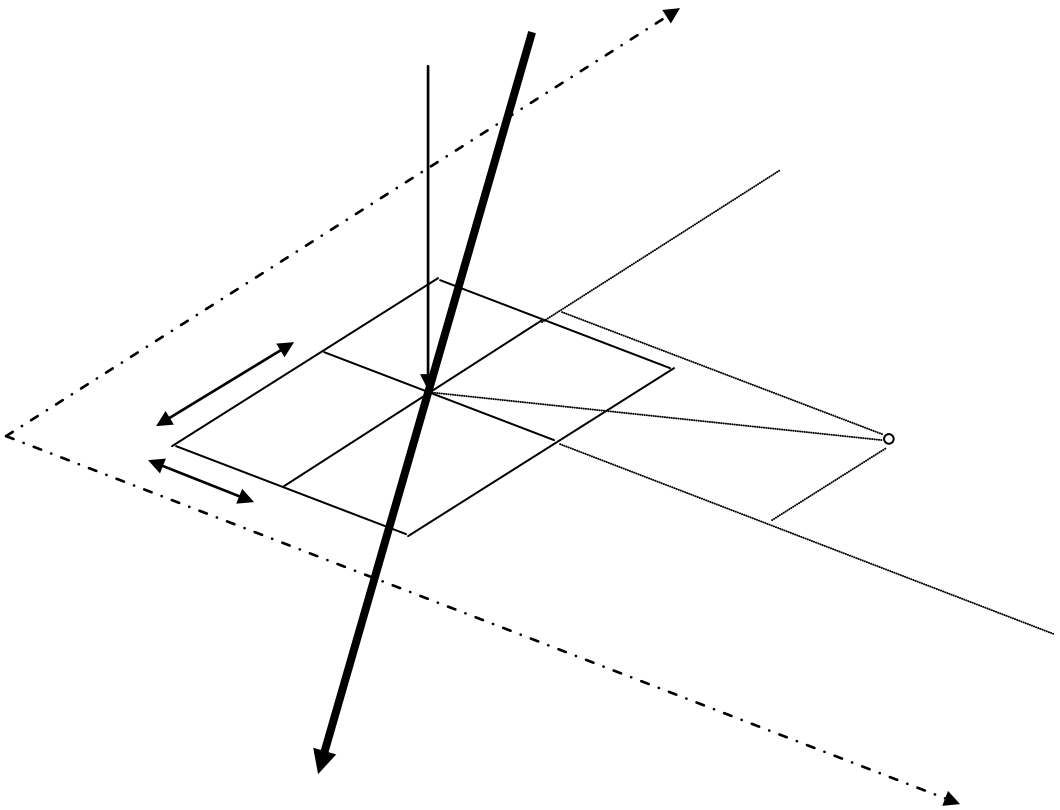


Figure 7-3: Pencil beam dose calculation geometry

Figure 7-4 shows a possible plot of TERMA along the central axis of a beamlet. TERMA is calculated according to the formulas.

$$\Psi(d) = \Psi_0 \cdot e^{-\frac{\overline{\mu}}{\rho} d}$$

$$TERMA = \frac{\overline{\mu}}{\rho} \cdot \Psi(d)$$

where

$\Psi(d)$	is the photon energy fluence at depth d
Ψ_0	is the incident photon energy fluence

$\overline{\frac{\mu}{\rho}}$ is the mass attenuation coefficient for the medium averaged over the photon energy fluence spectrum.

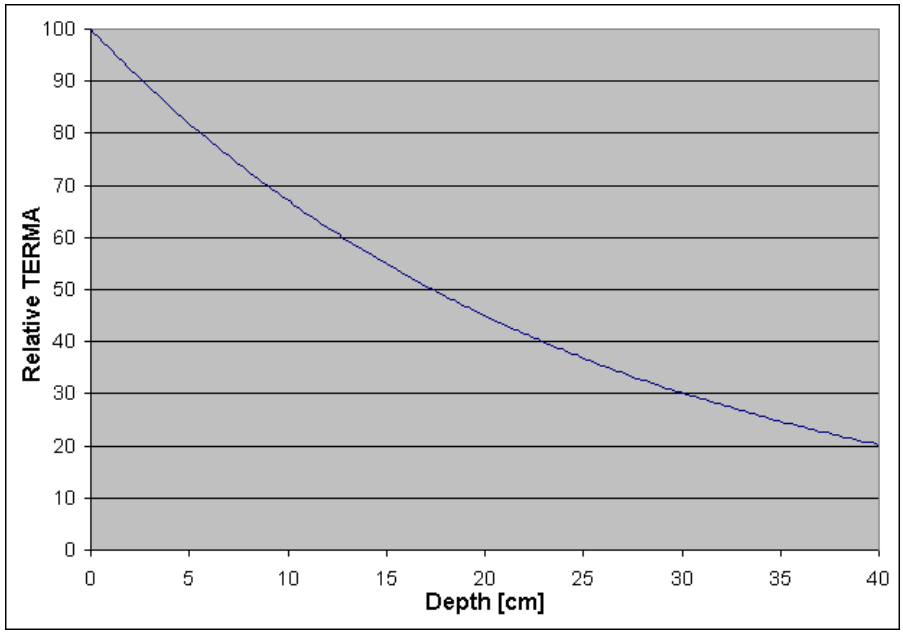


Figure 7-4 TERMA along the central axis of a beamlet

Combining the TERMA in Figure 7-4 with the lateral dose distribution shown in Figure 7-5, an intensity profile for a plane cut through the central axis of a beamlet can be calculated as shown in Figure 7-6.

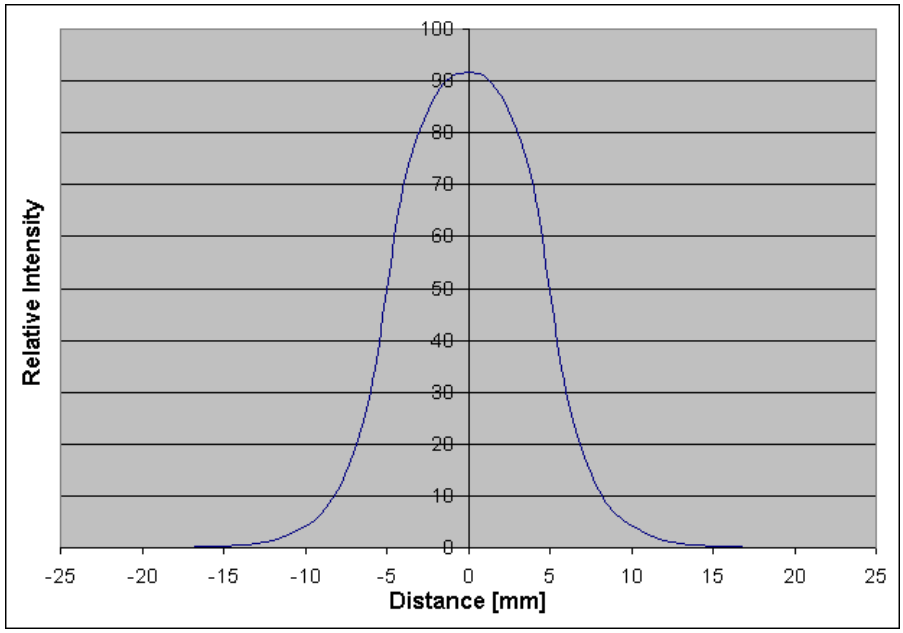


Figure 7-5: Beamlet profiles for half width = 5mm and* b *= 0.5

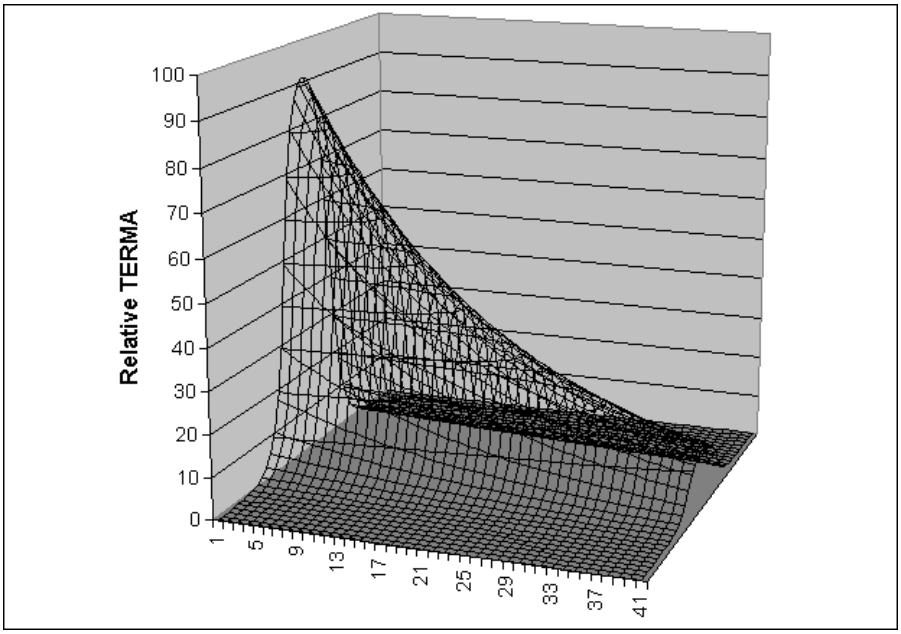


Figure 7-6: Beamlet plane TERMA intensity profile (not to scale)

Beamlet Calibration

The combined lateral dose distributions of all beamlets for a beam calculated using TERMA, as described above, is a very simple dose estimate that does not take into account photon and electron transport or the presence of electron contamination in the beamline of a linear accelerator. It has no buildup region and the dose at maximum is incorrect, among other things. In order to account for these phenomena, dose for each beam is calculated using one of the three-dimensional photon dose calculation algorithms available in XiO and this dose is used to *calibrate* the beamlet lateral dose distributions. After calibration, the combined beamlet dose exactly mimics the open field photon dose, accounting for transport phenomena and electron contamination.

All of the XiO three-dimensional dose calculation algorithms convert TERMA to dose by convolving it with a dose deposition kernel. Figure 7-7 shows a two-dimensional slice through the central axis of a dose deposition kernel produced by XiO for use with the FFT convolution algorithm. Note that this kernel is oriented as if the beam were entering a medium from the left. Dose calculation kernels, as shown in Figure 7-7, are essentially dose deposition probability density functions that are derived using Monte Carlo methods [Mackie et al., 1988]. That is, they describe the probability that energy released at some interaction point in a medium (shown in red in the figure) will be deposited as dose at some other point in the medium. Since it is a probability density function, kernels should sum to 1.0 or very near it. In more than a few cases, kernel "energies" will sum to 0.95 or so, because of energy that is deposited as dose outside of the region specified by the kernel's three-dimensional boundaries.

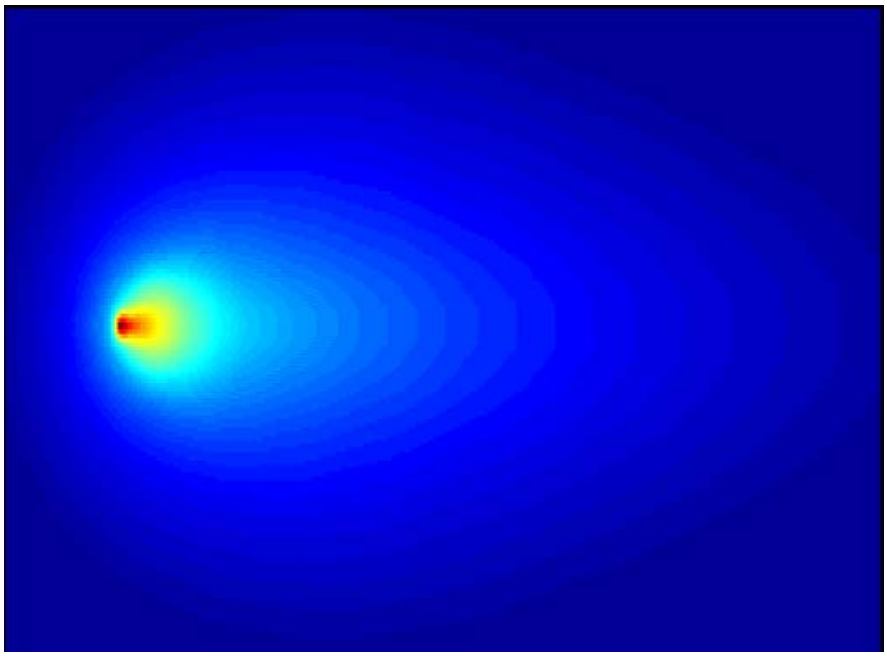


Figure 7-7: 18 MV dose kernel produced by XiO

Figure 7-8 is a schematic representing a one-dimensional intensity profile through the central axis of a kernel. It is made in the shape of a kernel profile, but not based on any measurements or first principles of physics. If we were to convolve the kernel in Figure 7-8 with the TERMA in Figure 7-9, we would arrive at the dose in the same figure. This calculation was, in fact, done using the same popular spreadsheet program that has been used to produce all of the graphs in this document. Figure 7-10 shows what the intensity profile of a dose-calibrated beamlet might look like using the "central axis" dose from Figure 7-9 (compare to Figure 7-6).

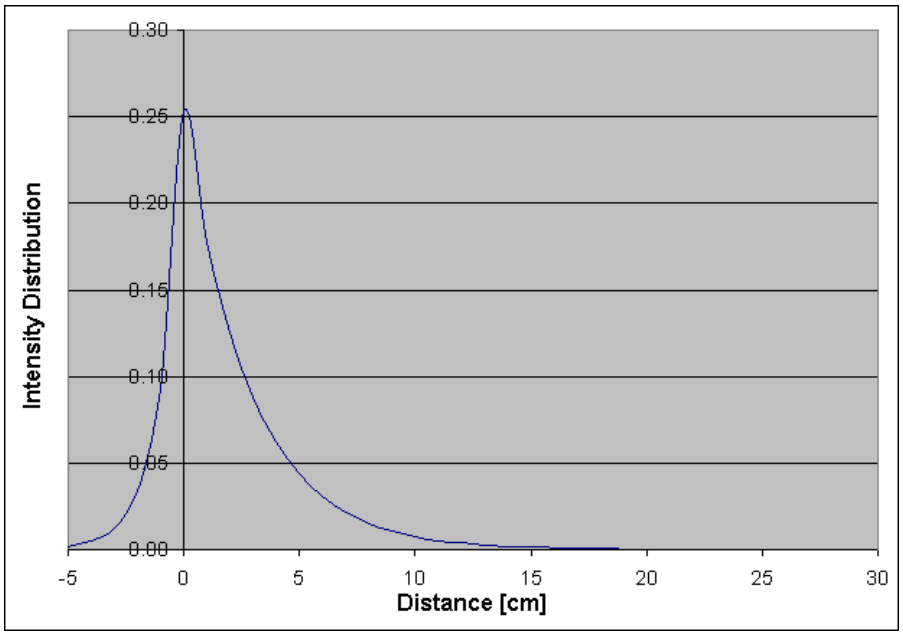


Figure 7-8: Kernel Central Axis Profile

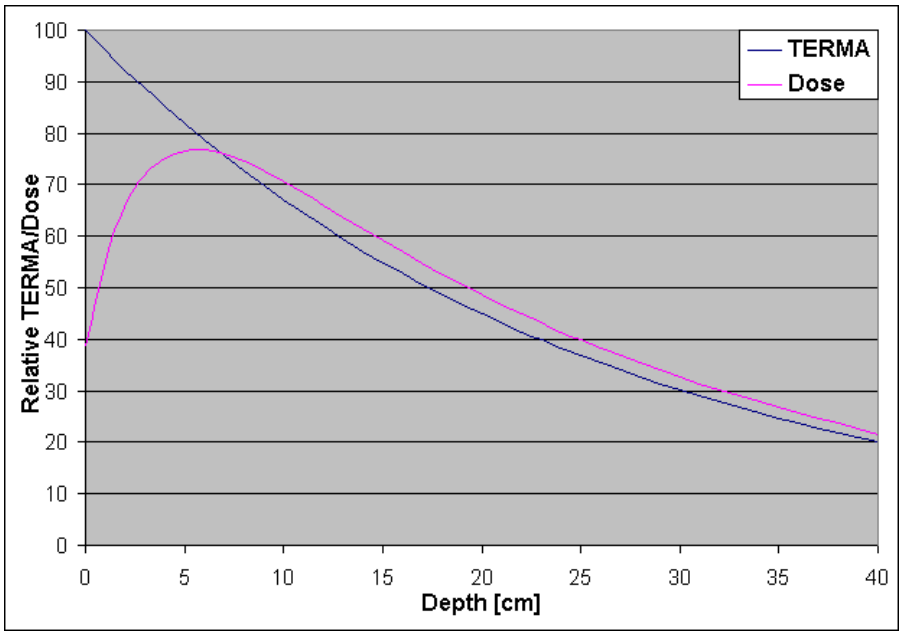


Figure 7-9: TERMA and Dose

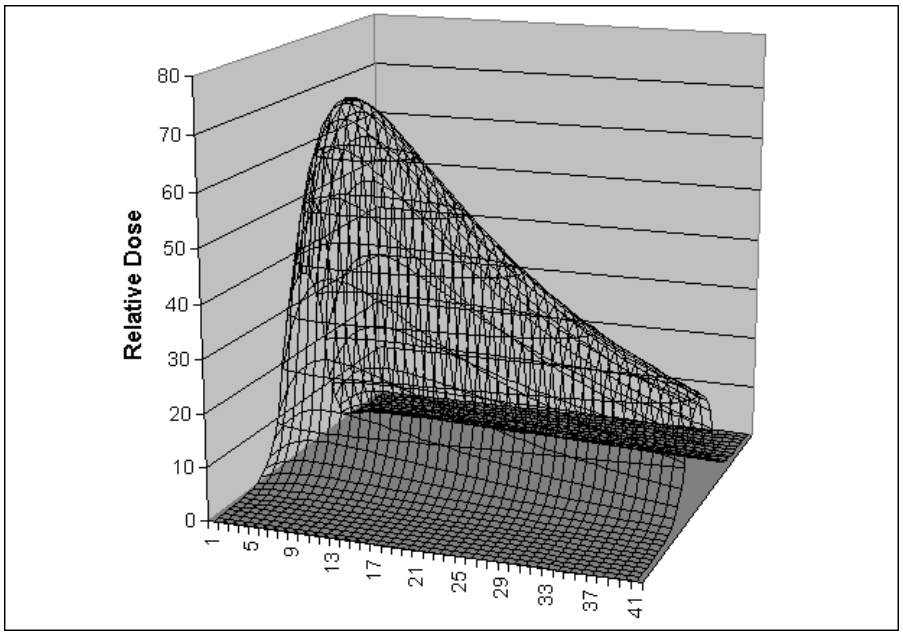


Figure 7-10: Beamlet plane dose intensity (not to scale)

Suppose that the dose in Figure 7-9 included electron contamination so we could use it to calibrate the TERMA-weighted lateral dose distribution of a beamlet. In the case of a single non-overlapping beamlet, we would simply set the TERMA-weighted lateral dose distribution point by point to the dose distribution. When many overlapping beamlets are involved, as shown in Figure 7-11, the beamlet lateral dose distributions are used to weight the contribution of dose to each beamlet.

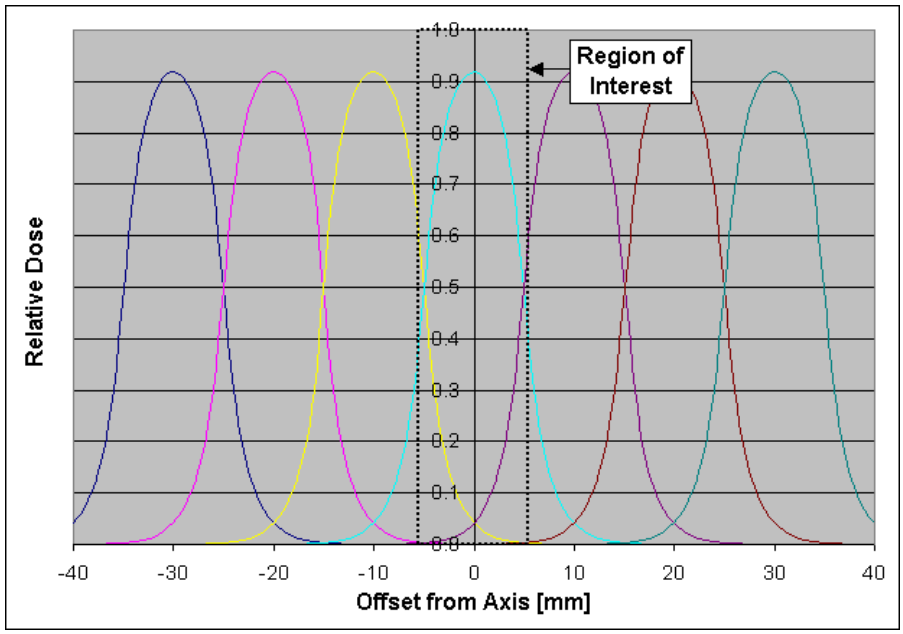


Figure 7-11: Calibration of many overlapping beamlets

Figure 7-12 shows the magnification of the region of interest from Figure 7-11. Suppose that it were possible to calibrate beamlets on a 1 mm calculation grid. The numbers associated with each of the lateral profiles show the fraction of each dose point calculated by a three-dimensional photon dose algorithm that would be allocated for each of the three beamlets along the line in the calculation volume that the figure represents.

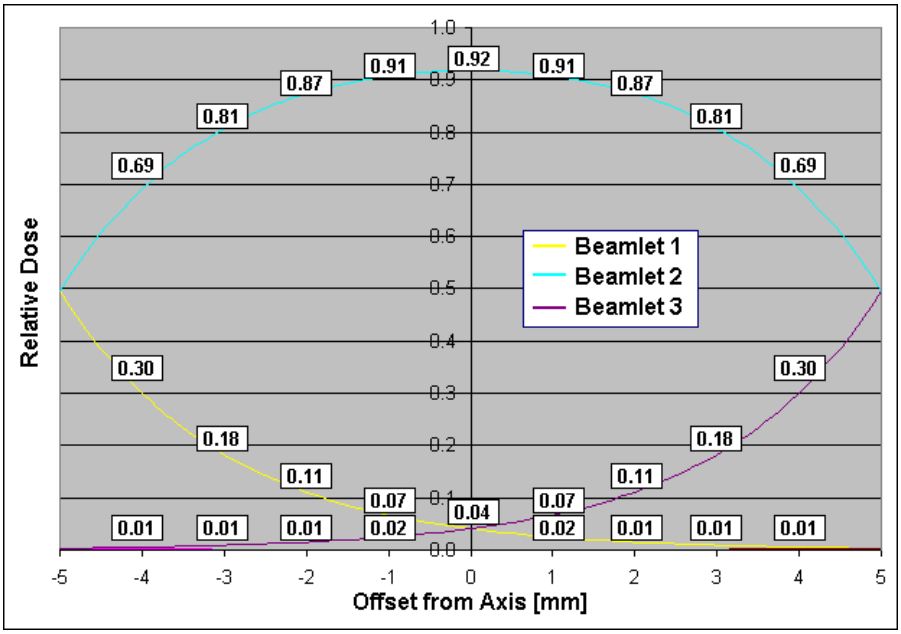


Figure 7-12: Magnification of detail of beamlet calibration

Figure 7-13 is meant to represent beam penumbra dose that might be calculated by one of XiO's photon dose calculation algorithms.

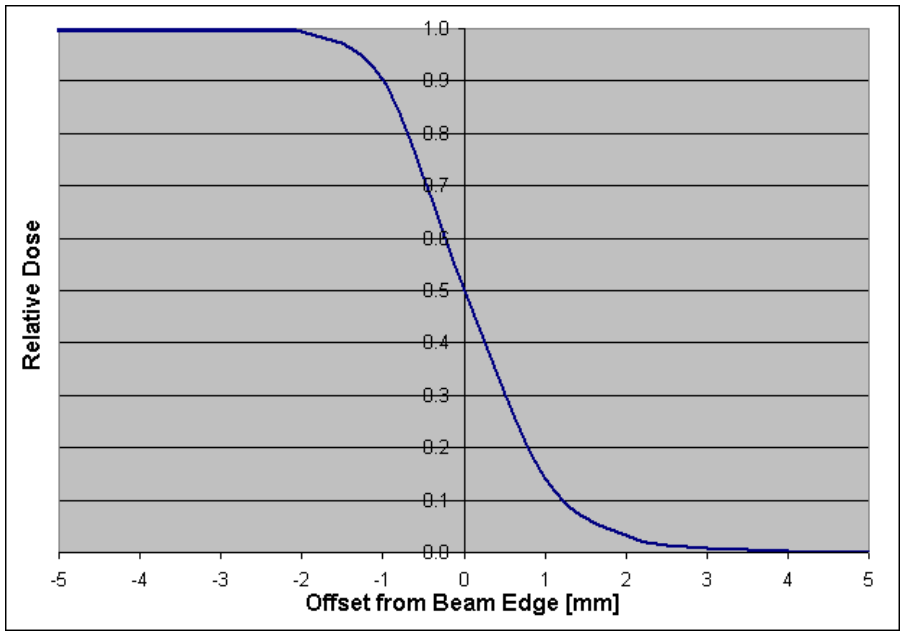


Figure 7-13: Beam edge photon dose profile

If we multiply the photon dose in Figure 7-13 by the beamlet fractional lateral distributions in Figure 7-12 we get the calibrated beamlet doses as shown in Figure 7-14

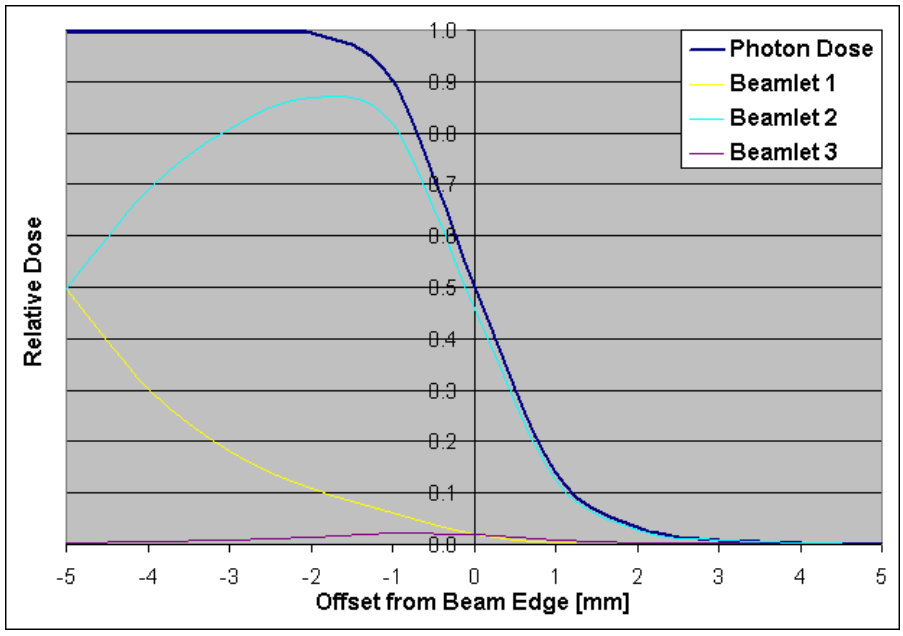


Figure 7-14: Beamlets calibrated by photon dose

Summary

XiO IMRT creates beamlets using a unique implementation of the pencil beam algorithm that is calibrated using three-dimensional dose calculation algorithms, such as Superposition, Fast

Superposition, or FFT Convolution. As such, beamlets in XiO IMRT have the advantage of being able to model dose in heterogeneous tissue in a way that is superior to pencil beam beamlet doses in other treatment planning systems. The lateral distributions of the XiO beamlets are described by an analytic function that allows variable beamlet half widths and variable length scatter tails, allowing it to model a variety of beamlet dose profiles. Like other pencil beam algorithms, however, it does not model the stretching out of lateral doses in low electron density regions or the contraction of lateral doses in high electron density regions.

8. Bibliography

Alber, ML, *A Concept for the Optimization of Radiotherapy*, Doctoral Thesis, University of Tuebingen, 2001.

Bar W, Alber M, Nusslin F. A variable fluence step clustering and segmentation algorithm for step and shoot IMRT. Phys Med Biol. 2001 Jul;46(7):1997-2007.

Bortfeld T, Stein J, Preiser K, Clinically relevant intensity modulation optimization using physical criteria, *XII International Conference on the Use of Computers in Radiation Therapy* Medical Physics Publishing: Madison (1997) 1-4

Ezzell GA, Galvin JM, Low D, Palta JR, Rosen I, Sharpe MB, Xia P, Xiao Y, Xing L, Yu CX; Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee. Med Phys. 2003 Aug;30(8):2089-115.

Deasy JO. Multiple local minima in radiotherapy optimization problems with dose-volume constraints. Med Phys. 1997 Jul;24(7):1157-61.

Gersem W, Claus F. An anatomy-based beam segmentation tool for intensity-modulated radiation therapy and its application to head-and-neck cancer.
Int J Radiat Oncol Biol Phys. 2001 Nov 1;51(3):849-59.

ICRU, Prescribing, recording, and reporting photon beam therapy. Report 50, Bethesda, International Commission on Radiation Units and Measurements, 1993.

ICRU, Prescribing, recording, and reporting photon beam therapy (Supplement to ICRU Report 50). Report 62, Bethesda, International Commission on Radiation Units and Measurements, 1999.

Khan, FM. *The Physics of Radiation Therapy*, 3rd Edition, Lippincott Williams & Wilkins, Philidelphia, 2003.

Mackie TR, Bielajew AF, Rogers DW, Battista JJ. Generation of photon energy deposition kernels using the EGS Monte Carlo code. Phys Med Biol. 1988 Jan;33(1):1-20.

Pirzkall A, Carol MP, Pickett B, Xia P, Roach M 3rd, Verhey LJ. The effect of beam energy and number of fields on photon-based IMRT for deep-seated targets.
Int J Radiat Oncol Biol Phys. 2002 Jun 1;53(2):434-42.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP, *Numerical Recipes in C++: The Art of Scientific Computing*, 2nd Edition, Cambridge University Press, Cambridge, 2002.

MathWorld http://www.mathworld.wolfram.com/

Nocedal J, Wright SJ, *Numerical Optimization*, Springer, New York, 2000.

Shepard DM, Earl MA, Li XA, Naqvi S, Yu C. Direct aperture optimization: a turnkey solution for step-and-shoot IMRT. Med Phys. 2002 Jun;29(6):1007-18.

Siebers JV, Lauterbach M, Keall PJ, Mohan R. Incorporating multi-leaf collimator leaf sequencing into iterative IMRT optimization. Med Phys. 2002 Jun;29(6):952-9.

Spirou SV, Chui CS. A gradient inverse planning algorithm with dose-volume constraints. Med Phys. 1998 Mar;25(3):321-33.

Spirou SV, Fournier-Bidoz N, Yang J, Chui CS, Ling CC. Smoothing intensity-modulated beam profiles to improve the efficiency of delivery. Med Phys. 2001 Oct;28(10):2105-12.

Wiesmeyer M. Effective cost functions as a means of improving user interaction in unconstrained gradient-based inverse planning, *XIV International Conference on the Use of Computers in Radiation Therapy* (2004).

Wu Q, Mohan R. Multiple local minima in IMRT optimization based on dose-volume criteria. Med Phys. 2002 Jul;29(7):1514-1527.

Zakarian C, Deasy JO, Beamlet dose distribution compression and reconstruction using wavelets for intensity modulated treatment planning. Med Phys. 2004 Feb;31(2):368-75.

13723 Riverport Drive, Suite 100, Maryland Heights, MO 63043

Manufacturer

Elekta Business Area Software Systems

IMPAC Medical Systems, Inc.

13723 Riverport Drive

Suite 100

Maryland Heights, MO 63043

USA

Phone: +1.800.878.4267

Fax: +1 314 812 4491

European Union Authorized Representative

EC REP

Elekta Limited

Linac House, Fleming Way

Crawley, West Sussex

RH10 9RR,

United Kingdom

Phone: +44 129 365 4242

Fax: +44 1293 471347

eu_ar@elekta.com