

Lecture 36

Intensity-Modulated Radiation Therapy

36.1 Lecture Objectives

- Recognize the basic components of Intensity-Modulated Radiation Therapy (IMRT).
- Identify the fundamental tradeoffs in IMRT, particularly between applying sufficient dose to the target volume (cancer), while minimizing the damage to other healthy tissues in the vicinity.
- Formulate and solve simple IMRT treatment planning optimization problems.

36.2 Background

A common way to deliver radiation therapy is so-called “conformal radiotherapy”. In 3D conformal radiotherapy, the tumor is treated with one radiation field from each beam direction. For each direction, the shape of the beam is the projection of the tumor volume along that direction. Importantly, conformal radiotherapy is highly limited in the shape of the target volume. For example, it is not able to target non-convex volumes. Consider a spherical target volume, but with a smaller spherical volume inside, which we wish to spare from radiation. Conformal radiotherapy would not be able to apply dose to the target sphere while sparing the inner sphere. With the advent of computed tomography (CT), which allows accurate delineation of the target volume in 3D, and the multi-leaf collimator, which allows improved shaping/modulation of the radiotherapy beam for each gantry angle, Intensity-Modulated Radiation Therapy (IMRT) emerged as an improved therapy delivery approach. IMRT is described briefly in this chapter, with a focus to its connection to mathematical optimization.

36.3 Overview of IMRT

36.3.1 Components of IMRT

IMRT methods modulate the fluence distribution in the plane perpendicular to the incident beam. For this purpose, the incident beam is partitioned into a set of small beam segments (also known as beamlets or bixels). Through the use of a multi-leaf collimator (MLC), the fluence can be controlled for each of these beamlets. The effect of the MLC is illustrated in Figure 36.1. Importantly, IMRT enables the delivery of dose distributions with much more sophisticated shapes compared to conformal radiotherapy, including non-convex shapes. In IMRT, the goal of therapy planning is to determine the fluence for each beamlet and each gantry angle that produce the best possible dose distribution in the patient. Thus, the connection to optimization problems is important and direct.

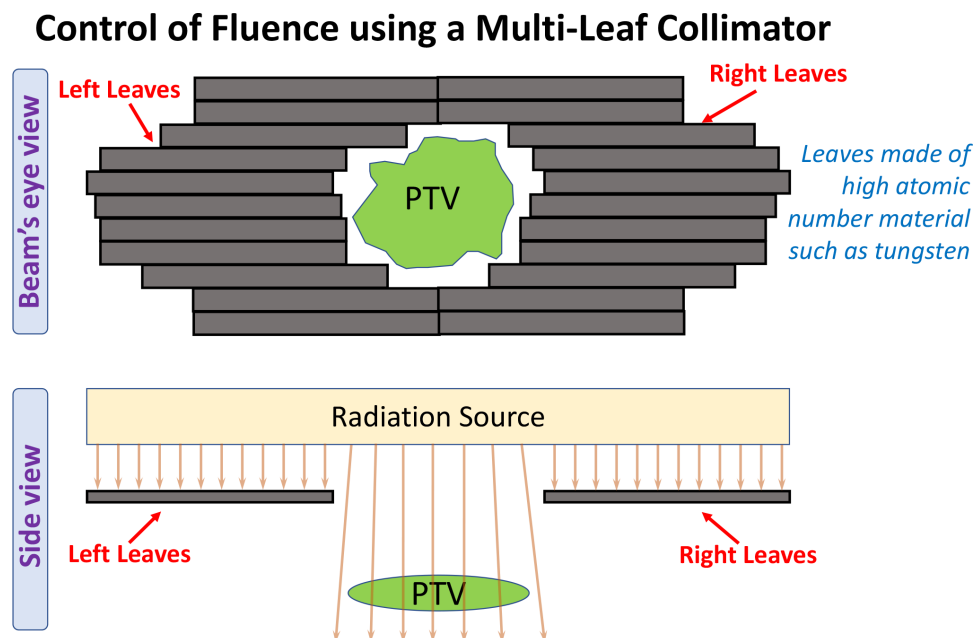


Figure 36.1: In IMRT, the fluence transmitted from the source to the patient can be controlled using a multi-leaf collimator (MLC). The MLC has multiple leaves made of a material with high atomic number, which largely block the radiation. By designing the position of each of the leaves in the MLC, as well as the timing, one can design the pattern of radiation beams. By designing the position of the MLC leaves at each of the gantry angles, as well as the corresponding timings, highly conformal dose distribution are feasible. In IMRT, the therapy planning can be performed in several steps, by first solving the “fluence optimization” problem (i.e., determining the desired fluence from each beamlet at each gantry angle), and then solving the “leaf sequencing” problem to determine the positions and timings of the MLC leaves in order to achieve the desired fluence pattern.

36.3.2 The dose deposition matrix

The dose deposition map associated with a specific fluence map can be written (under the corresponding linearity and discretization assumptions) as a matrix-vector operation. Let us introduce some notation:

- $\mathbf{x}$ is the vector containing the fluence map, i.e., the fluence for each beamlet and each gantry angle.
- $\mathbf{d}$ is the vector containing the resulting dose map, i.e., the dose deposited in each voxel within the patient.
- $\mathbf{D}$ is the dose deposition matrix, which relates $\mathbf{x}$ and $\mathbf{d}$ as follows:

$$\mathbf{d} = \mathbf{D} \mathbf{x}$$

In other words, each row m of $\mathbf{D}$ describes the contributions of all the beamlets to voxel m in the patient: $d_m = \sum_n D_{m,n} x_n$. Similarly, each column n of $\mathbf{D}$ describes how the fluence for the n th beamlet is distributed throughout the voxels in the patient. Note that the matrix $\mathbf{D}$ is generally highly sparse (i.e., most of its entries are zero) since each beamlet will generally lead to dose deposition in a small subset of voxels.

- Since dose is generally quantified in units of Gray (Gy)¹, and fluence is quantified in monitor units (MU), the dose deposition matrix values have units of Gy/MU.

The relationship between fluence at each beamlet and dose deposition at each voxel is illustrated graphically in Figure 36.2.

36.3.3 The fluence optimization problem

The determination of optimal fluence map constitutes an optimization problem with the following features:

- Seek a dose distribution $\mathbf{d} = \mathbf{D} \mathbf{x}$ that leads to a desired dose (e.g., 60 Gy) in the target volume, and low dose in healthy tissues.
- Satisfy physical constraints on the fluence map (at the very least, positivity with $x_n \geq 0$ for each beamlet n).
- Satisfy hard constraints in the dose distribution, such as $g_s(\mathbf{d}) \leq c_s$, where g_s is a function of the voxels included in some structure s , and c_s is the corresponding upper dose limit. For example, g_s may correspond to the maximum dose deposited over all the voxels in a certain structure s . Alternatively, g_s may correspond to the mean of the dose deposited over all the voxels in s . The choice of function g_s depends on the type of organ at risk captured by structure s and the corresponding mechanisms of radiation-induced damage.

¹A Gray is defined as the absorption of one joule of radiation energy per kilogram of matter

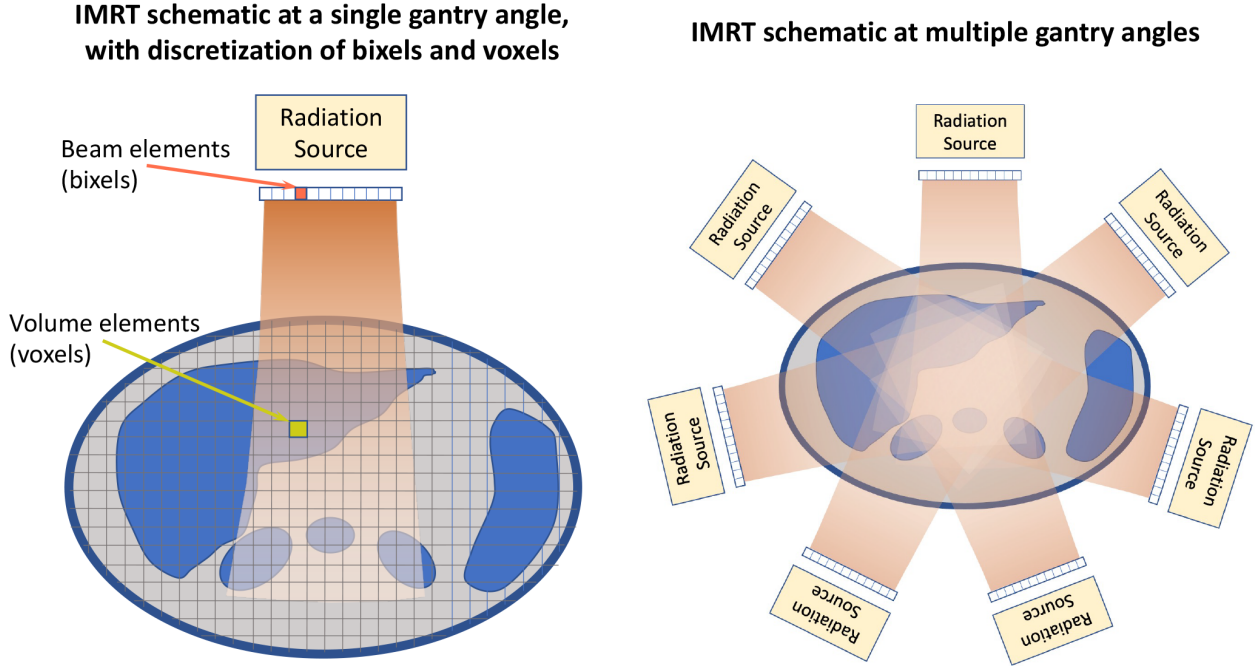


Figure 36.2: Schematic depiction of intensity modulated radiation therapy (IMRT), shown in 2D for simplicity. By controlling the fluence provided by each beam element (bixel) at each gantry angle, highly conformal dose deposition can be achieved. Note that the modulation of the fluence at each bixel is not depicted in this diagram, for simplicity of display.

Based on these elements, the fluence optimization problem can be posed as follows:

$$\begin{aligned}
 & \arg \min_{\mathbf{x}} \quad f(\mathbf{d}) \\
 & \text{subject to} \quad \mathbf{d} = \mathbf{D} \mathbf{x} \\
 & \quad \quad \quad g_s(\mathbf{d}) \leq c_s \quad (\text{for each relevant structure } s) \\
 & \quad \quad \quad x_n \geq 0
 \end{aligned} \tag{36.1}$$

In this optimization problem, the cost function $f(\cdot)$ penalizes deviation from the desired dose distribution:

$$f(\mathbf{d}) = \sum_m w_m f_m(d_m) \tag{36.2}$$

where $w_m \geq 0$ are positive weights, and $f_m(d_m)$ penalize deviation from the desired dose at voxel m .

For an organ we wish to spare from dose, the penalty functions $f_m(d_m)$ may take the form:

$$f_m(d_m) = (d_m - d_m^{max})_+^2 \tag{36.3}$$

$$= \begin{cases} (d_m - d_m^{max})^2 & , \text{ if } d_m > d_m^{max} \\ 0 & , \text{ else} \end{cases} \quad (36.4)$$

where our desire to spare the organ from dose is specified through a maximum dose d_m^{max} . Similarly, for a voxel m within a target, the penalty functions $f_m(d_m)$ may take the form:

$$f_m(d_m) = (d_m^{min} - d_m)_+^2 \quad (36.5)$$

$$= \begin{cases} (d_m^{min} - d_m)^2 & , \text{ if } d_m < d_m^{min} \\ 0 & , \text{ else} \end{cases} \quad (36.6)$$

where our desire to apply dose to the target is specified through a minimum dose d_m^{min} .

For some of the illustrative examples and exercises in this course, we will design our treatment plans using a basic quadratic penalty for mismatch with the desired dose $f_m(d_m) = (d_m - d_m^{desired})^2$, for simplicity. In this case, $d_m^{desired}$ may correspond to a high value (60 Gy) in the target, and zero in the remaining tissues. As an illustration, our desire to protect specific organs may be driven simply by the choice of w_m , with large w_m for organs we particularly would like to protect, and smaller w_m for organs that are less critical to protect from dose. As an alternative (or complement) to this weight-based approach, we may impose hard constraints on dose deposition in specific organs.

As can be seen, there are a large number of tunable parameters in this formulation. The details of the specific choices for various applications are beyond the scope of this course, but may be generally summarized as: *maximizing the tumor control probability, while minimizing the normal tissue complication probability*. For more details, the reader is referred to the book “Treatment planning for radiation oncology” (fifth edition, Sperduto and Gibbons, editors), and specifically chapter 25 on IMRT by Jan Unkelbach. Much of the material for this lecture was obtained from this chapter.

36.3.4 Two-pixel example

Some of the concepts described in this lecture can be nicely viewed with a small, two-voxel example, as depicted in figure [36.3](#). In this case, we have a “patient” consisting of only two voxels. One voxel contains a target volume where we wish to apply dose (desired dose of 60 Gy in the target volume) and the other voxel contains an adjacent organ at risk where we want to minimize the dose (desired dose of 0 Gy in the adjacent organ at risk). Further, we have a treatment consisting of only two beamlets, each arising from a different gantry angle.

Note that the space of achievable dose depositions is the region determined by $\mathbf{d} = x_1 \mathbf{b}^{(1)} + x_2 \mathbf{b}^{(2)}$, for $x_1 \geq 0$ and $x_2 \geq 0$. In this case, the desired dose is not exactly achievable with our system, and therefore a tradeoff must be made between reducing dose in the target volume (and potentially not treating the cancer sufficiently) and applying dose in the organ at risk (and potentially damaging this organ). Multiple considerations are generally included in this tradeoff (ie: different options are preferable in different scenarios), which are subsequently encoded in the cost function and constraints of the IMRT optimization problem, as discussed earlier in this lecture.

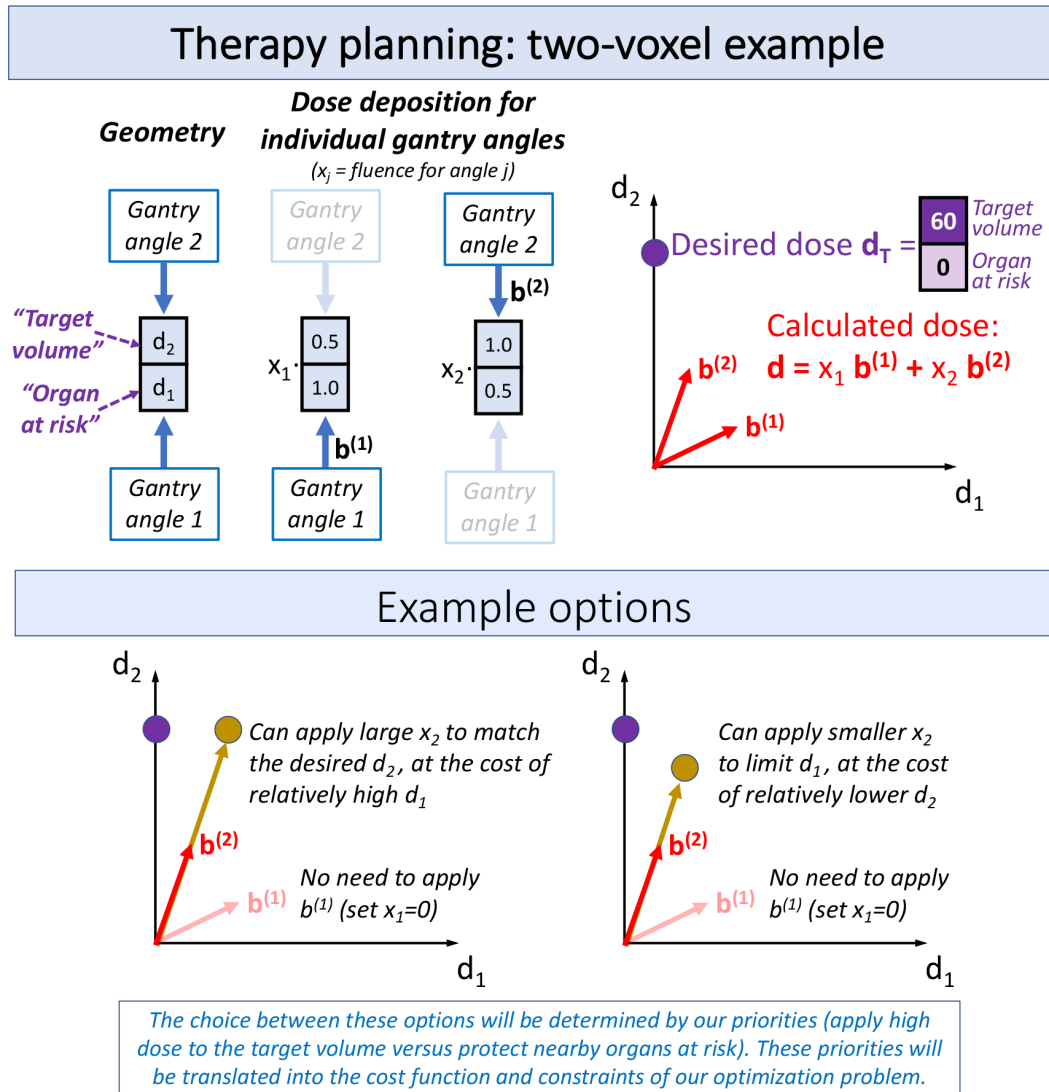


Figure 36.3: Illustrative example representing a two-voxel, two-beamlet IMRT problem. In this toy example, the space of possible solutions arising from our two beamlets is conveniently displayed on a 2D plot. Two potential solutions (choices of x_1 and x_2) are depicted in the bottom row. These two solutions correspond to different priorities in therapy planning, i.e., apply a sufficiently high dose to the target volume versus protect the adjacent organ at risk. In either case, the dose is delivered solely through the second beamlet (gantry angle 2), without using gantry angle 1. Note that fluence needs to be non-negative, which prevents an exact match of the desired dose deposition. The specific choice of fluence x_2 (assuming $x_1 = 0$ for any reasonable solution in this simple example) will be determined by the encoding of our priorities and goals into the cost function and constraints of the corresponding optimization problem.

36.3.5 Beyond fluence: the leaf sequencing problem

In this chapter, we have focused on the problem of determining the fluence map, i.e., the fluence for each beamlet and gantry angle, to obtain an optimal dose distribution. However, in practice the implementation of the desired fluence for each beamlet is achieved using MLCs, which have their own practical and timing limitations. The determination of the optimal sequence of MLC leaf positions is termed “leaf sequencing”. IMRT planning can be viewed as a two-step process where the optimal fluence map is determined first, and then a leaf-sequencing algorithm is applied to closely match the desired fluence map. However, IMRT planning can also be posed as a single-step direct aperture optimization problem, where the leaf sequencing is design directly. This direct approach has the advantage that it determines a feasible treatment plan in a single step (e.g., it may avoid fluence maps that are infeasible or undesirable based on MLC limitations). In addition, by inherently accounting for the leaf sequencing with the corresponding nonlinearities arising from the MLC, direct aperture optimization methods may be more accurate. However, direct aperture optimization generally leads to a non-convex optimization problem because the associated cost function is a non-convex function of the corresponding parameters (the MLC leaf positions).

36.4 Beyond IMRT

In IMRT, we consider the application of radiation from a discrete set of gantry angles, where the gantry remains stationary for each of these angles, while applying radiation, before moving on to the next angle. As an extension of this method, radiation can be applied continuously while the gantry rotates. In Volumetric Modulated Arc Therapy (VMAT), the MLC leaves are adjusted while the gantry rotates and while radiation dose is applied. Compared to IMRT, VMAT has the advantage that radiation is applied continuously from many angles, which can lead to better (more conformal) treatment plans, and may enable shorter overall treatment times. Note that VMAT is related to the Tomotherapy technique, originally proposed by TR Mackie in the early 1990s². In VMAT, the therapy planning (optimization) problem consists of determining the following parameters (each of them as a function of time): i) the MLC leaf trajectories, ii) gantry angle, and iii) dose rate. Importantly, IMRT optimization can be viewed as a simplified version and starting point for VMAT optimization. For more details on the VMAT optimization problem, please see (Unkelbach J, et al, Med Phys, 2015;42:1367-1377) as well as the book “Treatment planning for radiation oncology” mentioned above.

²See the original tomotherapy publication (Mackie TR et al, Med Phys 1993; 20:1709-1719), and a subsequent review paper (Mackie TR, Phys Med Biol, 2006; 51:R427-R453)

