

Lecture 35

Optimization for Radiation Therapy Planning

35.1 Lecture Objectives

- Recognize therapy planning as an optimization problem
- Appreciate the connections between therapy planning and imaging
- Formulate the therapy planning optimization problem in terms of its general components, i.e., the components of its cost function and constraints
- Connect this optimization problem to the concepts covered earlier in this course, including convexity and the properties of various algorithms

35.2 Introduction

Radiation therapy is a form of cancer therapy that uses high-energy radiation to kill cancer cells. This radiation may arise from external beams, or from sources inside the body.

Radiation therapy kills cancer cells by damaging their DNA, such that these cells are not able to replicate and may die. Unfortunately, this same mechanism may also damage healthy tissue. Therefore, therapy planning involves a trade-off between two conflicting goals: treating the cancer while sparing noncancerous tissue. This trade-off leads naturally to an optimization formulation for therapy planning.

The examples described in the following lecture are particularly relevant to external radiation therapy such as intensity modulated radiation therapy (IMRT) or volumetric modulated arc therapy (VMAT). These methods can be viewed as modulating the fluence of a set of beamlets, which in turn result in predictable dose deposition within each of the voxels within the patient. These examples are not meant to provide a comprehensive description of these radiotherapy methods. The overarching goal of this overview, with

corresponding examples and computational problems is to illustrate the direct mapping of the therapy planning problem to an optimization problem. The homework readings will provide an opportunity to learn more about the nuance of practical therapy planning.

35.3 Therapy Planning and Imaging

Modern radiation therapy planning relies heavily on imaging in order to identify the target volume as well as organs at risk ^[1]. Indeed, optimized use of imaging in therapy planning may enable the delivery of increased dose to the target, while minimizing damage to healthy tissues. Further, novel imaging methods have the potential to help identify “radio-resistant” regions of the tumor that require higher dose. Therapy planning typically starts with the acquisition of a CT scan, which provides a 3D view of the tumor as well as additional data needed for dose calculations. Other imaging modalities, particularly MRI and PET, can provide additional data for tumor delineation of characterization, which may be integrated into the therapy planning process. Errors in delineation can lead to insufficient dose delivered to the tumor, or excessive dose delivered to surrounding healthy tissues. For these reasons, medical imaging is critical to existing and emerging therapy planning methods ^[2].

35.4 Components of the Optimization Problem Formulation

To describe therapy planning as an optimization problem, we consider several central components, including the target volume and organs at risk, as well as hardware, timing, or other constraints.

35.4.1 Target Volume

The clinical target volume consists of the delineation of the entire tumor to be treated. As the central therapeutic goal of radiation therapy, an adequate dose needs to be delivered to the clinical target volume (CTV). In practice, the planning target volume (PTV) includes the clinical target volume plus a setup margin. This margin is meant to account for sources of uncertainty such as patient motion. These volumes are depicted in figure ^[35.1]

35.4.2 Organs at Risk

Therapy planning needs to consider other organs that may be damaged by the radiation. These organs are termed organs at risk (ORs). Different types of organs at risk require

¹NG Burnet et al, Defining the tumour and target volumes for radiotherapy, *Cancer Imaging*. 2004; 4(2): 153161.

²Beaton et al, How rapid advances in imaging are defining the future of precision radiation oncology, *British Journal of Cancer* volume 120, pages 779790 (2019).

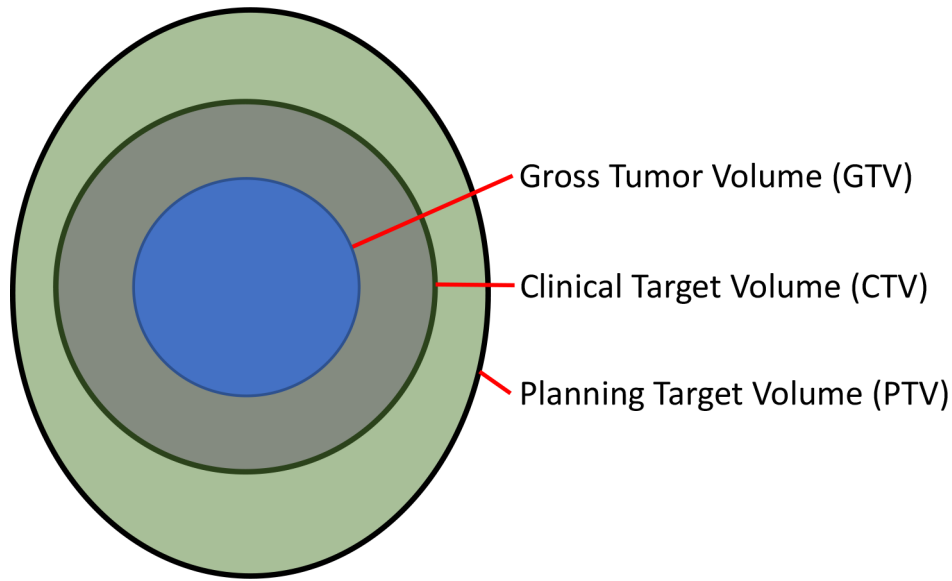


Figure 35.1: Treatment planning volumes: gross tumor volume, clinical target volume, and clinical target volume.

different considerations for therapy planning, as they may have different sensitivity not only to the total dose deposited in the organ, but to the spatial distribution of the dose. For example, organs at risk are often viewed as composed of multiple subunits, and classified as parallel or serial: in parallel organs, many of the subunits need to be disabled in order to cause organ failure; in serial organs, disabling a single subunit may lead to organ failure. The liver is an example of a parallel organ, whereas the spine is a serial organ.

35.4.3 Hardware and Other Constraints

The optimization-based design of a therapy plan is subject to a number of technique-specific constraints. For example, relevant constraints in therapy planning include:

- Limited leaf speed within the multi-leaf collimator (MLC). The MLC is used to modulate the fluence within each beamlet, for each gantry angle. This constraint places a limitation on the fluence distributions that are achievable in close succession.
- Gantry speed constraints.
- Dose rate constraints.
- Sometimes, hard constraints are placed on the dose deposition within organs at risk (see next chapter for details).
- A basic physical constraint is that fluence for each beamlet needs to be non-negative.

35.4.4 Overview of Optimization Formulation

Based on the above described components, the therapy planning problem can generally be viewed in terms of minimization of a cost function subject to a set of constraints:

$$\begin{aligned}
 &\text{minimize} \quad \text{Mismatch vs prescribed dose over PTV} + \\
 &\quad \text{Dose outside PTV} + \\
 &\quad \text{Dose in certain ORs} \\
 &\text{such that} \quad \text{Dose in other ORs is below given threshold} \\
 &\quad \text{Plan satisfies all relevant constraints} \\
 &\quad \text{(physical, hardware, geometric, timing, etc)}
 \end{aligned} \tag{35.1}$$

In other words, our cost function is a combination of multiple conflicting goals, and our constraints include both “hard constraints” on dose deposition to specific organs, as well as experimental constraints on the dose deposition (such as those related to the therapy device, geometry, basic physical constraints, etc). Designing the cost function requires establishing weights for each of its conflicting requirements (e.g., if we wish to very closely match the prescribed dose over the PTV, this may require higher dose outside the PTV). This cost function includes some seeming redundancy: for example, it seeks to minimize the dose outside the PTV, and also to minimize the dose to the ORs: this corresponds to the generic wish to minimize the dose everywhere outside the PTV, and the specific wish to *really* minimize the dose in certain ORs. A result of this approach is that the ORs will generally receive lower dose than other non-PTV tissues, which is the desired effect. Also, note that ORs appear twice in this formulation: once in the cost function and once in the constraints. The reason for this is that we can include our wish to minimize dose in ORs in two different, complementary ways, i.e., either by minimizing the dose to ORs (cost function), and/or by imposing a hard limit on the dose deposited in ORs (constraints). Note that the components of this optimization problem are purposely vague at this point, as there are many cost functions that can penalize mismatch vs prescribed dose over PTV as well as dose outside the PTV or in the ORs. In the following lecture, we will review specific formulations relevant to therapy planning, including specific choices for the components of the cost function and constraints.

35.5 Algorithms

35.5.1 Gradient Descent

The gradient descent algorithm is central to many treatment planning methods. This algorithm has the advantage of being relatively fast and simple to implement. Since gradient descent is a local search method, it will not guarantee convergence to the global optimum in non-convex optimization problems. Instead, gradient descent can generally

be expected to converge to a local optimum, which is adequate in many scenarios. Importantly, many cost functions and constraints used in therapy planning are convex, although this is not always the case.

The basic gradient descent algorithm does not include constraints, and therefore may lead to a solution that violates these constraints. For example, it may lead to a solution that is not realizable in terms of hardware constraints, or a non-physical solution such as one requiring negative fluence. There are a number of methods to deal with these constraints, including approximate handling of constraints via penalty functions, projection approaches (where solutions outside the constraint set get projected back to the constraint set), or variable transformation approaches (where the non-negative fluence may be formulated as x^2 or some other inherently non-negative value).

35.5.2 Second-order Algorithms

Second order algorithms (including second derivatives, as in Newton's method) may help improve the convergence of gradient descent algorithms. As we have discussed in previous lectures, calculation of second derivatives (or Hessian matrices) at each iteration can be computationally expensive, so approximated second order algorithms can be used instead.

35.5.3 Iteratively Refined Dose Deposition Models

Therapy planning methods often rely on linear dose deposition models where the relationship between fluence for each beamlet and dose deposition at each voxel is described by a matrix. This matrix-based linear formulation is highly convenient for simplicity and computational reasons. However, additional (non-linear) considerations are needed in practice to improve the accuracy of the predicted dose deposition, and therefore the quality of the treatment plans. These additional considerations include scatter, as well as inter-leaf leakage in the MLC. These factors result in a dose deposition that is not exactly a linear function of the applied fluence map.

For this reason, practical algorithms need to recompute the dose deposition matrix iteratively within the therapy planning process. Note that this update of the matrix inherently implies that the overall model is non-linear. However, the linear assumption is helpful and locally accurate (i.e., accurate upon small changes in the applied fluence map). Algorithmically, the linear model based on a fixed dose deposition matrix is therefore used as the “inner loop” within a more sophisticated method that accounts for model nonlinearities. For simplicity, the following chapter and corresponding computational exercises will focus on the therapy planning problem under the assumption of a given linear model (i.e., a dose deposition matrix) known a priori.

35.6 Who Plans the Therapy?

Each time radiotherapy is applied to a patient, a treatment plan needs to be designed. In the US, the radiation oncologist, medical physicist, dosimetrist, and radiation therapist

work closely together to design and apply optimal treatment plans. Under the supervision of the oncologist and physicist, the dosimetrist interacts with the computational treatment planning and simulation software, and also helps ensure adequate positioning and immobilization of the patient during treatment. Once the treatment plan is designed, the radiation therapist is in charge of verifying the feasibility of the treatment and applying it to the patient.

Although dosimetrists are able to plan most treatments from start to finish, the physicist provides oversight, and also provides specialized guidance for particularly challenging cases. In addition to overseeing the treatment planning, the medical physicist develops procedures followed by the dosimetrist, and is responsible for calibration and quality assurance, as well as radiation safety of the radiotherapy system.