J. Appl. Numer. Optim. 8 (2026), No. 1, pp. 127-156
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.8.2026.1.08

THE SHAPE OF YOU: INCORPORATING PATIENT GEOMETRY AND CONTROLLING EXTREME DOSAGE IN RADIATION TREATMENT PLANNING

AMANDA MONEGAT[1], ANGELA MORRISON[1], BITA SADEGHI[1], SARAH WEPPLER[1], YURIY ZINCHENKO[1,2,*]

[1]*Department of Mathematics and Statistics, University of Calgary, Alberta, Canada*
[2]*Gurobi Optimization LLC*

Abstract. Cancer treatment has advanced significantly over recent decades, with improvements in technology enabling radiation therapies that more precisely target tumors while sparing healthy tissue. One such approach is Intensity-Modulated Radiation Therapy or Intensity-Modulated Proton Therapy, which commonly relies on inverse planning. In inverse planning, clinicians specify acceptable radiation limits for healthy tissues while ensuring a sufficiently high dose to the tumor, and a computer-generated plan is produced to meet these criteria. In this work, we consider a key inverse-planning metric, the Dose-Volume Histogram (DVH). By interpreting the DVH as a probability distribution, we propose a framework to assess worst-case possible deviations from a reference DVH when generalized equivalent uniform dose constraints are imposed. We present several optimization models for solving the associated bounded moment problem, alongside closed-form bounds derived for four specific sequences of moments. To mitigate overly pessimistic outcomes associated with these formulations, we then introduce a mixed-integer programming approach that incorporates patient-specific anatomical and dosimetric information through a dose-deposition matrix. By embedding geometric considerations directly into the optimization model, the resulting DVH approximations achieve greater clinical relevance. Along with the incorporation of patient geometry, we also investigate strategies for controlling high-dose regions for organs at risk, comparing mean-tail dose and Conditional Value-at-Risk as two complementary approaches. Preliminary numerical experiments demonstrate how our formulations perform and highlight the tradeoffs between tractability and conservatism. The findings contribute to a clearer understanding of the effectiveness of moment-based DVH approximations in the context of radiation therapy planning.
Keywords. Conditional value-at-risk; Linear programming; Mixed-integer programming; Problem of moments; Radiation therapy; Semi-definite programming
2020 Mathematics Subject Classification. 90C90, 90C11, 92C50, 44A60, 90C15.

1. INTRODUCTION

Cancer has become one of the most prevalent diseases and leading causes of mortality worldwide. Among the various treatment methods, radiation therapy has been established as a cornerstone treatment option for a wide range of malignancies [1]. Radiation therapy involves the use of ionizing radiation through high-energy particles or waves, such as X-rays, gamma rays,

*Corresponding author.
E-mail address: yzinchen@ucalgary.ca (Y. Zinchenko).
Received 9 March 2026; Accepted 28 March 2026; Published online 1 April 2026.

electron beams, or protons, to induce damage or destroy tumor cells by creating small breaks in their DNA, thereby aimed at inhibiting their growth and division [2, 3]. At the macro-level, radiation treatment is primarily used to shrink tumors, eliminate residual cancer cells after surgery, manage recurrences, or alleviate symptoms caused by advanced-stage cancer [4].

While other cancer treatments often expose the entire body to therapeutic agents, within the spectrum of radiation-based modalities, external beam radiation therapy involves delivering radiation from a source external to the patient's body [1], focusing precisely on the area requiring intervention [2]. Despite this, adjacent healthy tissues may still be affected by radiation exposure. Consequently, the central objective of radiation treatment planning is to maximize tumor control while minimizing collateral damage to surrounding healthy structures [2]. In this work, we will focus on treatment planning as it relates to Intensity-Modulated Radiation Therapy (IMRT) and Intensity-Modulated Proton Therapy (IMPT). A schematic illustration of the treatment machine is given in Figure 2, where a rotating robotic-controlled radiation gantry revolves around a patient treatment couch to modulate radiation delivery from several angles.

One method in crafting such treatments is through inverse-planning. In conventional "forward" planning, the planner specifies the beam parameters—such as angle, energy, and intensity—and then calculates the resulting dose distribution. Inverse-planning reverses this process: the physician first prescribes the desired dose distribution, including minimum tumor coverage and maximum allowable doses to surrounding organs at risk (OARs), and the optimization algorithm then determines the beamlet intensities that best achieve these clinical goals. Today, inverse-planning with dose–volume objectives and constraints form the foundation of nearly all commercial IMRT planning systems, making it possible to achieve highly conformal treatments that would be impossible to design manually.

Besides clinically prescribed radiation dose targets, another consideration in these plans is the physical realizability of dose distributions. For example, if one is going through IMPT, the beamlets used are very precise and thus will not hit every single voxel in the treatment area. Therefore, there should be a large number of voxels that receive little to no radiation during treatment. These physical limitations are what we aim to include in our formulations along with the information derived from the prescribed dose distribution, referred to as the reference Dose-Volume Histogram (DVH).

Our Contributions. We begin by defining the necessary terminology in Section 2.1. This, in turn, allows us to build the connection between the prescribed dose distribution and the problem of moments. In Section 2.2, we formally define this problem of moments and the bounded moment problem. In particular, we describe the formulations of the bounded moment problem, which form the basis of the work in the later sections.

Next, we start Section 3 with closed form bounds for several specific sequences of moments applied to the bounded moment problem (Section 3.1.1). More generally, we present several alternative reformulations for solving the bounded moment problem. The first of these is the dual formulation, followed by a Semi-Definite Program (Section 3.1.2) and then a Linear Program (Section 3.2).

These methods are limited by the fact that they do not include information pertaining to the physical dose a patient may receive. As such, we show how one can include these physical limitations in the case of a single beam (Section 4.1). While this formulation allows for a direct connection between both the physically realizable dose distribution and the prescribed

dose distribution, we soon see that these connections do not scale well with the number of beamlets. Thus, we present two Mixed-Integer Programs (MIPs) that include information on both distributions of interest (Section 4.2).

In Section 5 we expand on these Mixed-Integer Programming formulations to control the higher-dose region. The high-dose tail of the distributions represents the parts of a patient's organs that are hit with the highest radiation dose. Controlling the tails of the distributions is particularly relevant for reducing damage to healthy organs. This tail control is achieved through the addition of mean-tail and Conditional Value-at-Risk constraints to one of the MIPs from Section 4.2. Finally, we conclude with a summary of each section as well as some directions for future work in the area of inverse-planning for IMRT and IMPT in Section 6.

2. Preliminaries

2.1. Connecting Dose Distributions and Generalized Equivalent Uniform Doses.

In inverse planning, to maintain functionality of the OARs, a physician designs a set of dose distributions which should not be exceeded during treatment. These desired (OAR) dose distributions are what we refer to as a *reference DVH*. Graphically, a DVH is plotted as the dose level, here normalized to $[0,1]$, on the x-axis and the fraction or percentage of the structure's volume receiving at least that dose on the y-axis. A point (τ, v) on the DVH curve means that a fraction v of the structure receives at least dose τ. For example, if at $\tau = 0.7$ the DVH value is 0.1, then 10% of the structure receives at least 70% of the normalized prescription dose; see Figure 1.

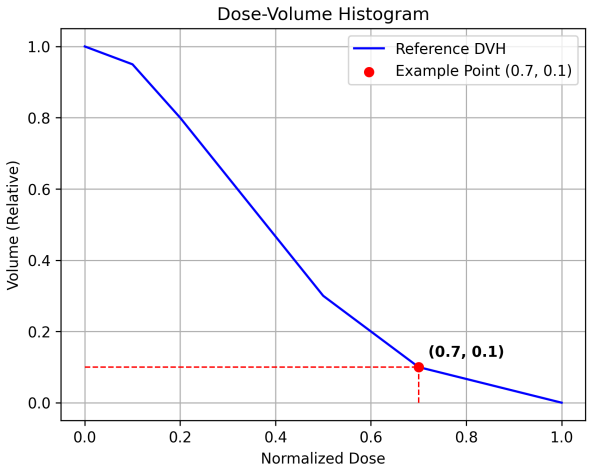


FIGURE 1. Reference DVH curve and example point

The patient treatment volume is discretized into small cubic elements called *voxels*, indexed by $\mathscr{V}$; a typical voxel size is approximately $2 \times 2 \times 2$ mm or similar, with the key idea that all cells within a single voxel receive nearly identical radiation dose. Furthermore, the treatment volume is partitioned into several anatomical structures, typically a target (or several targets) with respective voxels indexed by $\mathscr{T}$ and OARs to be preserved indexed by $\mathscr{H}_1, \mathscr{H}_2, \ldots$, so that $\mathscr{V} = (\bigcup_i \mathscr{H}_i) \cup \mathscr{T}$ and $\mathscr{T} \cap \mathscr{H}_i = \emptyset, \mathscr{H}_i \cup \mathscr{H}_j = \emptyset, \forall i, j$. In turn, during the external beam radiation treatment, the radiation is delivered from an external source to the patient via beams or (narrower, on the order of millimetres) *beamlets* for higher-precision modalities such as IMRT, IMPT, etc. In our context, we use the terms beams and beamlets interchangeably.

A simple inverse-planning problem with a single OAR can be formulated as

$$\begin{aligned} \text{minimize} \quad & \sum_{h\in\mathscr{H}} \mathbf{d}_h \\ \text{subject to:} \quad & L_{\mathscr{T}} \le \mathbf{d}_t \le U_{\mathscr{T}}, \ \ \forall t \in \mathscr{T} \\ & \mathbf{d}_h \le U_{\mathscr{H}}, \ \ \forall h \in \mathscr{H}, \\ & \mathbf{d} = \mathbf{A}\mathbf{w}, \ \ \mathbf{w} \ge \mathbf{0}, \\ & \mathbf{d} \in \mathscr{F}, \end{aligned} \tag{Planning}$$

where $U_{\mathscr{H}}$, $U_{\mathscr{T}}$, and $L_{\mathscr{T}}$ are the upper and lower bounds on the OAR and tumor voxels, respectively, and $\mathscr{F}$ is the set of dose vectors $\mathbf{d} \in \mathbb{R}^{|\mathscr{V}|}$ which satisfy the reference DVH constraints. The latter reference DVH conformity requirement turns out to be quite complex, and an oversimplified model can be posed by ignoring the dose distribution constraints and taking $\mathscr{F} \equiv \mathbb{R}^{|\mathscr{V}|}$. $\mathbf{A} \in \mathbb{R}_+^{|\mathscr{V}|\times b}$ represents the dose deposition matrix whose entries $\mathbf{A}_{ij}$ represent the dose voxel i receives from beamlet j, and $\mathbf{w} \in \mathbb{R}_+^b$ is the beamlet intensity vector. In this equation $\mathbf{A}$ maps the beamlet intensities to a voxel dose vector $\mathbf{d} \in \mathbb{R}_+^{|\mathscr{V}|}$. Figure 2 shows that the green beamlets on the left-side correspond to the entries of $\mathbf{w}$ whereas the area shown in orange corresponds to the voxels hit during treatment meaning $\mathbf{d}$ captures the dose received by each voxel. The machine itself determines what the entries of the matrix $\mathbf{A}$ will be based on the beam angles and position of the patient. Essentially, $\mathbf{d}$ represents a physically realizable dose distribution. Since there always exists a maximal tolerable radiation dose limit, for simplicity we assume that the maximal dose value is normalized to 1.

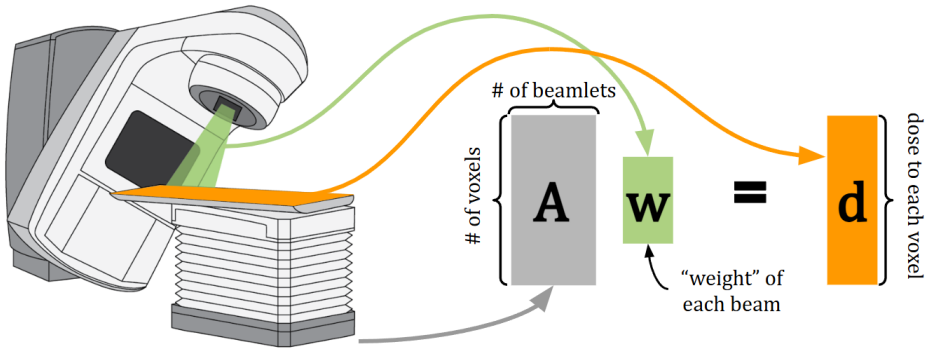


FIGURE 2. Illustration of how an IMRT machine is captured by the relation $\mathbf{A}\mathbf{w} = \mathbf{d}$

To expand on the DVH constraints mentioned previously, we first define the DVH$(\cdot)$ function

$$\mathrm{DVH}_{\mathscr{H}}(\tau) = \frac{1}{|\mathscr{H}|} \sum_{j\in\mathscr{H}} \mathbf{1}_{\{\mathbf{d}_j \ge \tau\}},$$

for a structure of interest indexed by $\mathscr{H} \subseteq \mathscr{V}$ and a dose threshold $\tau \ge 0$. One can interpret $\mathrm{DVH}_{\mathscr{H}}(\tau)$ as the fraction of $\mathscr{H}$ that receive dose at least τ. Thus, a DVH constraint that enforces "for a given structure, no more than ν% of the volume of $\mathscr{H}$ may receive a dose at least τ is written as

$$\mathrm{DVH}_{\mathscr{H}}(\tau) \le \nu.$$

Sparing the functionality of healthy organs that receive collateral radiation damage from treatment can be captured by taking an intersection of critically important prescribed DVH constraints for the OARs. This, combined with a near-uniform high target dose requirement, yields the feasible set $\mathscr{F}$, which is generally non-convex. The latter fact led to research into formulations which allow one to still control the dose distribution yet maintain some form of computational tractability describing $\mathscr{F}$, for instance, by using convex constraints.

Lee, Fox, and Crocker investigated applying integer programming to IMRT treatment planning. They proposed specifying critical DVH points of the form (τ, ν), where τ denotes a threshold dose value, and ν represents the maximum allowable volume fraction of an organ that may receive a dose that exceeds τ [5]. In their formulation, they replace the DVH constraints of (Planning) with,

$$\begin{aligned} &\mathbf{d}_h \leq \tau + z_h(1-\tau), \ \ \forall h \in \mathscr{H}, \\ &\frac{1}{|\mathscr{H}|} \sum_{h \in \mathscr{H}} z_h \leq \nu, \\ &z_h \in \{0,1\} \ \ \forall h \in \mathscr{H}, \end{aligned}$$

where 1 corresponds to the maximum admissible dose across all voxels. For a given OAR, ensuring that the resulting DVH lies below a specified point (τ, ν) requires the introduction of $|\mathscr{H}|$ binary indicator variables z_i, resulting in a large-scale MIP formulation.

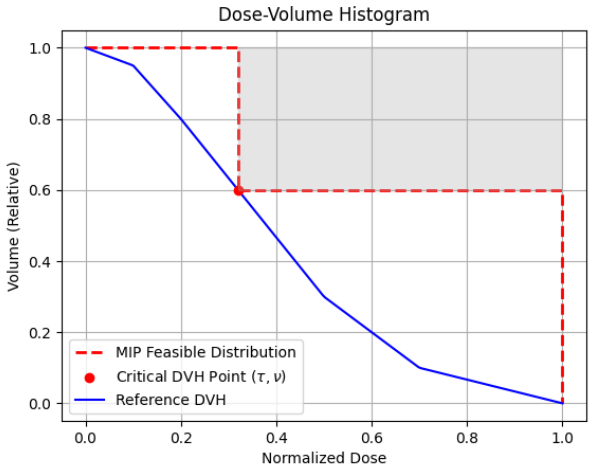


FIGURE 3. Example of a refined feasible region using (τ, ν)-DVH critical point

When solving this MIP for a single OAR with a single constraint at (τ, ν), the feasible dose–volume region is reduced only by the shaded area shown in Figure 3. All other dose distributions remain feasible. For example, the red (dashed) curve in Figure 3 satisfies the MIP constraint at (τ, ν) but clearly violates the intended proximity or non-dominance requirement to the reference DVH shown in blue (solid). To avoid a scenario like this one may need to introduce many critical DVH control points, where each point needs to be addressed individually via its own set of variables and constraints, resulting in an even larger and harder non-convex MIP. Despite substantial advancements in computational power and algorithmic frameworks, this formulation remains prohibitive in time-constrained clinical environments.

As an alternative to the purely DVH-based dose prescription, the use of the so-called generalized equivalent uniform dose (gEUD), introduced by Niemierko [6], was advocated by Quiuwen, Mohan and Niemierko [7] as well as by Beong and Deasy [8]. This concept stems from the assumption that a tumor consists of a large population of clonogenic cells (cells capable of proliferating and forming a multicellular colony derived from a single cell), with cell survival following a Poisson distribution. In this framework, tumor control is determined by the expected number of surviving clonogens after irradiation. Consequently, two dose distributions are considered radio-biologically equivalent if they yield the same expected number of surviving cells [6]. Similar multicellular colony considerations may be applied to OARs to estimate normal tissue complication probabilities.

This led to the notion of an equivalent uniform dose: a hypothetical uniform dose that, when delivered to the target volume over the same fractionation schedule and total treatment time, would produce the same biological effect as the original nonuniform distribution. The main advantage of this formulation lies in its ability to summarize the biological impact of a spatially heterogeneous dose using a single representative value. Despite the potential complexities of a comprehensive probabilistic equivalent uniform dose model, one much simplified surrogate particularly stood out in many empirical studies. Consider a single volume of interest discretized into a finite set of voxels $\mathscr{H}$. Let $\mathbf{d}_i \in [0,1]$ denote the normalized dose delivered to each voxel $i \in \mathscr{H}$, and let $a \in \mathbb{R}$ be a given parameter. The gEUD is defined as

$$\mathrm{gEUD}_a = \frac{1}{|\mathscr{H}|} \sum_{i \in \mathscr{H}} \mathbf{d}_i^a. \tag{2.1}$$

This function is convex in $\mathbf{d}_i$ when $a \in (-\infty, 0) \cup (1, \infty)$, and concave when $a \in [0,1]$. Intuitively, gEUDs can be used to impose restrictions on the resulting DVH, with the parameter a modulating the influence of each constraint on the dose distribution. The values of $a < 0$ emphasize the avoidance of under-dosing by penalizing the presence of many voxels receiving low doses and are suitable for target volumes. For $a > 1$, the focus shifts to penalizing high-dose exposure, making such constraints appropriate for sparing OARs from overdose. The key benefit using multiple gEUD-type constraints is that, by working with suitably-chosen gEUD inequalities, the planning formulation results in a convex optimization problem. This structure is computationally advantageous, enabling the use of efficient algorithms, often with polynomial-time complexity in the problem dimensions.

Zinchenko et al. [9] established a fundamental connection between the gEUD and the dose distribution using probability theory. Their analysis relies on the observation that voxel-wise doses can be interpreted as random variables. Recall that a random variable is a measurable function that maps outcomes from a sample space to the real numbers. A random variable assigns a numerical value to each possible outcome, allowing one to quantify the likelihood of said outcomes. For a continuous random variable X, the distribution is usually described by a probability density function (PDF), f_X. One can also define the probability that X lies in a given interval $[\alpha, \beta]$, i.e., $\Pr(\alpha \leq X \leq \beta)$, and use the cumulative distribution function (CDF), F_X.

Using this interpretation of the voxel-wise dose, one is then able to view the DVH as a probability distribution. In turn, Equation (2.1) corresponds to a probability moment. There is, of course, nothing inherently random about the dose a given voxel receives. However, we adopt this probabilistic perspective on the dose distribution represented by the reference DVH. We treat the set of voxels within a structure of interest, e.g., an OAR, as the sample space,

and consider the random variable D corresponding to selecting one voxel with dose value $\mathbf{d}_i$, uniformly at random. In this way, the distribution of doses across voxels can be viewed as the distribution of a random variable D [9]. Assuming the DVH is continuous, which practically can be thought of as having a large number of voxels, the complement $1-\text{DVH}(\tau)$ corresponds to the CDF of the voxel-dose random variable D: it gives the fraction of voxels receiving dose at most τ, see Figure 4. Consequently, the inverse problem of finding a DVH curve that satisfies a prescribed set of gEUDs becomes mathematically equivalent to the problem of recovering a distribution subject to moment constraints.

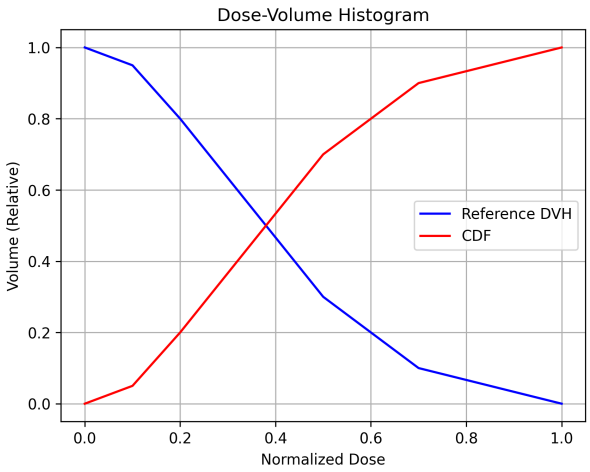


FIGURE 4. Reference DVH compared to CDF: $\text{DVH}(\tau) = 1 - F_D(\tau)$

2.2. **The Problem of Moments.** In 1895, Stieltjes published the first formalization of the problem of moments [10]. In particular, he defines the problems as finding a bounded non-decreasing function $F(x)$ defined over $\mathbb{R}_+$ such that

$$\mu_s = \int_{\mathbb{R}^+} x^s F(dx), \quad s = 0, 1, 2, \ldots$$

for a predetermined sequence of raw moments $\{\mu_0, \mu_1, \mu_2, \ldots, \mu_p\}$, which we will write as $\{\mu_p\}$. He employed the theory of continued fractions to determine the existence and uniqueness of such an $F(x)$. Stieltjes further established necessary and sufficient conditions for a sequence $\{\mu_p\}$ to be a feasible moment sequence, i.e., to correspond to a valid probability distribution. This was later extended to functions defined over $\mathbb{R}$ and compact intervals such as $[0,1]$ by Hamburger and Hausdorff, respectively (see [11, 12, 13, 14] for further details).

When the entire infinite sequence of moments $\{\mu_\infty\}$ is known, and the moment generating function exists, the corresponding probability distribution $F(x)$ is uniquely determined. However, when only a finite number of moments are specified, a set of admissible probability measures may satisfy the moment constraints. Thus, the classical moment problem can be stated as the problem of determining whether a sequence $\{\mu_p\}$ can be realized as the moments of some probability distribution, and if so, whether this distribution is unique.

In our setting, grounded in practical consideration of later solving the corresponding (Planning) model, we work with a finite sequence of moments. This gives rise to the bound problem, which

aims to determine tight lower and upper bounds over the set of all such feasible probability distributions. The bound problem can be seen from an optimization perspective [15]. If we assume F corresponds to an absolutely continuous measure, the Stieltjes measure $F(dx)$ is replaced by $f(x)\,dx$, allowing us to formulate the Hausdorff bound problem as follows.

Definition 2.1. Consider the set of absolutely continuous probability measures on $[0,1]$ with density f. Given two finite index sets $\{1,2,...,n\}$ and $\{1,2,...,m\}$ and measurable functions $g_s : [0,1] \to \mathbb{R}$ and $h_t : [0,1] \to \mathbb{R}$, the **univariate Hausdorff bound problem** [16] is defined as

$$\begin{aligned} &\sup / \inf_{f(x)} \int_0^\tau f(x)\,dx \\ \text{subject to: } &\int_0^1 g_s(x) f(x)\,dx = \delta_s, \quad \forall s \in \{1,2,...,n\} \\ &\int_0^1 h_t(x) f(x)\,dx \le \eta_t, \quad \forall t \in \{1,2,...,m\} \\ &\int_0^1 f(x)\,dx = 1,\ f(x) \ge 0. \end{aligned}$$

The first and second constraints ensure that $f(x)$ adheres to the prescribed sequence of moments $\{\delta_n\}$ and $\{\eta_m\}$ while the last one explicitly states that $f(x)$ is a probability density. Note the objective corresponds to $\Pr(D \le \tau)$ for a random variable $D \sim f$.

We investigate a specific formulation for the Hausdorff bound which utilizes both raw and shifted moments. In particular, we use shifted first moments,

$$\tilde{\mu}_t = \int_0^1 |x - \psi_t| f(x)\,dx \quad \forall t \in \{1,2,...q\},$$

where, without loss of generality, $\psi_1 < \psi_2 < ... < \psi_q$ for all $\psi_t \in [0,1]$, $t \in \{1,2,...,q\}$. Using this, we can rewrite the upper bound formulation from Definition 2.1 as

$$\begin{aligned} \text{val}_\tau := \inf_{f(x)} &\int_0^\tau f(x)\,dx \\ \text{subject to: } &\int_0^1 x^0 f(x)\,dx = 1 \\ &\int_0^1 x^s f(x)\,dx = \mu_s, \quad \forall s \in \{1,2,3,...,p\} \\ &\int_0^1 |x - \psi_t| f(x)\,dx = \tilde{\mu}_t, \quad \forall t \in \{1,2,...,q\} \\ &f(x) \ge 0, \end{aligned} \tag{Bound}$$

where p is the number of raw moments and q is the number of shifted first-order moments.

Unfortunately, except for a handful of cases with very few moments (Bound) has no closed-form solution, and one must rely on numerical methods. One main computational difficulty with (Bound) is that the forced equality of the raw and shifted moments introduces nonconvex constraints. To make our formulation tractable, we transform most of the moment equalities to

inequalities. This yields the following formulation:

$$\begin{aligned}
\rho_\tau := \inf_{f(x)} & \int_0^\tau f(x)\,dx \\
\text{subject to: } & \int_0^1 x^0 f(x)\,dx = 1 \\
& \int_0^1 x f(x)\,dx = \mu_1 \\
& \int_0^1 x^s f(x)\,dx \le \mu_s, \quad \forall s \in \{2,3,...,p\} \qquad \text{(Relaxed Bound)} \\
& \int_0^1 |x-\psi_t| f(x)\,dx \le \tilde{\mu}_t, \quad \forall t \in \{1,2,...,q\} \\
& f(x) \ge 0.
\end{aligned}$$

The objective value ρ_τ corresponds to the tightest possible upper bound on $\Pr(D \le \tau)$ consistent with the given set of raw and shifted moments. An analogous supremum problem yields the corresponding upper bound. We are primarily interested in the minimization problem since the lower bound on CDF $\Pr(D \le \tau)$ corresponds to the upper bound on $\mathrm{DVH}(\tau)$ and therefore is relevant to the OAR sparing.

3. Solving the Bound Problem

In this section, we present methods for finding solutions to (Bound). First, we present closed form bounds for (Bound) given a specific sequence of raw and shifted moments. In particular, we present formulations for $\{\mu_1, \tilde{\mu}_1, \tilde{\mu}_2\}$, $\{\mu_1, \tilde{\mu}_1\}$, $\{\mu_1, \mu_2\}$, and $\{\mu_1\}$. Empirically, we have observed that working with a handful of raw and shifted moments already has a significant effect on controlling the upper bound on DVH, while adding a much larger number of moment constraints exhibits rapidly diminishing returns. As such, we believe that the presented closed-form bounds are of practical interest. Because the closed-form bounds are restricted to such a small sequence of raw and shifted moments, we then present two numerical methods for solving (Bound) given any sequence of raw and shifted moments. The first is a Semi-Definite Program (SDP) and the second is a Linear Program (LP).

3.1. Duality and Semi-Definite Programming Bounds.

The results presented in this subsection are derived from the dual formulation of (Bound). As such, we provide said dual formulation:

$$\begin{aligned}
\mathrm{val}^*_\tau := \max\ & \mathbf{y}_0 + \sum_{s=1}^{p} \mu_s \mathbf{y}_s + \sum_{t=1}^{q} \tilde{\mu}_t \mathbf{y}_{s+t} \\
\text{subject to: } & \sum_{s=0}^{p} x^s \mathbf{y}_s + \sum_{t=1}^{q} |x-\psi_t| \mathbf{y}_{p+t} \le \mathbf{1}_{[0,\tau]}, \qquad \text{(Bound Dual)}
\end{aligned}$$

where $\mathbf{1}_{[0,\tau]}$ is the indicator function which is 1 when $x \in [0,\tau]$ and 0 otherwise.

We note that the dual and subsequent SDP and LP formulations can be extended to $g_s(x), h_t(x)$ in Definition 2.1 as piece-wise rational functions. However, the corresponding notation becomes convoluted and we find that working with the above simpler equality-based dual suffices for our narrative.

3.1.1. *Closed-Form Bounds.* It may be noted that the left-hand side of the constraint in (Bound Dual) is the expression for a polynomial curve. For several special cases, which are also of practical interest, we can derive analytic Hausdorff probability bounds by positioning the curve below the indicator step function, as per the constraints in the dual formulation. Namely, we derive analytic probability bounds for the following selections of shifted and raw moments,

(i) $\mu_1, \tilde{\mu}_1, \tilde{\mu}_2$,
(ii) $\mu_1, \tilde{\mu}_1$,
(iii) μ_1, μ_2,
(iv) μ_1,

with the proof sketch given in the Appendix A.

(i) raw and shifted moments $\mu_1, \tilde{\mu}_1, \tilde{\mu}_2$: assume that the first order raw moment μ_1 and two shifted moments $\tilde{\mu}_1$, $\tilde{\mu}_2$, $\psi_1 < \psi_2$ of a Hausdorff probability distribution are specified. We consider the equality-constrained problem (Bound).

$$\begin{aligned} &\inf_{f(x)} \int_0^\tau f(x)\,dx \\ \text{subject to: } &\int_0^1 x f(x)\,dx = \mu_1 \\ &\int_0^1 |x-\psi_1| f(x)\,dx = \tilde{\mu}_1 \\ &\int_0^1 |x-\psi_2| f(x)\,dx = \tilde{\mu}_2 \\ &\int_0^1 f(x)\,dx = 1,\ f(x) \geq 0. \end{aligned}$$

For completeness, we include the expressions for both upper and lower bounds. With

$$\theta_1 = \frac{(\psi_2-\psi_1)\mu_1 - \psi_2\tilde{\mu}_1 + \psi_1\tilde{\mu}_2}{\psi_2-\psi_1+\tilde{\mu}_2-\tilde{\mu}_1}$$

and

$$\theta_2 = \frac{\psi_1\psi_2 + \psi_1(1-2\psi_2)\mu_1 - \psi_1\tilde{\mu}_2}{\psi_1 + (1-\psi_2-\psi_1)\mu_1 + (\psi_2-1)\tilde{\mu}_1 - \psi_1\tilde{\mu}_2},$$

$$\theta_3 = \frac{(\psi_2-\psi_1)\mu_1 + \psi_2\tilde{\mu}_1 - \psi_1\tilde{\mu}_2}{\psi_2-\psi_1+\tilde{\mu}_1-\tilde{\mu}_2},$$

we have

$$\Pr(D \leq \tau) \geq \begin{cases} 0, & 0 \leq \tau < \theta_1 \\ \frac{\tau(\psi_1-\psi_2)+(\psi_2-\psi_1)\mu_1+(\tau-\psi_2)\tilde{\mu}_1+(\psi_1-\tau)\tilde{\mu}_2}{2\tau(\psi_1-\psi_2)}, & \theta_1 \leq \tau < \psi_1 \\ \frac{\psi_1-\mu_1+\tilde{\mu}_1}{2\psi_1}, & \psi_1 \leq \tau < \theta_2 \\ \frac{\psi_2(1+\tau-\psi_1)+\psi_1-2\tau+(\tau-\psi_2)\mu_1+(1-\psi_2)\tilde{\mu}_1+(\tau-1)\tilde{\mu}_2}{2(\tau-\psi_1)(\psi_2-1)}, & \theta_2 \leq \tau < \psi_2 \\ \frac{\psi_2-\psi_1-\tilde{\mu}_1+\tilde{\mu}_2}{2(\psi_2-\psi_1)}, & \psi_2 \leq \tau < \theta_3 \\ \frac{2\tau-\psi_2-\mu_1-\tilde{\mu}_2}{2(\tau-\psi_2)}, & \theta_3 \leq \tau \leq 1 \end{cases} \tag{3.1}$$

and

$$\Pr(D \le \tau) \le \begin{cases} \frac{-\psi_1+\mu_1-\tilde{\mu}_1}{2(\tau-\psi_1)}, & 0 \le \tau < \theta_1 \\ \frac{\psi_2-\psi_1-\tilde{\mu}_1+\tilde{\mu}_2}{2(\psi_2-\psi_1)}, & \theta_1 \le \tau < \psi_1 \\ \frac{\psi_1(\tau-\psi_2)+(\psi_1-\tau)\mu_1+\tau\tilde{\mu}_1-\psi_1\tilde{\mu}_2}{2\psi_1(\tau-\psi_2)}, & \psi_1 \le \tau < \theta_2 \\ \frac{2-\psi_2-\mu_1-\tilde{\mu}_2}{2(1-\psi_2)}, & \theta_2 \le \tau < \psi_2 \\ \frac{(2-\tau)(\psi_2-\psi_1)+(\psi_1-\psi_2)\mu_1+(\tau-\psi_2)\tilde{\mu}_1+(\psi_1-\tau)\tilde{\mu}_2}{2(\psi_2-\psi_1)(1-\tau)}, & \psi_2 \le \tau < \theta_3 \\ 1, & \theta_3 \le \tau \le 1 \end{cases} \tag{3.2}$$

The validity of the newly established bounds has been confirmed with subsequent numerical approaches based on SDP and LP; for example, it can be verified that the SDP and LP bound in Figure 7a matches our closed-form solution curve with $\mu_1 = 0.405$, $\psi_1 = 0.35, \tilde{\mu}_1 = 0.1825$, $\psi_2 = 0.5, \tilde{\mu}_2 = 0.2049$.

(ii) raw and shifted moments $\mu_1, \tilde{\mu}_1$: given μ_1 and one shifted moment $\tilde{\mu}_1$, similarly we can compute a closed form solution to (Bound). Namely, with

$$\phi_1 = \frac{\psi_1 + (1-2\psi_1)\mu_1 - \tilde{\mu}_1}{2-\psi_1-\mu_1-\tilde{\mu}_1}$$

and

$$\phi_2 = \frac{2\psi_1\mu_1}{\psi_1+\mu_1-\tilde{\mu}_1},$$

we have

$$\Pr(D \le \tau) \ge \begin{cases} 0, & 0 \le \tau < \phi_1 \\ \frac{2\tau-\psi_1(1+\tau)+(-1+2\psi_1-\tau)\mu_1+(1-\tau)\tilde{\mu}_1}{2\tau(1-\psi_1)}, & \phi_1 \le \tau < \psi_1 \\ \frac{\psi_1-\mu_1+\tilde{\mu}_1}{2\psi_1}, & \psi_1 \le \tau < \phi_2 \\ \frac{2\tau-\psi_1-\mu_1-\tilde{\mu}_1}{2(\tau-\psi_1)}, & \phi_2 \le \tau \le 1 \end{cases} \tag{3.3}$$

and

$$\Pr(D \le \tau) \le \begin{cases} \frac{\psi_1-\mu_1+\tilde{\mu}_1}{2(\psi_1-\tau)}, & 0 \le \tau < \phi_1 \\ \frac{2-\psi_1-\mu_1-\tilde{\mu}_1}{2(1-\psi_1)}, & \phi_1 \le \tau < \psi_1 \\ \frac{\psi_1(2-\tau)+(\tau-2\psi_1)\mu_1-\tau\tilde{\mu}_1}{2\psi_1(1-\tau)}, & \psi_1 \le \tau < \phi_2 \\ 1, & \phi_2 \le \tau \le 1 \end{cases} \tag{3.4}$$

(iii) raw moments μ_1, μ_2: similarly to the celebrated Chebyshevs' one-sided inequality,

$$\Pr(X \le \mu_1 - \chi) \le \frac{\mu_2 - \mu_1^2}{\chi^2 + \mu_2 - \mu_1^2}, \chi > 0,$$

we have

$$\Pr(D \le \tau) \ge \begin{cases} 0, & \tau \le \frac{\mu_1-\mu_2}{1-\mu_1} \\ \frac{\tau-(1+\tau)\mu_1+\mu_2}{\tau}, & \frac{\mu_1-\mu_2}{1-\mu_1} < \tau \le \frac{\mu_2}{\mu_1} \\ \frac{(\tau-\mu_1)^2}{\mu_2-2\mu_1\tau+\tau^2}, & \frac{\mu_2}{\mu_1} < \tau \le 1 \end{cases}$$

and

$$\Pr(D \le \tau) \le \begin{cases} \frac{\mu_2-\mu_1^2}{\mu_2-2\mu_1\tau+\tau^2}, & \tau \le \frac{\mu_1-\mu_2}{1-\mu_1} \\ \frac{1-\tau+\tau\mu_1-\mu_2}{1-\tau}, & \frac{\mu_1-\mu_2}{1-\mu_1} < \tau \le \frac{\mu_2}{\mu_1} \\ 1, & \frac{\mu_2}{\mu_1} < \tau \le 1 \end{cases}$$

The validity of the newly established bounds has been confirmed with subsequent numerical approaches based on SDP and LP; for example, it can be verified that the SDP bound in Figure 5 matches our closed-form solution curve with $\mu_1 = 0.405$, $\mu_2 = 0.2143$.

We observe that these bounds do not coincide with Chebyshev's inequality. To examine the role played by the support of a probability distribution, we proceed to derive the bounds for the Stieltjes moment problem, that is, for a non-negative random variable, and the Hamburger moment problem with a random variable supported on the whole of $\mathbb{R}$. Our results are summarized in Table 1. For the Hamburger moment problem with two raw moments, indeed the bounds

TABLE 1. Tight analytic raw moment inequalities given μ_1, μ_2

Stieltjes, supported on $\mathbb{R}_+$	**Hamburger, supported on $\mathbb{R}$**
$\Pr(D \le \tau) \ge \begin{cases} 0, & \tau \le \mu_1 \\ 1-\frac{\mu_1}{\tau}, & \mu_1 < \tau \le \frac{\mu_2}{\mu_1} \\ \frac{(\tau-\mu_1)^2}{\mu_2-2\mu_1\tau+\tau^2}, & \frac{\mu_2}{\mu_1} < \tau \le 1 \end{cases}$	$\Pr(D \le \tau) \ge \begin{cases} 0, & \tau \le \mu_1 \\ \frac{(\tau-\mu_1)^2}{\mu_2-2\mu_1\tau+\tau^2}, & \tau > \mu_1 \end{cases}$
$\Pr(D \le \tau) \le \begin{cases} \frac{\mu_2-\mu_1^2}{\mu_2-2\mu_1\tau+\tau^2}, & \tau \le \mu_1 \\ 1, & \tau > \mu_1 \end{cases}$	$\Pr(D \le \tau) \le \begin{cases} \frac{\mu_2-\mu_1^2}{\mu_2-2\mu_1\tau+\tau^2}, & \tau \le \mu_1 \\ 1, & \tau > \mu_1 \end{cases}$

coincide with Chebyshev's one-sided inequality. However, note that in our bounds are tighter with specified support.

(iv) raw moment μ_1: the derived bounds for Hausdorff, Hamburger and Stieltjes random variables are summarized in the table below. At the same time, note that from Markov's inequality $\Pr(X \ge \chi) \le \frac{\mu_1}{\chi}$, for a random variable supported on $[0,1]$ we can also infer that

$$\Pr(D < \tau) = 1 - \Pr(D \ge \tau) \ge 1 - \frac{\mu_1}{\tau} \tag{3.5}$$

and

$$\Pr(D < \tau) = \Pr(1 - D \ge 1 - \tau) \le \frac{1-\mu_1}{1-\tau}. \tag{3.6}$$

Thus, in the case of a single first moment, refining the support essentially does not improve the probability bounds over a classical Markov's inequality.

TABLE 2. Tight analytic raw moment inequalities given μ_1

Hausdorff $[0,1]$	**Stieltjes** $\mathbb{R}_+$	**Hamburger** $\mathbb{R}$
$\Pr(D \le \tau) \ge \begin{cases} 0, & \tau \le \mu_1 \\ 1 - \frac{\mu_1}{\tau}, & \tau > \mu_1 \end{cases}$	$\Pr(D \le \tau) \ge \begin{cases} 0, & \tau \le \mu_1 \\ 1 - \frac{\mu_1}{\tau}, & \tau > \mu_1 \end{cases}$	$\Pr(D \le \tau) \ge 0$
$\Pr(D \le \tau) \le \begin{cases} \frac{1-\mu_1}{1-\tau}, & \tau \le \mu_1 \\ 1, & \tau > \mu_1 \end{cases}$	$\Pr(D \le \tau) \le 1$	$\Pr(D \le \tau) \le 1$

3.1.2. *SDP Bounds.* As mentioned earlier, in general, there are no closed-form bounds for any given sequence of raw and shifted moments. Therefore, we present numerical procedures for computing the solution to (Bound). Similar to Bertsimas and Popescu [15], we aim to express the problem of determining probability bounds by an SDP.

With auxiliary $\psi_0 = 0, \psi_{q+1} = 1$, expanding the absolute value notation in (Bound Dual) yields

$$\text{val}^*_\tau = \max \mathbf{y}_0 + \sum_{s=1}^{p} \mu_s \mathbf{y}_s + \sum_{t=1}^{q} \tilde{\mu}_t \mathbf{y}_{s+t}$$

subject to:

$$\left(\mathbf{y}_0 - \sum_{t=1}^{\ell} \psi_t \mathbf{y}_{p+t} + \sum_{t=\ell}^{q} \psi_t \mathbf{y}_{p+t}\right) + x\left(\mathbf{y}_1 - \sum_{t=1}^{\ell} \mathbf{y}_{p+t} - \sum_{t=\ell}^{q} \mathbf{y}_{p+t}\right) \quad \text{(Expanded DB)}$$
$$+ \sum_{s=2}^{p} x^s \mathbf{y}_s \le \mathbf{1}_{[0,\tau]}, \qquad \forall x \in [\psi_\ell, \psi_{\ell+1}],\ \ell \in \{0,2,...,q\}$$

The constraints in (Expanded DB) require that a polynomial curve, defined by each left-hand side expression, be contained below the step function $\mathbf{1}_{[0,\tau]}$ over the respective support. Next, we apply the SDP characterizations for polynomial non-negativity due to Nesterov [17].

For the subsequent SDP formulation, for a given τ we pick k such that

$$\psi_k < \tau \le \psi_{k+1}.$$

With this, we have

$$
\begin{aligned}
\mathrm{val}_\tau^* = \max\ & \mathbf{y}_0 + \sum_{s=1}^{p} \mathbf{y}_s \mu_s + \sum_{t=1}^{q} \mathbf{y}_{p+t} \tilde{\mu}_t \\
\text{subject to:}\ & \sum_{a,b:a+b=2r-1} \mathbf{X}_{ab} = 0, \quad \forall r \in \{1,\dots,p\} \\
& 1 - \mathbf{y}_0 + \sum_{j=1}^{\ell} \psi_j \mathbf{y}_{p+j} - \sum_{j=\ell+1}^{q} \psi_j \mathbf{y}_{p+j} = \mathbf{X}_{00} \\
& \forall \ell \in \{0,\dots,k\},\ x \in [0,\tau] \begin{cases} \binom{p}{r} \left(1 + \sum_{j=1}^{\ell} \psi_j \mathbf{y}_{p+j} - \sum_{j=\ell+1}^{q} \psi_j \mathbf{y}_{p+j} \right) \\ + \binom{p-1}{r-1} \tau \left(\sum_{j=\ell+1}^{q} \mathbf{y}_{p+j} - \sum_{j=1}^{\ell} \mathbf{y}_{p+j} \right) \\ - \sum_{j=0}^{r} \mathbf{y}_j \tau^j \binom{p-j}{r-j} = \sum_{a,b:a+b=2r} \mathbf{X}_{ab},\ \forall r \in \{1,\dots,p\} \end{cases} \\
& \sum_{a,b:a+b=2r-1} \mathbf{Z}_{ab} = 0, \quad \forall r \in \{1,\dots,p\} \qquad \text{(SDP Bound)} \\
& \sum_{j=0}^{p} \mathbf{y}_j \tau^j + \sum_{j=1}^{\ell} \psi_j \mathbf{y}_{p+j} - \sum_{j=\ell+1}^{q} \psi_j \mathbf{y}_{p+j} \\
& + \tau \left(\sum_{j=\ell+1}^{q} \mathbf{y}_{p+j} - \sum_{j=1}^{\ell} \mathbf{y}_{p+j} \right) = \mathbf{Z}_{00} \\
& \forall \ell \in \{k,\dots,q\},\ x \in [\tau,1] \begin{cases} \binom{p}{r} \left(\sum_{j=1}^{\ell} \psi_j \mathbf{y}_{p+j} - \sum_{j=\ell+1}^{q} \psi_j \mathbf{y}_{p+j} \right) \\ + \left[\binom{p-1}{r} \tau + \binom{p-1}{r-1} \right] \left(\sum_{j=\ell+1}^{q} \mathbf{y}_{p+j} - \sum_{j=1}^{\ell} \mathbf{y}_{p+j} \right) \\ - \sum_{m=0}^{r} \sum_{j=m}^{p+m-r} \binom{j}{m} \binom{p-j}{r-m} \mathbf{y}_j \tau^{j-m} \\ \qquad = \sum_{a+b=2r} \mathbf{Z}_{ab},\ \forall r \in \{1,\dots,p-1\} \end{cases} \\
& \sum_{m=0}^{p} \sum_{j=m}^{m} \binom{j}{m} \binom{p-j}{p-m} \mathbf{y}_j \tau^{j-m} + \cdots = \sum_{a,b:a+b=2p} \mathbf{Z}_{ab} \\
& \mathbf{X}, \mathbf{Z} \succeq 0
\end{aligned}
$$

An analogous formulation for the upper probability bound on $\Pr(D \leq \tau)$ can be obtained by minimizing rather than maximizing the objective function, or considering a complementing event $\Pr(D \geq \tau)$ with the corresponding $\mathbf{1}_{[\tau,1]}$ in the derivation. The SDP formulation can be extended to accommodate piece-wise rational moment functions in the constraints, as previously mentioned.

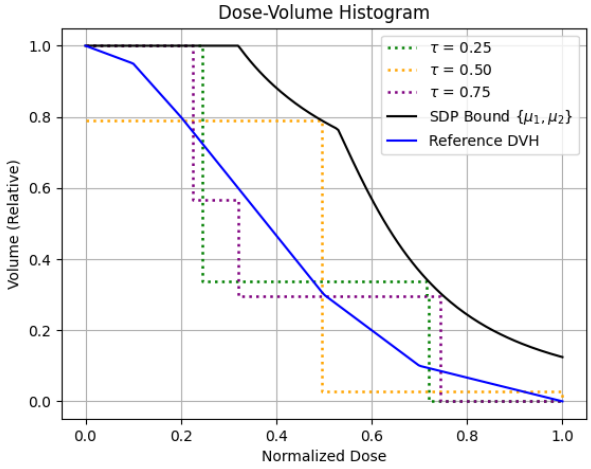


FIGURE 5. Upper SDP Bound with markers for $\tau = 0.25, 0.5, 0.75$

The bounds for the reference DVH are obtained by taking the union of val_τ^* for each $\tau \in [0,1]$. Figure 5 shows several extreme distributions –solutions to the primal (Bound), extracted through complementarity– that achieve these objective values for a select few τ.

Strictly speaking, we should address the potential duality gap issue. In short, this is not problematic. From the practical –OAR sparing– point of view, we are primarily interested in establishing a valid upper bound on $\text{DVH}(\tau)$. Recalling the DVH and CDF complementarity relationship, we note that $\text{DVH}(\tau) = 1 - \Pr(D \leq \tau) = 1 - \text{val}_\tau \leq 1 - \text{val}_\tau^*$ since universally by weak duality we have $\text{val}_\tau \geq \text{val}_\tau^*$. So, indeed, (SDP Bound) provides a valid upper bound on $\text{DVH}(\tau)$ regardless of duality gap. Moreover, we expect strong duality to hold and thus the SDP-based bound to be tight since strict feasibility of the moment problem suffices. In turn, strict feasibility, that is, the feasibility of (Bound) under infinitesimal perturbations to the reference moment values $\mu_s, \tilde{\mu}_t$ is expected to hold since the reference DVH_{ref} that gives rise to these moments is generally not "extreme" in our setting, as explained at the end of Section 3, whereas extreme DVHs correspond to atomic distributions and are step-function like.

A shortcoming of the SDP approach is the potential numerical instability, primarily due to the binomial coefficients appearing in the formulation when higher-order moments are utilized. While one could remedy this by changing the polynomial basis, it is not clear how to incorporate patient geometric information into the SDP dual-based formulation.

3.2. **LP Bounds.** We now present a more numerically stable LP formulation to approximate the exact probability bounds on $\Pr(D \leq \tau)$. Unlike the SDP approach, the LP works directly with the primal (Relaxed Bound). Therefore, solving the LP may readily be preferred for determining tight upper bounds on $\text{DVH}(\tau)$.

Let $0 = x_0 < x_1 < x_2 < \ldots < x_N = 1$, be a partition of $[0,1]$ into N equal intervals, $x_j = j/N$. Using this partition, any integral over $[0,1]$ can be expressed as

$$\int_0^1 g(x)f(x)\,dx = \sum_{j=1}^{N} \int_{x_{j-1}}^{x_j} g(x)f(x)\,dx.$$

For each sub-interval $[x_{j-1}, x_j]$, the corresponding probability mass is

$$\mathbf{f}_j = \int_{x_{j-1}}^{x_j} f(x)\, dx.$$

Each integral over $[x_{j-1}, x_j]$ can be interpreted as the probability mass assigned to a single representative point, e.g., x_{j-1} or x_j, effectively replacing a continuous distribution with an atomic measure.

Now, consider the moment integrals in (Bound). By exploiting the monotonicity of x^s and the fact that $\mathbf{f}_j$ captures the total probability mass on $[x_{j-1}, x_j]$, we have

$$\left(\frac{j-1}{N}\right)^s \mathbf{f}_j \leq \int_{\frac{j-1}{N}}^{\frac{j}{N}} x^s f(x)\, dx \leq \left(\frac{j}{N}\right)^s \mathbf{f}_j, \quad \forall s \in \{1, 2, ..., p\}. \tag{3.7}$$

These inequalities express the fact that the contribution of the integral over each interval lies between the two extreme cases: all probability mass concentrated at x_{j-1} or at x_j. In particular, for $s = 1$ the right-hand side yields

$$\int_0^1 x f(x)\, dx \leq \sum_{j=1}^{N} \frac{j-1}{N} \mathbf{f}_j + \frac{1}{N}.$$

With these in mind, we formulate two LP models that approximate (Relaxed Bound). For simplicity, we describe the model with raw moments only; the extension to shifted moments is straightforward. We start with *(discrete) inner approximation* LP.

For some fixed

$$\tau = \frac{\tau_N}{N}, \quad \tau_N \in \{0, 1, \ldots, N\},$$

consider

$$\begin{aligned} \overline{\rho_\tau} := \min & \sum_{j=1}^{\tau_N} \mathbf{f}_j \\ \text{subject to: } & \sum_{j=1}^{N} \mathbf{f}_j = 1, \\ & \sum_{j=1}^{N} \frac{x_{j-1} + x_j}{2} \mathbf{f}_j = \mu_1, \\ & \sum_{j=1}^{N} x_j^s \mathbf{f}_j \leq \mu_s, \quad \forall s \in \{2, 3, ..., p\}, \\ & \mathbf{f} \geq \mathbf{0} \end{aligned} \tag{LP Inner}$$

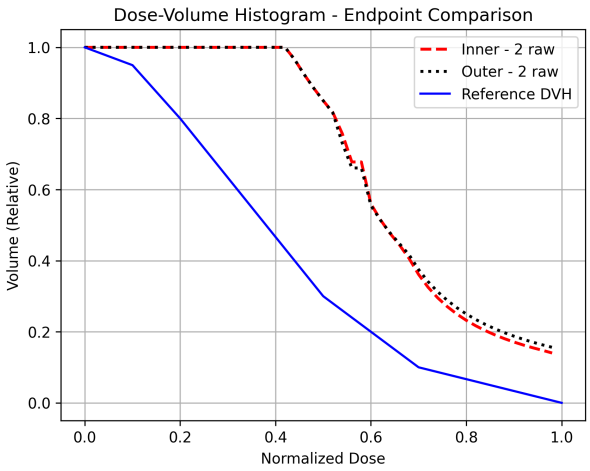


FIGURE 6. Comparing inner and outer LP approximation bounds with $N = 20$

Next, accounting for the first moment equality constraint as a two-sided inequality, we formulate *(discrete) outer approximation* LP,

$$\begin{aligned} \underline{\rho_\tau} := \min & \sum_{j=1}^{\tau_N} \mathbf{f}_j \\ \text{subject to: } & \sum_{j=1}^{N} \mathbf{f}_j = 1, \qquad \text{(LP Outer)} \\ & \mu_1 - \frac{1}{N} \le \sum_{j=1}^{N} x_{j-1}\mathbf{f}_j \le \mu_1, \\ & \sum_{j=1}^{N} x_{j-1}^s \mathbf{f}_j \le \mu_s, \quad \forall s \in \{2,3,...,p\}, \\ & \mathbf{f} \ge \mathbf{0} \end{aligned}$$

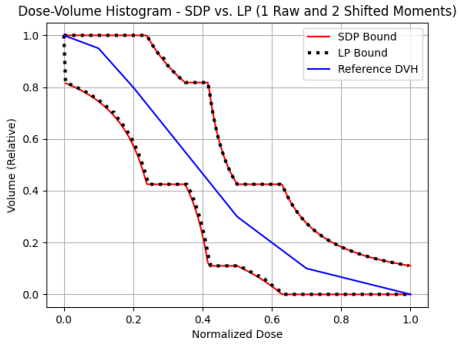
By construction, the feasible region of (Relaxed Bound) is "sandwiched" between the (discretized) feasible regions of (LP Inner) and (LP Outer). Note that if $\mathbf{f}$ is a solution for (LP Inner), we can take the corresponding $f(x)$ to be a piece-wise step function with $f(0) = 0$ and $f(x) = N\mathbf{f}_j$, $x \in (x_{j-1}, x_j]$, feasible with respect to (Relaxed Bound). Thus, we have the following

Theorem 3.1.

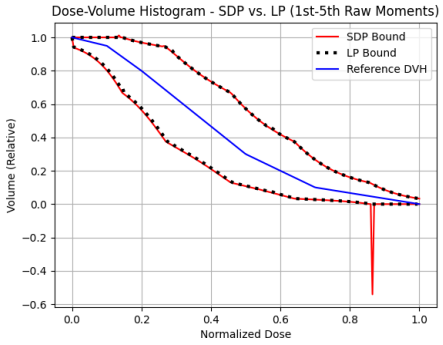
(1) If (LP Inner) *is feasible, then* (Relaxed Bound)*) is also feasible.*
(2) If (Relaxed Bound) *is feasible, then* (LP Outer) *is also feasible.*

Consequently, we have $\underline{\rho}_\tau \le \rho_\tau \le \overline{\rho}_\tau$.

Intuitively, as N grows larger, our approximations improve in quality and get closer to the original continuous model. Indeed, even with a modest N it is hard to distinguish the two LP bounds visually, see Figure 6. Figure 7a shows how these LP bounds compare to those computed by the SDP and Figure 7b highlights the potential numerical instability of the SDP. (LP Inner) and (LP Outer) formulations can be easily extended to accommodate shifted, piece-



(A) DVH bounds from the SDP and LP based on the sequence $\{\mu_1, \tilde{\mu}_1, \tilde{\mu}_2\}$



(B) DVH bounds from the SDP and LP based on the sequence $\{\mu_1, \mu_2, \mu_3, \mu_4, \mu_5\}$

FIGURE 7. SDP vs LP formulation bounds

wise rational moments, etc. Note that we can also refine the partitioning for a specific dose subinterval, departing from the uniform choice of x_j, to better accommodate the LP approximation towards some desired moment functions. For instance, if we are primarily interested in the higher-order moments, we can pick non-uniform x_j with higher concentration around 1.

Note on the reference DVH and moments: recall that $\mathbf{d}$ can be associated to a random variable D so that (discrete-valued, nonnegative) D takes values $\mathbf{d}_i$ with equal probability $1/|\mathscr{H}|$. With this probabilistic interpretation of the dose vector, we have $\mathrm{DVH}(\tau) = \Pr(D \geq \tau)$. As the number of voxels $\mathscr{H}$ increases, the DVH approaches the complement of the standard CDF function $\mathrm{DVH}(\tau) \approx \Pr(D > \tau) = 1 - F_D(\tau)$. In practicality, we work with a sufficiently large number of voxels to consider the DVH to be the exact complement to the CDF.

The latter limiting case provides a mechanism to compute the moments from the reference DVH using integration by parts. For instance, for the raw moments we have

$$\mu_s = \int_0^1 x^s f_{\mathrm{ref}}(x)dx = s\int_0^1 x^{s-1}(1 - F_{\mathrm{ref}}(x))dx = s\int_0^1 x^{s-1}\,\mathrm{DVH}_{\mathrm{ref}}(x)dx. \tag{3.8}$$

Shifted moments can be computed similarly.

We also note that in our setup, we construct the reference $\mathrm{DVH}_{\mathrm{ref}}(x)$ as a piece-wise linear function, connecting the available clinically relevant DVH points (τ_i, ν_i). For instance, in Figure 7a one can see that the reference DVH consists of four linear segments. In turn, this can be further used to simplify the calculation of the reference moments by integration.

4. INCLUDING PATIENT GEOMETRY

When considering (Relaxed Bound), we only take into account the reference DVH. In reality, there are physical limitations to the dose that a voxel may receive during radiation treatment, captured via $\mathbf{Aw} = \mathbf{d}$ and here referred to as the patient geometry. In what follows, we refer to the above physically-realizable dose as the *empirical dose*. For notational simplicity, we focus on a single OAR with voxels indices in $\mathscr{H}$, and thus have $\mathbf{d} \in \mathbb{R}^{|\mathscr{H}|}$, $\mathbf{A} \in \mathbb{R}^{\mathscr{H} \times b}$ with b representing the number of beamlets.

The methods presented in Section 3 do not take the potentially significant patient geometry into account. Here, we propose two MIP formulations to address this shortcoming. As we will see, in large part, these formulations are motivated by the limits of the more elementary approaches developed for the 1-beam and 2-beam cases.

4.1. **The 1-Beam and 2-Beam Cases.** We begin by discussing the single beam case, i.e., when the treatment plan utilizes only one beam $\mathbf{w} = [\mathbf{w}_1]$ to deliver radiation to the OAR. Here, it is more convenient to rely on a purely probabilistic interpretation of the dose. Let A be the random variable corresponding to the empirical dose distribution with $\mathbf{w}_1$ at unit intensity. For simplicity, we assume A to be continuously distributed and consider only standard (non-shifted) moments. For $\mathbf{w}_1 > 0$, we define a scaled random variable $D = \mathbf{w}_1 A$, which represents the dose delivered by the beamlet with intensity $\mathbf{w}_1$. This directly corresponds to the earlier discussed affine relation $\mathbf{d} = \mathbf{A}\mathbf{w}$.

For $\tau > 0$ the cumulative distribution function F_D satisfies the standard scaling identity

$$F_D(\tau) = \Pr(D \leq \tau) = \Pr(\mathbf{w}_1 A \leq \tau) = \Pr\left(A \leq \frac{\tau}{\mathbf{w}_1}\right) = F_A\left(\frac{\tau}{\mathbf{w}_1}\right). \tag{4.1}$$

In other words, the CDF of D is obtained by a horizontal rescaling of the baseline empirical dose CDF (F_A). This transformation incorporates the patient geometry through the factor $\mathbf{w}_1$ and preserves the underlying distributional structure, while shifting the evaluation of the distribution along the dose axis. Differentiating Equation 4.1 gives us D-distribution density

$$f_D(\tau) = \frac{1}{\mathbf{w}_1} f_A\left(\frac{\tau}{\mathbf{w}_1}\right).$$

Building on this relationship, and recalling that the DVH is defined as the complementary CDF of the dose distribution, we note that any moment condition enforced on D translates into an equivalent requirement on the original dose distribution through this scaling relationship. Namely, the s^{th} raw moment of the scaled dose variable D can be expressed as

$$\mu_s^D = \int_0^\infty x^s f_D(x)\, dx = \int_0^\infty x^s \frac{1}{\mathbf{w}_1} f_A\left(\frac{x}{\mathbf{w}_1}\right) dx.$$

Applying the change of variables $x = \mathbf{w}_1 y$ and $dx = \mathbf{w}_1\, dy$, this then becomes

$$\mu_s^D = \int_0^\infty (\mathbf{w}_1 y)^s \frac{1}{\mathbf{w}_1} f_A(y) \mathbf{w}_1\, dy = \mathbf{w}_1^s \int_0^\infty y^s f_A(y)\, dy = \mathbf{w}_1^s \mu_s^A,$$

where μ_s^A denotes the s^{th} raw moment of the empirical dose distribution at unit intensity. Thus, any moment constraint imposed on D can be equivalently translated into a constraint on the empirical unit dose distribution A.

Therefore, the single-beamlet continuous bound problem (with raw moments) can be formulated as

$$\begin{aligned} &\inf F_A\left(\frac{\tau}{\mathbf{w}_1}\right) \\ &\text{subject to: } \mathbf{w}_1 \mu_1^A = \mu_1, \\ &\qquad \mathbf{w}_1^s \mu_s^A \leq \mu_s,\ \forall s \in \{2, \ldots, p\}, \end{aligned} \tag{Single Beam}$$

which can be easily solved by picking a suitable $\mathbf{w}_1 > 0$ that maximizes the objective function.

The situation with the moment constraints with 2 beams is much less favorable but is still straightforward. Unlike the 1-beam case, it is more convenient to switch over to the discrete voxel-based setup. For simplicity, let us consider how one can impose the second-order raw moment inequality constraint.

Let $\mathbf{A}_{i,1}$ denote unit-intensity dose deposition coefficients for beam-1, and $\mathbf{A}_{i,2}$ — for beam-2, and $\mathbf{w}_1, \mathbf{w}_2$ – the intensities of beams 1 and 2, respectively. Assuming there are a total of $|\mathscr{H}|$ voxels, we can write the respective second moment as

$$\begin{aligned} f(\mathbf{w}_1,\mathbf{w}_2) &= 1/|\mathscr{H}| \sum_i (\mathbf{A}_{i,1}\mathbf{w}_1 + \mathbf{A}_{i,2}\mathbf{w}_2)^2 \\ = 1/|\mathscr{H}| \sum_i \mathbf{A}_{i,1}^2\mathbf{w}_1^2 + 1/|\mathscr{H}| \sum_i \mathbf{A}_{i,2}^2\mathbf{w}_2^2 + 2/|\mathscr{H}| \sum_i \mathbf{A}_{i,1}\mathbf{A}_{i,2}\mathbf{w}_1\mathbf{w}_2 \\ &= a_1\mathbf{w}_1^2 + a_2\mathbf{w}_2^2 + 2a_{12}\mathbf{w}_1\mathbf{w}_2 \end{aligned}$$

with a_1, a_2, a_{12} denoting the respective sums of dose deposition coefficients. In turn, the desired second raw moment inequality constraint can be written as

$$\alpha\mathbf{w}_1^2 + \beta\mathbf{w}_2^2 + 2a_{12}\mathbf{w}_1\mathbf{w}_2 \leq \mu_2.$$

Despite being elementary, the above approach highlights at least one potential issue already: with a larger number of beamlets –practically in the order of thousands– we have to consider a potentially large, with regards to the number of non-zeros, quadratic form to represent a relatively simple second-order moment constraint. The number of non-zero cross-terms would increase as we have more beamlets interacting with one another, which in turn is not uncommon. Furthermore, constructing higher-order raw moments would necessitate collecting even more cross-terms, and quickly becomes impractical.

Finally, with the above approach, the key issue is pinning down the objective – counting the number of voxels with the dose in excess of some fixed value τ. Presently, we do not see an alternative to using some form of discrete variables. Also, it is not clear at all if encompassing the patient geometry is possible within the dual or SDP models. These considerations motivate our switch to the MIP-based formulation next.

4.2. **Geometry Inclusive MIP Formulations.** While the single beam formulation shows that there are direct ways to connect the distribution of the values of $\mathbf{d}$ to the moments of the reference DVH, the two presented approaches above do not scale well with the number of beams. Thus, we look at other formulations to allow incorporating the geometry with multiple beamlets.

Let the radiation field be discretized into b beamlets ($b > 1$), and the OAR represented by a finite set of $|\mathscr{H}|$ voxels. Similar to the discretization in Section 3.2, let the domain of the DVH be partitioned into N equally spaced intervals, where each interval $[x_{j-1}, x_j]$ for $j = 1, 2, \ldots, N$, corresponds to a discrete DVH bin. With the dose-deposition matrix $\mathbf{A} \in \mathbb{R}^{|\mathscr{H}| \times b}$, the linear relation $\mathbf{d} = \mathbf{A}\mathbf{w}$ provides the mechanism that directly links the beamlet intensity vector $\mathbf{w} \in \mathbb{R}^b$, to the resulting voxel-wise dose vector $\mathbf{d} \in \mathbb{R}^{|\mathscr{H}|}$. The experiments in this chapter are conducted using a toy matrix $\mathbf{A}$. Figure 8 shows the voxels from a synthetic 2D OAR and the 5 beams (arrows) used during a hypothetical treatment; the grid corresponds to the voxels of the OAR while the red (dashed) lines represent the beams.

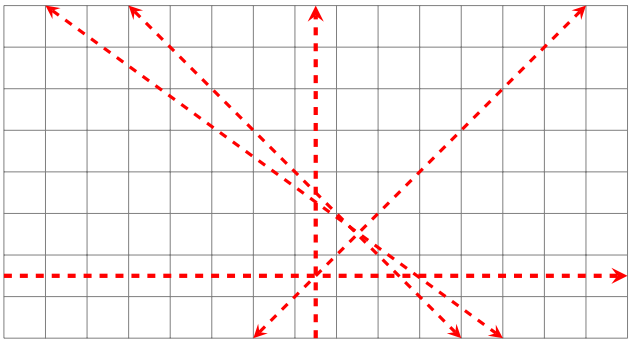


FIGURE 8. Synthetic 2D OAR used in experiments

For simplicity we assume that whenever a unit-intensity beam intersects a voxel, its dose contribution to that voxel is 1. $\mathbf{A}$ subsequently used for illustration purposes is the matrix representation of our toy OAR example.

This already leads to an important practical observation that, based on this illustration, specifically the total fraction of voxels affected by the beamlets clearly not exceeding 50%, we see that the pure moment-based bounds without the geometry taken into account can be quite pessimistic. For instance, with this geometry, the bound in excess of 80% of the organ receiving the dose of .4 or greater, as per Figure 7a, is clearly a gross overstatement.

Unlike the single-beam case, the multi-beam case requires that the dose distribution be explicitly coupled to the moment-based formulation. To relate voxel doses to a discretized DVH, each $\mathbf{d}_i$ must be assigned to the respective bin over the DVH domain. For this purpose, the binary decision variables z_{ij} are introduced. The bin association is enforced through the constraint $x_{j-1}z_{ij} \leq \mathbf{d}_i \leq x_j z_{ij} + (1 - z_{ij})$, which guarantees that the voxel dose $\mathbf{d}_i$ lies within the interval $[x_{j-1}, x_j]$ if and only if $z_{ij} = 1$. In particular, when $z_{ij} = 1$, the inequalities reduce to $x_{j-1} \leq \mathbf{d}_i \leq x_j$, forcing $\mathbf{d}_i$ to lie within the specified interval. Otherwise, $0 \leq \mathbf{d}_i \leq 1$ which imposes no effective restriction on $\mathbf{d}_i$ since the dose is normalized to $[0,1]$. To ensure that each voxel dose is assigned to exactly one bin, the additional constraint $\sum_{j=1}^{N} z_{ij} = 1$ for all $i \in \{1,2,...,|\mathscr{H}|\}$ is imposed.

To establish the connection between the dose bin assignments at the voxel-level and the probability histogram, the following constraint is introduced: $\sum_{i=1}^{|\mathscr{H}|} z_{ij} = |\mathscr{H}|\mathbf{f}_j, \forall j$. This constraint aggregates the binary assignment over all voxels, where the summation of z_{ij} represents the number of voxels whose doses fall within that interval. Dividing by the total number of voxels, the variable $\mathbf{f}_j$ is interpreted as the empirical frequency associated with the j^{th} DVH bin. Equivalently, $\mathbf{f} \in \mathbb{R}^N$ defines a discrete probability mass function induced by the voxel-wise dose distribution, where $\mathbf{f}_j$ represents the fraction of the organ volume receiving a dose in $[x_{j-1}, x_j]$.

In summary, the spatial dose information is mapped to DVH intervals via the binary variables z_{ij}, and subsequently aggregated into the frequencies $\mathbf{f}_j$ that appear explicitly in the moment constraints. The moment, inequalities are enforced on a dose-volume distribution that is not abstractly prescribed, but instead induced by a physically realizable dose field consistent with the underlying anatomy and beam configuration. We can express the bound problem, which incorporates this geometry information, as follows

$$
\begin{aligned}
\min \quad & \sum_{j=1}^{\tau_N} \mathbf{f}_j \\
\text{subject to:} \quad & \sum_{j=1}^{N} \mathbf{f}_j = 1, \qquad \text{(MIP Geometry)} \\
& \text{Moment Constraints} \begin{cases}
\sum_{j=1}^{N} x_{j-1}\mathbf{f}_j \le \mu_1, \quad \sum_{j=1}^{N} x_{j-1}\mathbf{f}_j \ge \mu_1 - \frac{1}{N} & \\
\sum_{j=1}^{N} x_{j-1}^s \mathbf{f}_j \le \mu_s, & \forall s \in \{2,\dots,p\} \\
\sum_{j:x_j<\psi_t} |\psi_t - x_j|\mathbf{f}_j + \sum_{j:x_{j-1}>\psi_t} |x_{j-1}-\psi_t|\mathbf{f}_j & \\
\quad + \sum_{j:\psi_t \in [x_{j-1},x_j]} \min(x_j-\psi_t, \psi_t - x_{j-1})\mathbf{f}_j \le \tilde{\mu}_t, & \forall t \in \{1,\dots,q\}
\end{cases} \\
& \text{Empirical Bin Assignment} \begin{cases}
\mathbf{A}\mathbf{w} = \mathbf{d}, & \\
x_{j-1}z_{ij} \le \mathbf{d}_i \le x_j z_{ij} + (1-z_{ij}), & \forall i \in \{1,\dots,|\mathscr{H}|\}, \forall j \in \{1,\dots,N\} \\
\sum_{j=1}^{N} z_{ij} = 1, & \forall i \in \{1,\dots,|\mathscr{H}|\} \\
\sum_{i=1}^{|\mathscr{H}|} z_{ij} = |\mathscr{H}|\mathbf{f}_j, & \forall j \in \{1,\dots,N\}
\end{cases} \\
& \mathbf{f} \ge \mathbf{0},\ \mathbf{w} \ge \mathbf{0},\ z_{ij} \in \{0,1\},
\end{aligned}
$$

where the last notationally cumbersome moment constraint is simply a relaxed version of the shifted moment requirement. Since we are primarily interested in recovering an upper bound on DVH(τ), this model is formulated on the basis of (LP Outer) (to provide a valid lower bound on $\Pr(D \le \tau)$).

The formulation in (MIP Geometry) forces both the lower and the upper moment-induced bounds to be satisfied, and thus it is possible for there to exist a dose distribution $\mathbf{d}$ which is not feasible because of the moment constraints, but is feasible in terms of the clinical preference. In particular, OAR sparing constraints specify upper limits on dose-volume while delivering a much lower dose is even more favorable. For example, if one executed a perfect treatment in which the OAR is never hit with any radiation, this would be preferred over a plan in which the OAR receives any amount of radiation; see Figure 9. However, the DVH where the OAR never receives any radiation does not fall between the lower and upper moment-induced bounds on the DVH. To address this, we propose two alternative formulations.

Since we are concerned with preserving only the upper bound on DVH(τ), the idea is to introduce a spurious dose distribution $\tilde{\mathbf{f}}$, governed by the moment constraints only, and use it as a dominating histogram for the physically realizable dose $\mathbf{d}$.

The first reformulation achieves this through a voxel-based approach. Let $\tilde{\mathbf{f}}$ denote the auxiliary DVH distribution designed to satisfy the prescribed raw and shifted moment constraints, and let $\tilde{\mathbf{d}}$ denote the dose values associated with the moment-consistent representation. We also introduce an additional set of binary variables $\tilde{z}_{ij}$ to associate each dose $\tilde{\mathbf{d}}_i$ to a unique DVH bin. The resulting reformulation is

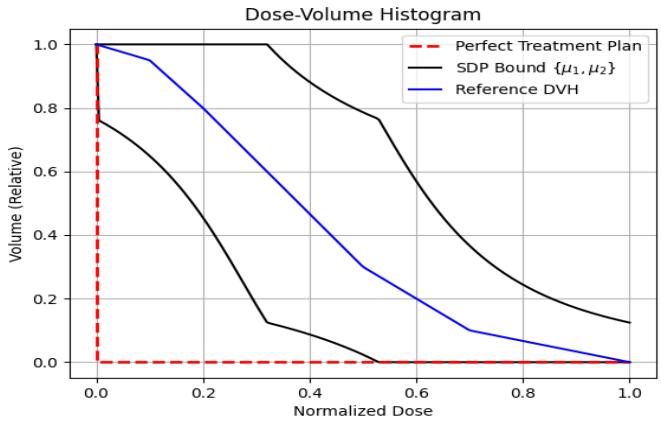


FIGURE 9. Moment-infeasible but otherwise clinically preferred DVH

$$
\begin{aligned}
\min \quad & \sum_{j=1}^{\tau_N} \mathbf{f}_j \\
\text{subject to:} \quad & \sum_{j=1}^{N} \mathbf{f}_j = 1, \quad \sum_{j=1}^{N} \tilde{\mathbf{f}}_j = 1, \qquad \text{(MIP Voxel-Geometry)} \\
& \text{Moment Constraints} \begin{cases}
\sum_{j=1}^{N} x_{j-1}\tilde{\mathbf{f}}_j \le \mu_1, \quad \sum_{j=1}^{N} x_{j-1}\tilde{\mathbf{f}}_j \ge \mu_1 - \frac{1}{N} & \\
\sum_{j=1}^{N} x_{j-1}^{s}\tilde{\mathbf{f}}_j \le \mu_s, & \forall s \in \{2,\dots,p\} \\
\sum_{j:x_j<\psi_t} |\psi_t - x_j|\tilde{\mathbf{f}}_j + \sum_{j:x_{j-1}>\psi_t} |x_{j-1}-\psi_t|\tilde{\mathbf{f}}_j & \\
\quad + \sum_{j:\psi_t\in[x_{j-1},x_j]} \min(x_j-\psi_t, \psi_t - x_{j-1})\tilde{\mathbf{f}}_j \le \tilde{\mu}_t, & \forall t \in \{1,\dots,q\}
\end{cases} \\
& \text{Spurious Bin Assignment} \begin{cases}
x_{j-1}\tilde{z}_{ij} \le \tilde{\mathbf{d}}_i \le x_j\tilde{z}_{ij} + (1-\tilde{z}_{ij}), & \forall i \in \{1,\dots,|\mathscr{H}|\}, \forall j \in \{1,\dots,N\} \\
\sum_{j=1}^{N} \tilde{z}_{ij} = 1, & \forall i \in \{1,\dots,|\mathscr{H}|\} \\
\sum_{i=1}^{|\mathscr{H}|} \tilde{z}_{ij} = |\mathscr{H}|\tilde{\mathbf{f}}_j, & \forall j \in \{1,\dots,N\}
\end{cases} \\
& \text{Empirical Bin Assignment} \begin{cases}
\mathbf{A}\mathbf{w} = \mathbf{d}, & \\
x_{j-1}z_{ij} \le \mathbf{d}_i \le x_j z_{ij} + (1-z_{ij}), & \forall i \in \{1,\dots,|\mathscr{H}|\}, \forall j \in \{1,\dots,N\} \\
\sum_{j=1}^{N} z_{ij} = 1, & \forall i \in \{1,\dots,|\mathscr{H}|\} \\
\sum_{i=1}^{|\mathscr{H}|} z_{ij} = |\mathscr{H}|\mathbf{f}_j, & \forall j \in \{1,\dots,N\}
\end{cases} \\
& \text{Coupling} \quad \{ \ \tilde{\mathbf{d}} \ge \mathbf{d} \\
& \mathbf{f} \ge \mathbf{0},\ \tilde{\mathbf{f}} \ge \mathbf{0},\ \mathbf{w} \ge \mathbf{0},\ z_{ij} \in \{0,1\},\ \tilde{z}_{ij} \in \{0,1\},
\end{aligned}
$$

where some of the constraints may be deduced as redundant, such as $\sum_{j=1}^{N} \mathbf{f}_j = 1$ implied by the empirical bin assignment, but are kept in the formulation for ease of interpretation.

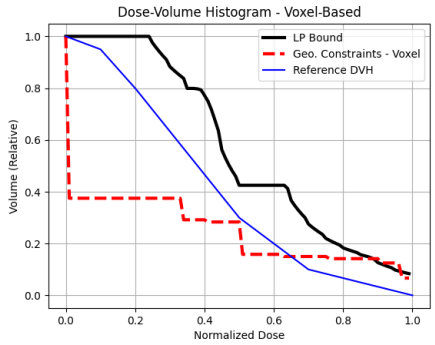
As we see in Figure 10a, the inclusion of the patient geometry can significantly reduce the area of the DVH curve which lies above the reference DVH. The physical limitations can greatly restrict the space of realizable DVH distributions that satisfy a set of moment constraints.

The second reformulation is frequency-based. This model re-establishes feasibility with respect to the lower moment-based DVH bound by directly linking the geometry-induced dose distribution to the moment-consistent frequency representation. To ensure that, in DVH terms, the empirical dose distribution remains bounded above by the envelope prescribed through the moment constraints, we add the inequalities of the form $\sum_{j=1}^{\ell} \tilde{\mathbf{f}}_j \le \sum_{j=1}^{\ell} \mathbf{f}_j$. This frequency-based model is formulated as

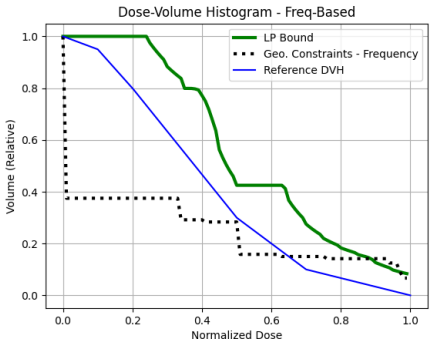
$$
\begin{aligned}
\min \quad & \sum_{j=1}^{\tau_N} \mathbf{f}_j \\
\text{subject to:} \quad & \sum_{j=1}^{N} \mathbf{f}_j = 1, \quad \sum_{j=1}^{N} \tilde{\mathbf{f}}_j = 1, \qquad \text{(MIP Frequency-Geometry)} \\
& \text{Moment Constraints} \begin{cases}
\sum_{j=1}^{N} x_{j-1}\tilde{\mathbf{f}}_j \le \mu_1, \quad \sum_{j=1}^{N} x_{j-1}\tilde{\mathbf{f}}_j \ge \mu_1 - \frac{1}{N} & \\
\sum_{j=1}^{N} x_{j-1}^{s}\tilde{\mathbf{f}}_j \le \mu_s, & \forall s \in \{2,\dots,p\} \\
\sum_{j:x_j<\psi_t} |\psi_t - x_j|\tilde{\mathbf{f}}_j + \sum_{j:x_{j-1}>\psi_t} |x_{j-1}-\psi_t|\tilde{\mathbf{f}}_j & \\
\quad + \sum_{j:\psi_t\in[x_{j-1},x_j]} \min(x_j-\psi_t, \psi_t - x_{j-1})\tilde{\mathbf{f}}_j \le \tilde{\mu}_t, & \forall t \in \{1,\dots,q\}
\end{cases} \\
& \text{Empirical Bin Assignment} \begin{cases}
\mathbf{A}\mathbf{w} = \mathbf{d}, & \\
x_{j-1}z_{ij} \le \mathbf{d}_i \le x_j z_{ij} + (1-z_{ij}), & \forall i \in \{1,\dots,|\mathscr{H}|\}, \forall j \in \{1,\dots,N\} \\
\sum_{j=1}^{N} z_{ij} = 1, & \forall i \in \{1,\dots,|\mathscr{H}|\} \\
\sum_{i=1}^{|\mathscr{H}|} z_{ij} = |\mathscr{H}|\mathbf{f}_j, & \forall j \in \{1,\dots,N\}
\end{cases} \\
& \text{Coupling} \begin{cases}
\sum_{j=1}^{\ell} \tilde{\mathbf{f}}_j \le \sum_{j=1}^{\ell} \mathbf{f}_j, & \forall j \in \{1,\dots,N\},
\end{cases} \\
& \mathbf{f} \ge \mathbf{0},\ \tilde{\mathbf{f}} \ge \mathbf{0},\ \mathbf{w} \ge \mathbf{0},\ z_{ij} \in \{0,1\},
\end{aligned}
$$

with, again, some redundant constraints kept for better interpretability of the model. The effects of the geometric information inclusion via the frequency domain are illustrated in Figure 10b. Comparing Figures 10a and 10b, we see that both formulations, as expected, produce near identical results. The frequency-based approach took roughly 650 seconds to find a solution, while the voxel-based approach took 270 seconds. However, we hypothesize that this runtime trend will get reversed as the size of $\mathbf{A}$ increases, due to the voxel-based model relying on both

$\mathbf{d}$ and $\tilde{\mathbf{d}}$ and consequently becoming much larger. Thus, we elect to use the frequency-based approach when including patient geometry in the experiments performed in Section 5.



(A) Upper DVH bound based on (MIP Voxel-Geometry) for $\{\mu_1, \mu_2, \tilde{\mu}_1, \tilde{\mu}_2\}$



(B) Upper DVH bound based on (MIP Frequency-Geometry) for $\{\mu_1, \mu_2, \tilde{\mu}_1, \tilde{\mu}_2\}$

FIGURE 10. Voxel-based vs frequency-based MIP formulations

5. CONTROLLING EXTREME RADIATION DOSAGE THROUGH TAIL CONSTRAINTS

In this section, we concern ourselves with extreme dosage control. Recall that the tail of the DVH curve corresponds to the highest dosage that an OAR may receive during treatment. Thus, a high tail indicates that more voxels in the OAR are being hit with a high dose of radiation. In this section, we look at two methods for controlling the tail of the DVH: mean-tail and Conditional Value-at-Risk (CVaR).

5.1. Mean-Tail Constraints.

The mean-tail dose constraint limits the average dose delivered to voxels receiving doses above a specified threshold γ. Assume a random variable D is supported on $[0,1]$ with density f_D and $\gamma \in [0,1]$ is fixed. We denote $(D-\gamma)^+ = \max\{0, D-\gamma\}$. Then one can write the mean-tail dose $E[(D-\gamma)^+]$ as

$$\begin{aligned} E[(D-\gamma)^+] &= \int_0^1 \max\{0, x-\gamma\} f_D(x)\, dx \\ &= \int_\gamma^1 (x-\gamma) f_D(x)\, dx \\ &= (x-\gamma)F_D(x)|_\gamma^1 - \int_\gamma^1 F_D(x)\, dx \\ &= (1-\gamma) - \int_\gamma^1 F_D(x)\, dx \\ &= \int_\gamma^1 (1-F_D(x))\, dx \\ &= \int_\gamma^1 \mathrm{DVH}(x)\, dx, \end{aligned}$$

where F_D is the CDF of the dose across voxels, and $\mathrm{DVH}(x) = 1 - F_D(x)$. Thus, the mean-tail dose is simply the area under the DVH curve tail beyond the threshold γ. For the reference DVH, we denote this tail quantity as

$$\alpha_\gamma = \int_\gamma^1 \mathrm{DVH}_{\mathrm{ref}}(x)dx.$$

There are two ways to incorporate these tail constraints into (MIP Frequency-Geometry): using auxiliary dose variables or probability mass terms. For the first approach, the mean-tail condition appears through the auxiliary variables $\tilde{\mathbf{d}}_i$, which measure dose in excess of γ. We can write our mean-tail constraints as

$$\begin{aligned} &\tilde{\mathbf{d}}_i \geq 0, \quad \forall i \in \{1,2,...,|\mathscr{H}|\}, \\ &\tilde{\mathbf{d}}_i \geq \mathbf{d}_i - \gamma, \quad \forall i \in \{1,2,...,|\mathscr{H}|\}, \qquad \text{(Mean-tail Voxel)} \\ &\frac{1}{|\mathscr{H}|}\sum_{i=1}^{|\mathscr{H}|} \tilde{\mathbf{d}}_i \leq \alpha_\gamma. \end{aligned}$$

Alternatively, we can include the mean-tail dose constraints through the use of frequency variables $\mathbf{f}_j$. For simplicity, assume γ is such that $\gamma = \frac{\gamma_N}{N}$ for some integer $\gamma_N \in \{0,1,...,N-1\}$. In this case,

$$\sum_{j=\gamma_N+1}^{N-1} (x_j - \gamma)\,\mathbf{f}_{j+1} \leq \alpha_\gamma, \qquad \text{(Mean-tail Frequency)}$$

provides a (outer discrete approximation to the) constraint guaranteeing that the weighted contribution of the bins above γ does not exceed the tail area under the reference DVH.

In our numerical experiments, we investigate whether—and to what extent—these mean-tail constraints improve upon the DVH bounds from (MIP Frequency-Geometry). We specifically focus on improvements in the tail of the distribution. Figure 11 shows that the tail of the DVH curve produced from combining (MIP Frequency-Geometry) and (Mean-tail Frequency) is lower than that of the curve produced in Figure 10b, illustrated with $\gamma = 0.7$, i.e., we are looking at the last 30% of the dose distribution.

5.2. **CVaR Constraints.** The Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR) are mainly used in portfolio optimization and actuarial science for measuring quantities associated with the tail of a loss distribution. In the optimization context, unlike generally non-convex VaR, CVaR is convex, and admits a well-known LP characterization [18].

Simply put, in our context VaR at some fixed level $(1-\beta)$ corresponds to $\gamma(\beta)$ so that

$$\mathrm{DVH}_{\mathrm{ref}}(\gamma(\beta)) = 1 - \beta,$$

that is, $\gamma(\beta)$ is the smallest of the (highest or) worst-$(1-\beta)$% dose values; for instance, the reference DVH in Figure 10b tells us that the worst-10% dose values start at $\gamma(.9) = .7$. With this, the corresponding conditional worst-$(1-\beta)$% (tail) average dose can be computed as

$$\gamma(\beta) + \frac{\alpha(\beta)}{1-\beta},$$

where

$$\alpha(\beta) = \int_{\gamma(\beta)}^1 \mathrm{DVH}_{\mathrm{ref}}(x)dx.$$

To go back to our illustration with Figure 10b, the average dose above $\gamma(.9) = .7$ is .85 and matches $.7 + \frac{.1\times.3}{2\times.1}$.

It is worth noting that the mean-tail dose and CVaR serve a very similar purpose –controlling the high-end tail of the distribution– and are related constructs. In fact, the LP representation of the CVaR constraint is similar to (Mean-tail Voxel) with the exception that now we have to treat γ as a variable:

$$\begin{aligned} \tilde{\mathbf{d}}_i &\geq 0, \quad \forall i \in \{1,2,...,|\mathscr{H}|\}, \\ \tilde{\mathbf{d}}_i &\geq \mathbf{d}_i - \gamma, \quad \forall i \in \{1,2,...,|\mathscr{H}|\} \qquad \text{(CVaR)} \\ \gamma + \frac{1}{(1-\beta)|\mathscr{H}|} \sum_{i=1}^{|\mathscr{H}|} \tilde{\mathbf{d}}_i &\leq \gamma(\beta) + \frac{\alpha(\beta)}{1-\beta}, \\ \gamma &\geq 0. \end{aligned}$$

As we saw with the mean-tail constraints, the CVaR constraints provide similar control of the extreme dose, see Figure 10b for $\beta = .9$. Mean-tail constraints establish the tail-end control by picking a threshold over the dose values, while CVaR is driven by the tail-end volume percentage. If a higher degree of tail distribution control is desired, one can consider using several

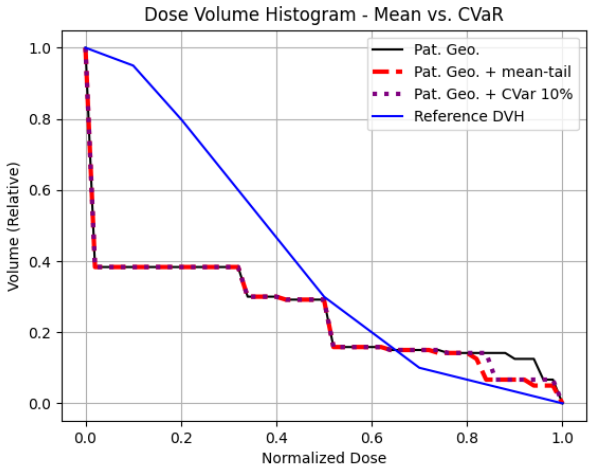


FIGURE 11. Upper DVH bounds based on (MIP Frequency-Geometry) and both tail controlling formulations for the sequence $\{\mu_1, \mu_2, \tilde{\mu}_1, \tilde{\mu}_2\}$

mean-tail and CVaR sets of constraints simultaneously along with (possibly piece-wise constant) higher degree moments to emphasize high-dose regions.

6. CONCLUSION

In our findings, in Section 3, we begin by providing closed-form expressions for (Bound) for four specific sequences of raw and shifted moments. In general, such closed forms are not obtainable. We present three optimization reformulations for solving (Bound): the dual, a semi-definite program, and a linear program.

A shortcoming of these models is the lack of inclusion of patient geometry information from $\mathbf{d} = \mathbf{A}\mathbf{w}$. To this end, we present two MIPs in Section 4 that take both the reference DVH and

geometric information into account. These two models rely on either a voxel-based approach or a frequency-based approach in order to connect the distribution represented by the reference DVH and the physically realizable distribution captured by **d**. In both cases, we show that the inclusion of the patient geometry has the potential to drastically improve the bounds on DVH being approximated with (convex) moment constraints. In turn, this could provide a viable pathway to assess the efficacy of moment constraints to control the DVH and the dose distribution in clinical settings.

Lastly, we contrast two LP-based approaches to improving control over the tails of the distributions. The high-dose tail reflects the voxels in the OAR that receive the highest radiation doses, whereas these high-dose regions are particularly important for reducing toxicity in healthy tissues. We propose controlling the impact on these voxels through the addition of the relatively standard mean-tail or CVaR constraints to the novel frequency-based MIP. In our experiments, we observed that such high dose constraints are effective in reducing the number of voxels that receive high doses.

Our future work aims to address the computational difficulties faced by the models presented in this paper. We need to further investigate how our methods scale up with more realistic and finer voxel resolution. Our ultimate goal is to provide a computational tool to enable the quick selection of effective and tractable constraints to control the dose distribution in the context of inverse radiotherapy planning.

Data Availability

The code used for these experiments are at https://github.com/angela-r-morrison/radiation-therapy. All experiments done here used a toy dose deposition matrix **A** and were run on a laptop with AMD Ryzen 7 7730U with Radeon Graphics and 16GB of RAM.

Acknowledgements

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), RGPIN-2019-07199. The initial inspiration for this work was drawn during one of the author's post-doctoral appointment with Professor Tamás Terlaky.

References

[1] S. Gianfaldoni, R. Gianfaldoni, U. Wollina, J. Lotti, G. Tchernev, T. Lotti, An overview on radiotherapy: From its history to its current applications in dermatology, Macedonian J. Medical Sci. 5 (2017), 521–525.

[2] American Cancer Society, How radiation therapy is used to treat cancer, https://www.cancer.org/cancer/managing-cancer/treatment-types/radiation/basics.html, 2024. Last Revised: December 17, 2024. Accessed: May 20, 2025.

[3] National Cancer Institute, Radiation therapy to treat cancer, https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy, 2025. Accessed: May 20, 2025.

[4] Canadian Cancer Society, Radiation therapy, https://cancer.ca/en/treatments/treatment-types/radiation-therapy, 2025. Accessed: May 19, 2025.

[5] E. K. Lee, T. Fox, I. Crocker, Integer programming applied to intensity-modulated radiation therapy treatment planning, Ann. Oper. Res. 119 (2003), 165–181.

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

[7] W. Qiuwen, R. Mohan, A. Niemierko, Imrt optimization based on the generalized equivalent uniform dose (eud), In: Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, volume 1, pages 710–713. IEEE, 2000.

[8] C. Beong, J. Deasy, The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning, Phys. Med. Biol. 47 (2002), 3579–3589.
[9] Y. Zinchenko, T. Craig, H. Keller, T. Terlaky, M. Sharpe, Controlling the dose distribution with geud-type constraints within the convex radiotherapy optimization framework, Phys. Med. Biol. 53 (2008), 3231–3250.
[10] T. J. Stieltjes, Recherches sur les fractions continues, vol. 9, Annales de la Faculté des Sciences de Toulouse, 1895.
[11] H. Hamburger, Ueber eine Erweiterung der Stieltjes'schen Momentenproblems, I, Math. Ann. 81 (1920), 235–319.
[12] H. Hamburger, Ueber eine Erweiterung der Stieltjes'schen Momentenproblems, II, Math. Ann. 82 (1921), 168–187.
[13] F. Hausdorff, Summationsmethoden und momentfolgen I, Math. Zeitschrift, 9 (1921), 74–109.
[14] F. Hausdorff, Summationsmethoden und Momentfolgen. II, Math. Zeitschrift, 9 (1921), 280–299.
[15] D. Bertsimas, I. Popescu, Optimal inequalities in probability theory: A convex optimization approach, SIAM J. Optim. 15 (2005), 780–804.
[16] S. Weppler, Tight probability bounds for hausdorff random variables with applications to optimal cancer radiotherapy, Master's thesis, University of Calgary, Calgary, Canada, 2015.
[17] Y. Nesterov, Structure of non-negative polynomial and optimization problems, Technical report, Preprint DP 9749, Louvain-la-Neuve, Belgium, 1997.
[18] R. T. Rockafellar, S. Uryasev, Optimization of conditional value-at-risk, J. Risk, 2 (2000), 21–42.

APPENDIX A. CLOSED-FORM BOUNDS PROOF SKETCH

The proofs follow the approach outlined in [15]; while the complete proof is quite lengthy and tedious, here for brevity we present the key steps of the argument; more details can be found in [16]. The constraints in (Bound Dual) specify that a certain piece-wise polynomial curve must reside below a suitably defined step function on $[0.1]$. We exploit this fact and decompose the problem into a sequence of more elementary (linear) optimization models, where, in turn, each elementary model can be solved explicitly.

We start illustrating the case of having one raw and two shifted moments, $\{\mu_1, \tilde{\mu}_1, \tilde{\mu}_2\}$, with $\psi_1 < \psi_2$. Our objective is to maximize the expression $\mathbf{y}_0 + \mu_1\mathbf{y}_1 + \tilde{\mu}_1\mathbf{y}_2 + \tilde{\mu}_2\mathbf{y}_3$. The vertices of the absolute value-like function $g(x) = \mathbf{y}_0 + \mathbf{y}_1 x + \mathbf{y}_2|x-\psi_1| + \mathbf{y}_3|x-\psi_2|$ are given by $(\psi_1, \mathbf{y}_0 + \mathbf{y}_1\psi_1 + \mathbf{y}_3(\psi_2-\psi_1))$, $(\psi_2, \mathbf{y}_0 + \mathbf{y}_1\psi_2 + \mathbf{y}_2(\psi_2-\psi_1))$. Next, we examine the potential positioning of these vertices relative to $[0,\tau]$. Defining two planar regions $\mathscr{R}_1 = \{(\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^2 : \mathbf{x}_1 \leq \tau, \mathbf{x}_2 \leq 1\}$ and $\mathscr{R}_2 = \{(\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^2 : \mathbf{x}_1 \geq \tau, \mathbf{x}_2 \leq 0\}$ we consider the following three cases:

a. $\mathscr{R}_1$ contains both vertices,
b. $\mathscr{R}_1$ and $\mathscr{R}_2$ contain one vertex each, and
c. $\mathscr{R}_2$ contains both vertices.

Within each of these three cases, we consider 7 possibilities for $g(x)$ passing under $\mathbf{1}_{[0,\tau]}$:

1) $g(0) = 1, g(\tau) \leq 0, g(1) \leq 0$,
2) $g(0) \leq 1, g(\tau) = 0, g(1) \leq 0$,
3) $g(0) \leq 1, g(\tau) \leq 0, g(1) = 0$,
4) $g(0) = 1, g(\tau) = 0, g(1) \leq 0$,
5) $g(0) \leq 1, g(\tau) = 0, g(1) = 0$,
6) $g(0) = 1, g(\tau) \leq 0, g(1) = 0$,
7) $g(0) = 1, g(\tau) = 0, g(1) = 0$,

noting that in the remaining feasible case of having $g(0) < 1, g(\tau) < 0, g(1) < 0$, the corresponding $\mathbf{y}_i$ remain in the strict interior and the objective may be further improved. The figure below illustrates one generic feasible positioning of $g(x)$ in sub-case a.1). In total, this gives

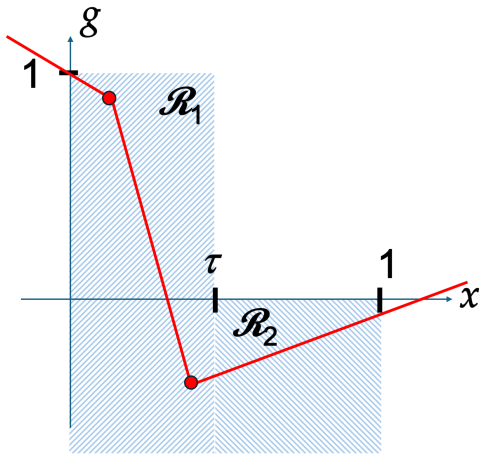


FIGURE 12. Generic $g(x)$ positioning in sub-case a.1)

us 21 possibilities to consider, where maximizing $\mathbf{y}_0 + \mu_1\mathbf{y}_1 + \tilde{\mu}_1\mathbf{y}_2 + \tilde{\mu}_2\mathbf{y}_3$ over the respective sub-region corresponds to solving an LP in three variables or less with no more than four inequalities (after the substitution of the boundary equality constraint to eliminate at least one of the $\mathbf{y}_i$ variables). The resulting LPs can each be solved explicitly, by either vertex enumeration or further parametrization with respect to the objective value to reduce the dimension by one. Combining the optimal solutions yields the result.

Similarly, in the case of two raw moments μ_1, μ_2, we fit a parabola with the vertex

$$\left(\frac{-\mathbf{y}_1}{2\mathbf{y}_2}, \frac{4\mathbf{y}_0\mathbf{y}_2 - \mathbf{y}_1^2}{4\mathbf{y}_2}\right)$$

while maximizing $\mathbf{y}_0 + \mu_1\mathbf{y}_1 + \mu_2\mathbf{y}_2$ and form a respective sub-case partitioning, etc.