

On the non-uniqueness of optimal radiation treatment plans

Christoph Börgers and Eric Todd Quinto

Department of Mathematics, Tufts University, Medford, MA 02155, USA

E-mail: borgers@math.tufts.edu and equinto@math.tufts.edu

Received 29 January 1999, in final form 5 May 1999

Abstract. The possibility of multiple locally optimal dose distributions in radiation treatment planning has been discussed and documented in the literature. Here we study a different question related to uniqueness: Is it possible for different treatment plans to generate the same dose distribution? For greatly simplified two-dimensional model problems, we show that the answer is ‘yes’ in regions where two or more beams intersect. In realistic problems, those are of course not the only regions of interest. However, as a result of cancellations in regions of intersection, substantial perturbations of beam profiles in certain directions may still have only small effects on the dose distribution. This is interesting because it offers an opportunity to optimize some other useful property, for instance simplicity, among all treatment plans generating a desired dose distribution with sufficient accuracy. We take a first step beyond our model problems by proving the stability of our results with respect to small perturbations of problem parameters. Since realistic problems differ from our model problems by much more than small perturbations, we plan to present a numerical study of more realistic examples in a sequel to this article.

1. Introduction

The possibility of multiple locally optimal dose distributions in radiation treatment planning has been discussed and documented in the literature; see [2, 6, 9, 14]. The source of this non-uniqueness is non-convexity. The set of realizable dose distributions is non-convex for instance, and most typically, if a prescribed number of beams is to be used but the beam positions and directions can be chosen freely. The objective function being minimized may be non-convex as well, for instance, and most typically, if a large dose to a small volume of healthy tissue is acceptable, but a moderate dose to a larger volume of healthy tissue is not.

In this paper, we study a different question related to uniqueness: Given an optimal dose distribution, is it possible that several different treatment plans generate it? For a single beam, reconstruction of the radiation intensity profile from the deposited dose is possible although ill-posed if a realistic model of dose deposition, including scattering, is used. For the simple models considered in this paper, which do not include scattering, it is trivial and well-posed. Our question therefore makes sense, strictly speaking at least, only for regions in which two or more beams intersect. In a realistic problem, these are of course not the only regions of interest. Typically, every beam will irradiate some region of the patient’s body that is not affected by any of the other beams, and as a result, different treatment plans will typically generate different dose distributions. (This is true even in the very simple example discussed in section 10.) However, as a result of cancellations in regions of intersection, substantial perturbations of beam profiles in certain directions may still have small effects on the dose distribution. One should then search, among all treatment plans generating a desired dose

distribution with acceptable accuracy, for one that optimizes some other useful property, for instance simplicity. Simplicity of radiation treatment plans is important because simpler plans are less error-prone and result in shorter treatment sessions.

An illustration is given by figures 1–5. A target, taken to be two-dimensional here and throughout this article, and a surrounding region Ω are irradiated from four different directions (figure 1). An optimization algorithm yields oscillatory beam intensity profiles (figure 2) generating the best dose distribution achievable with the given beam directions (figure 3). However, there exists a set of smoother intensity profiles (figure 4) yielding a nearly identical, but smoother dose distribution (figure 5). The details of the computational experiment underlying these figures are given in section 10.

In our model problems, a finite number n of broad beams is directed at a region Ω in the plane. We call an n -tuple of beam intensity profiles a *treatment plan*, and the mapping A from treatment plans to dose distributions the *dose operator*. A is linear, even in realistic problems. Its nullspace \mathcal{N} yields the directions in which beam intensity profiles can be perturbed without changing the dose distribution. We call the elements of \mathcal{N} the *null profiles*. Non-zero null profiles must involve negative intensities. Nevertheless they are of physical interest, since a *perturbation* of a physical, non-negative intensity profile by a null profile may of course result in another physical, non-negative profile.

In sections 2 and 3, we determine the null profiles for *semi-discrete* (discrete beam directions, continuous space variables) and *fully discrete* model problems with neither attenuation nor scattering of radiation. In all cases, the dimension of \mathcal{N} is substantial. In the fully discrete problems, in addition to discrete analogues of null profiles already present in the semi-discrete problems, we find spurious null profiles with oscillations on the scale of the spatial grid.

One might fundamentally object to studying nullspaces of simplified model operators, since nullspaces are unstable objects. For instance, the nullspaces of the matrices

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \tilde{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \epsilon \end{bmatrix} \quad \epsilon \neq 0$$

are perpendicular to each other, and the nullspaces of

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \tilde{A} = \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix} \quad \epsilon \neq 0$$

do not even have the same dimension. So what right do we have to hope that the nullspaces of our model dose operators have anything to do with reality?

We answer this objection as follows. Although *nullspaces* are unstable in the sense illustrated by these examples, *approximate nullspaces*, namely spans of right singular vectors associated with *isolated clusters of singular values near zero*, are stable. In section 4, we present two simple ways of making this precise. Thus, although realistic dose operators will not have *nullspaces* resembling those of our model dose operators, they may still have *approximate nullspaces* resembling the nullspaces of our model operators.

The estimates in section 4 depend on the inner products used. To apply them to our problem, we must therefore define inner products of treatment plans and dose distributions. The choices should be such that the induced norms sensibly measure the significance of perturbations in treatment plans and dose distributions. It is difficult to say anything rigorous about this, but in section 5 we discuss a heuristically reasonable way of defining inner products.

The gap σ between zero and the positive singular values is crucial in the perturbation estimates of section 4. In section 6, we therefore analyse σ for the semi-discrete model

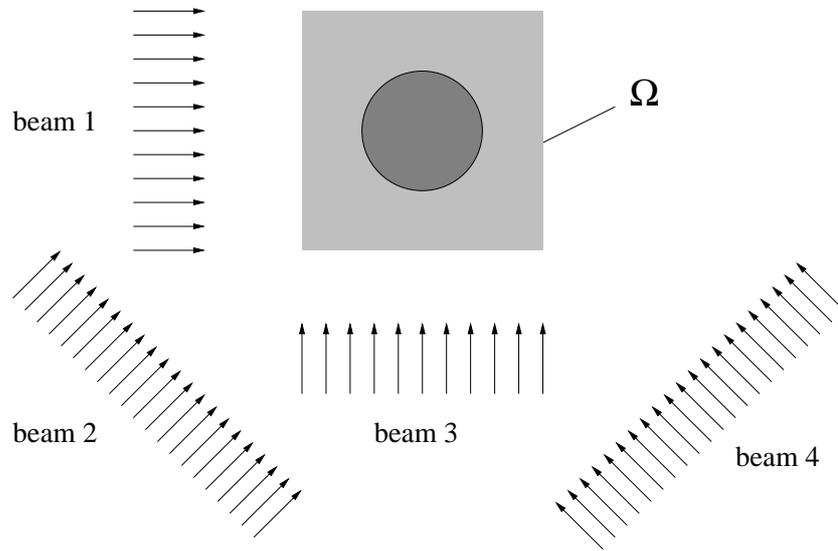


Figure 1. Four broad beams directed at a disk-shaped target inside a square region Ω .

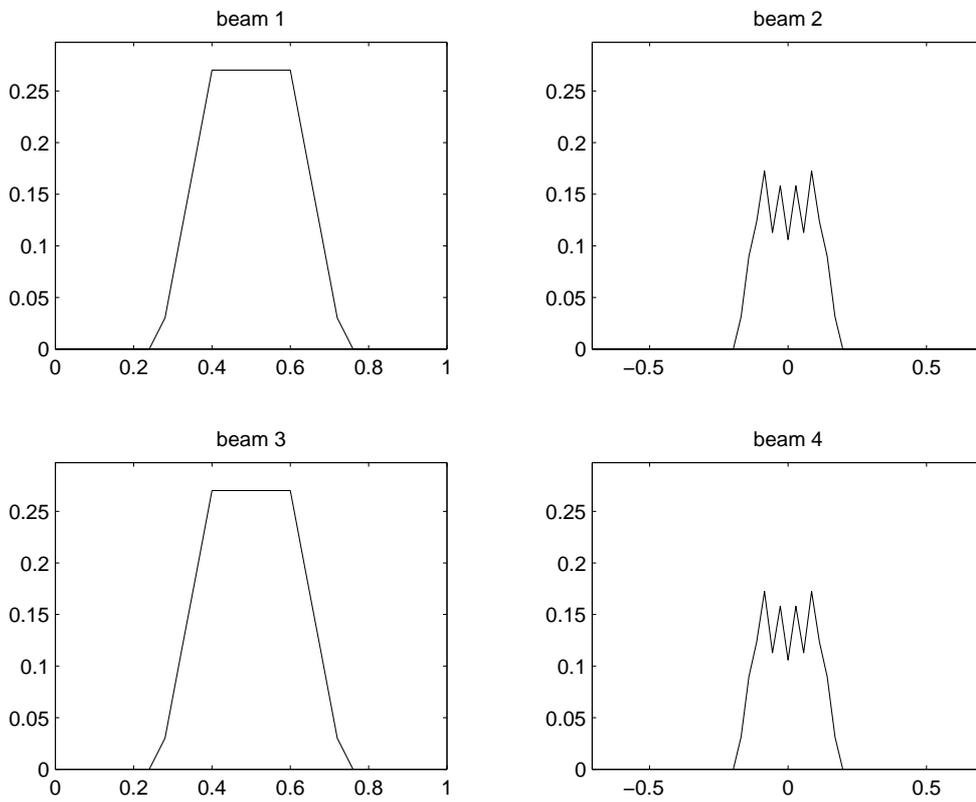


Figure 2. Optimal beam profiles.

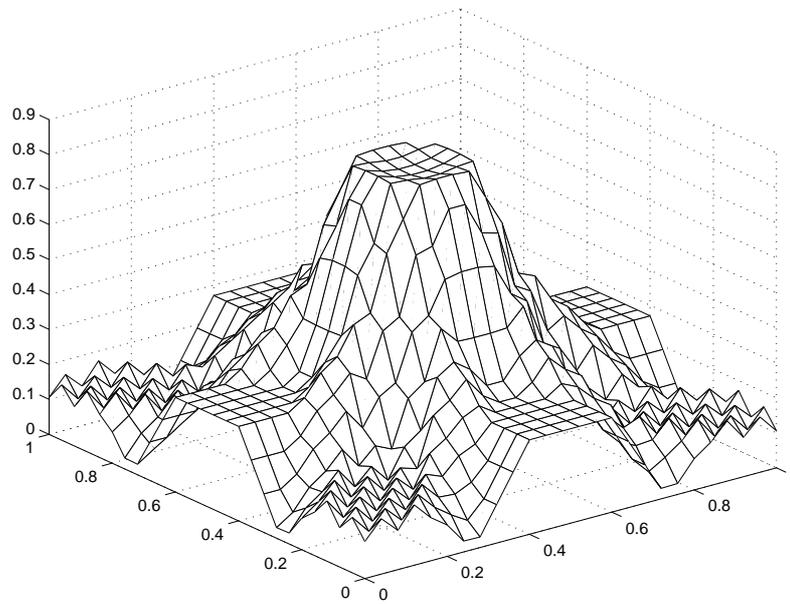


Figure 3. Optimal dose distribution.

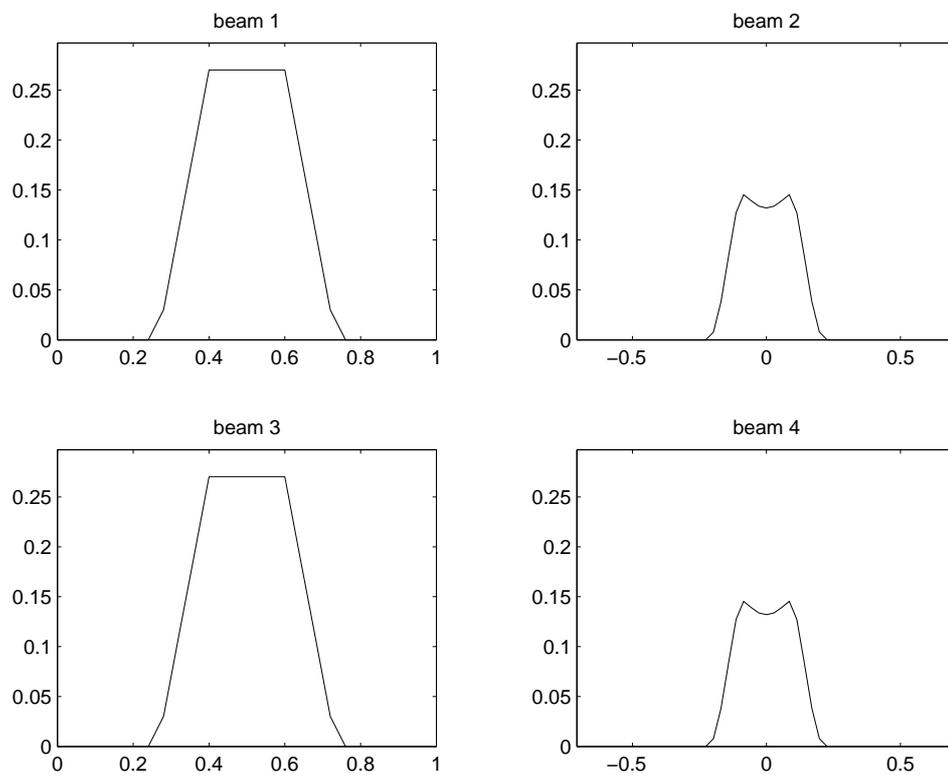


Figure 4. Smoother, nearly optimal beam profiles.

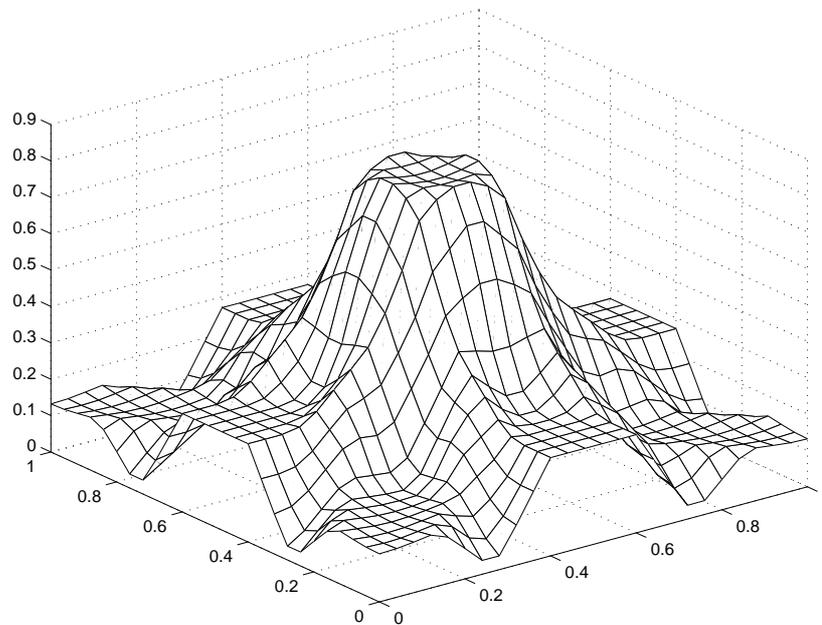


Figure 5. Smoother, nearly optimal dose distribution.

problem for the special case when the region Ω is a disk, and in particular show $\sigma > 0$. (This is non-obvious *a priori* because the domain of the semi-discrete dose operator is infinite-dimensional.) For fully discrete model problems, we restrict ourselves to a numerical study of the size of σ . In section 7, we present numerical experiments confirming that the nullspaces of our model dose operators turn into nearby approximate nullspaces when small perturbations in problem parameters are introduced.

The main thread of our article ends with section 7. The material in sections 8–10 is supplementary. In sections 8 and 9, we explicitly determine the nullspaces of model dose operators including constant positive beam attenuation. In section 10, we present the details of the optimization experiment referred to earlier (see figures 1–5).

In this paper, we only analyse model dose operators and small perturbations of those operators. However, realistic dose operators differ from our highly idealized ones by much more than just small perturbations. In particular, we say nothing about three dimensions, and nothing about scattering of radiation. Jointly with Dr James Satterthwaite (ADAC Laboratories and Tufts Medical School), we therefore plan to present in [3] a numerical study of more realistic dose operators from the point of view of this paper.

2. Semi-discrete model problem without attenuation

Let $\Omega \subseteq \mathbb{R}^2$ be open, and

$$0 \leq \theta_1 < \theta_2 < \dots < \theta_n < 2\pi.$$

Think of broad beams in the directions

$$\omega_\nu = (\cos \theta_\nu, \sin \theta_\nu) \quad \nu = 1, \dots, n$$

aimed at Ω , and let

$$\omega_v^\perp = (-\sin\theta_v, \cos\theta_v).$$

For $1 \leq v \leq n$, let g_v be real-valued functions or distributions defined on the open set

$$I_v = \{\omega_v^\perp \cdot \mathbf{x} : \mathbf{x} = (x, y) \in \Omega\} \subseteq \mathbb{R}.$$

Think of g_v as the radiation intensity profiles of the broad beams. We call the n -tuple

$$\mathbf{g} = (g_1, \dots, g_n)$$

a treatment plan, even if the g_v assume negative values. (As pointed out in the introduction, \mathbf{g} is of physical interest even if some g_v assume negative values because it is a possible *perturbation* of non-negative treatment plans.) To find the total dose deposited at a point $\mathbf{x} \in \Omega$, we sum the doses contributed by the individual beams:

$$A\mathbf{g}(\mathbf{x}) = \sum_{v=1}^n g_v(\omega_v^\perp \cdot \mathbf{x}). \quad (1)$$

Our aim is to determine

$$\mathcal{N} = \{\mathbf{g} : A\mathbf{g} \equiv 0 \text{ in } \Omega\}.$$

If $1 \leq v < \mu \leq n$ and $\theta_\mu = \theta_v + \pi$, then $\mathbf{g} = (0, \dots, 0, g_v, 0, \dots, 0, g_\mu, 0, \dots, 0) \in \mathcal{N}$ whenever $g_\mu(p) \equiv -g_v(-p)$. Thus pairs of parallel beams in opposite directions contribute to the nullspace in a trivial way, and we assume, for the remainder of section 2, that there are no such pairs.

We begin by observing that

$$(\omega_\mu \cdot \nabla)g_v(\omega_v^\perp \cdot \mathbf{x}) = (\omega_\mu \cdot \omega_v^\perp)g'_v(\omega_v^\perp \cdot \mathbf{x}) = \sin(\theta_\mu - \theta_v)g'_v(\omega_v^\perp \cdot \mathbf{x}).$$

We note that $\sin(\theta_\mu - \theta_v) \neq 0$ for $\mu \neq v$ because there are no parallel beams in opposite directions. Applying the directional derivatives $\omega_\mu \cdot \nabla$, $\mu \neq v$, to the equation $A\mathbf{g} \equiv 0$, we therefore obtain

$$g_v^{(n-1)} \equiv 0.$$

Thus g_v is a polynomial of degree $\leq n - 2$:

$$g_v(p) = \sum_{k=0}^{n-2} c_{vk} p^k. \quad (2)$$

We now need to examine which choices of polynomials g_v of degree $\leq n - 2$ actually yield null profiles. Consider first the example $n = 4$, $\theta_1 = 0$, $\theta_2 = \pi/4$, $\theta_3 = \pi/2$, $\theta_4 = 3\pi/4$. For a null profile, the four individual intensity profiles must be of the forms $a_1 + b_1 y + c_1 y^2$, $a_2 + b_2 x + c_2 x^2$, $a_3 + b_3(x - y) + c_3(x - y)^2$, and $a_4 + b_4(x + y) + c_4(x + y)^2$ and $A\mathbf{g} \equiv 0$ means

$$(a_1 + b_1 y + c_1 y^2) + (a_2 + b_2 x + c_2 x^2) + (a_3 + b_3(x - y) + c_3(x - y)^2) + (a_4 + b_4(x + y) + c_4(x + y)^2) \equiv 0.$$

By grouping together the coefficients in front of 1 , x , y , x^2 , xy and y^2 , this translates into the following six independent conditions

$$\begin{aligned} a_1 + a_2 + a_3 + a_4 &= 0 \\ b_2 + b_3 + b_4 &= 0 \\ b_1 - b_3 + b_4 &= 0 \\ c_2 + c_3 + c_4 &= 0 \\ -c_3 + c_4 &= 0 \\ c_1 + c_3 + c_4 &= 0. \end{aligned}$$

The generalization to n beams is straightforward. The k th powers in the equation $Ag = 0$ yield $k + 1$ constraints, for a total of $1 + 2 + 3 + \dots + (n - 1) = (n - 1)n/2$ constraints on the $(n - 1)n$ coefficients in the polynomials g_ν . The only remaining question is whether these are linearly independent constraints for all n . As illustrated by the example, it suffices to consider the k th powers, and examine, for a fixed k with $0 \leq k \leq n - 2$, whether the $k + 1$ conditions imposed by

$$\sum_{\nu=1}^n c_{\nu k} (-\sin \theta_\nu x + \cos \theta_\nu y)^k \equiv 0$$

on the coefficients $c_{\nu k}$ are linearly independent. These conditions can be written in the form

$$\mathcal{M}_{nk} \begin{bmatrix} c_{1k} \\ c_{2k} \\ \cdot \\ \cdot \\ c_{nk} \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \tag{3}$$

with

$$\mathcal{M}_{nk} = \begin{bmatrix} \sin^k \theta_1 \cos^0 \theta_1 & \cdot & \cdot & \cdot & \sin^k \theta_n \cos^0 \theta_n \\ \sin^{k-1} \theta_1 \cos^1 \theta_1 & \cdot & \cdot & \cdot & \sin^{k-1} \theta_n \cos^1 \theta_n \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sin^0 \theta_1 \cos^k \theta_1 & \cdot & \cdot & \cdot & \sin^0 \theta_n \cos^k \theta_n \end{bmatrix} \in \mathbb{R}^{(k+1) \times n}. \tag{4}$$

If $\sin \theta_\nu = 0$ or $\cos \theta_\nu = 0$ for some ν , the convention to be used in (4) is $0^0 = 1$. The question is whether the matrices \mathcal{M}_{nk} have full rank for all $k \in \{0, 1, \dots, n - 2\}$. By rotating the coordinate system if necessary, we may assume $\theta_1 = 0$. Then the first column in (4) is

$$\begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

and our question is reduced to whether the matrices

$$\begin{bmatrix} \sin^k \theta_2 \cos^0 \theta_2 & \cdot & \cdot & \cdot & \sin^k \theta_n \cos^0 \theta_n \\ \sin^{k-1} \theta_2 \cos^1 \theta_2 & \cdot & \cdot & \cdot & \sin^{k-1} \theta_n \cos^1 \theta_n \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sin^1 \theta_2 \cos^{k-1} \theta_2 & \cdot & \cdot & \cdot & \sin^1 \theta_n \cos^{k-1} \theta_n \end{bmatrix} \tag{5}$$

have full rank for $0 \leq k \leq n - 2$. If $\theta_1 = 0$, then $\sin \theta_\nu \neq 0$ for $\nu \neq 1$. Therefore, the rank of matrix (5) equals the rank of

$$\begin{bmatrix} \sin^{k-1} \theta_2 \cos^0 \theta_2 & \dots & \dots & \sin^{k-1} \theta_n \cos^0 \theta_n \\ \sin^{k-2} \theta_2 \cos^1 \theta_2 & \dots & \dots & \sin^{k-2} \theta_n \cos^1 \theta_n \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \sin^0 \theta_2 \cos^{k-1} \theta_2 & \dots & \dots & \sin^0 \theta_n \cos^{k-1} \theta_n \end{bmatrix}$$

which is of precisely the same form as (4), with the numbers of rows and columns reduced by one. By induction, we conclude that for any n , the matrices (4) have full rank for all $k \in \{0, 1, \dots, n - 2\}$. In summary, we have proved the following theorem.

Theorem 1. *For the semi-discrete model problem without attenuation, the null profiles are of the form given by equation (2), with coefficients $c_{\nu k}$, $0 \leq \nu \leq k - 2$, satisfying equation (3). The matrices M_{nk} in equation (3) (defined in equation (4)) have full rank. The dimension of \mathcal{N} consequently is $n(n - 1)/2$.*

Theorem 1 can be viewed as a special case of results that are standard in tomography and integral geometry; see, for instance [5, 11, 12]. We have presented a simple, self-contained proof because we wish to generalize this proof to fully discrete problems (section 3), and to problems with constant positive attenuation (sections 8 and 9).

3. Fully discrete model problem without attenuation

In real radiation treatment planning, the target domain and the beam profiles must be discretized. In this section, we consider a model problem in which the discrete target domain is the grid

$$\Gamma = h\mathbb{Z} \times h\mathbb{Z}$$

with $h > 0$. We restrict ourselves to beam directions that are either aligned with the coordinate axes, or form angles of 45° with them. As in section 2, and for the same reason, we assume that there are no parallel beams in opposite directions. Thus, we use the four angles

$$\theta_\nu = (\nu - 1)\pi/4 \quad 1 \leq \nu \leq 4.$$

The beam profiles are discretized on the grids

$$\Gamma_\nu = \{jh_\nu : j \in \mathbb{Z}\} \subseteq \mathbb{R}$$

with

$$h_\nu = \begin{cases} h & \text{for odd } \nu \\ h/\sqrt{2} & \text{for even } \nu. \end{cases}$$

Figure 6 illustrates this.

Let g_ν be functions from Γ_ν into \mathbb{R} , and $\mathbf{g} = (g_1, g_2, g_3, g_4)$. For $\mathbf{x} \in \Gamma$, we define

$$A\mathbf{g}(\mathbf{x}) = \sum_{\nu=1}^4 g_\nu(\boldsymbol{\omega}_\nu^\perp \cdot \mathbf{x}). \tag{6}$$

The terms in this sum are well-defined because $\boldsymbol{\omega}_\nu^\perp \cdot \mathbf{x} \in \Gamma_\nu$ for $\mathbf{x} \in \Gamma$, $1 \leq \nu \leq 4$. Our aim in this section is to determine

$$\mathcal{N} = \{\mathbf{g} : A\mathbf{g}(\mathbf{x}) = 0 \text{ for all } \mathbf{x} \in \Gamma\}.$$

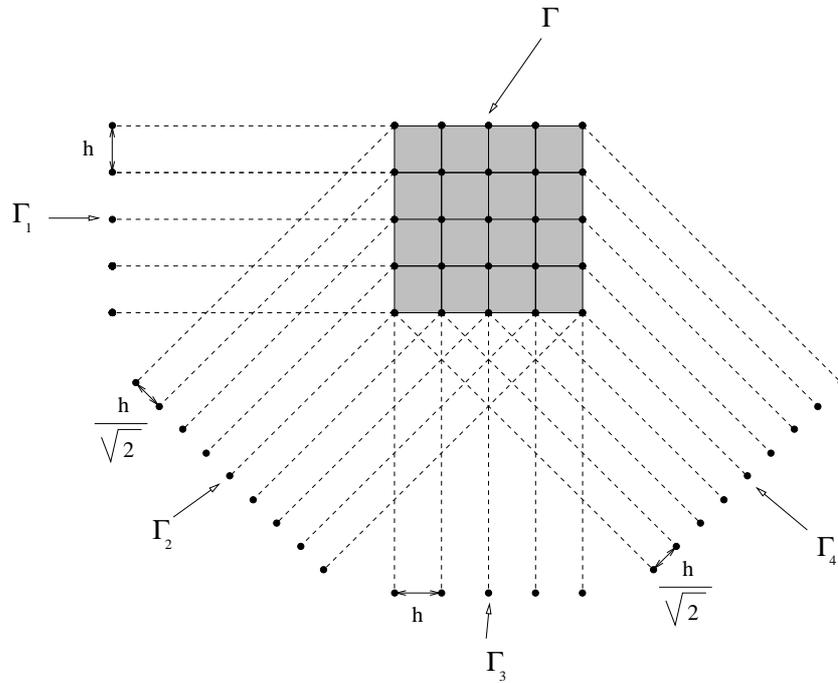


Figure 6. Grids used in the fully discrete model problems.

Our argument will be a discrete analogue of that in section 2, with the directional derivatives $\omega_v \cdot \nabla$ replaced by upstream difference quotients.

Assume that $Ag(x) = 0$ for all $x \in \Gamma$. For $(k, l) \in \mathbb{Z} \times \mathbb{Z}$, we write $x_{k,l} = (kh, lh)$ and

$$D_{k,l} = Ag(x_{k,l}) = \sum_{v=1}^4 g_v(\omega_v^\perp \cdot x_{k,l}).$$

We define ∂_ν , $\nu = 1, 2, 3, 4$, to be the upstream first-order difference discretizations of the directional derivatives $\omega_\nu \cdot \nabla$. That is, for a function u from Γ to \mathbb{R} , writing $u_{k,l} = u(kh, lh)$:

$$\partial_1 u_{k,l} = \frac{u_{k,l} - u_{k-1,l}}{h} \tag{7}$$

$$\partial_2 u_{k,l} = \frac{u_{k,l} - u_{k-1,l-1}}{\sqrt{2}h} \tag{8}$$

$$\partial_3 u_{k,l} = \frac{u_{k,l} - u_{k,l-1}}{h} \tag{9}$$

$$\partial_4 u_{k,l} = \frac{u_{k,l} - u_{k+1,l-1}}{\sqrt{2}h}. \tag{10}$$

For $\nu \in \{1, 3\}$, by applying the operators ∂_μ with $\mu \neq \nu$ to the equation $D_{k,l} = 0$, we obtain after a small amount of algebra:

$$\frac{g_\nu(q + 3h/2) - 3g_\nu(q + h/2) + 3g_\nu(q - h/2) - g_\nu(q - 3h/2)}{h^3} = 0 \tag{11}$$

if $q \pm h/2 \in \Gamma_\nu$. Similarly, for $\nu \in \{2, 4\}$, by applying the operators ∂_μ with $\mu \neq \nu$ we obtain:

$$\frac{g_\nu(p + 2h/\sqrt{2}) - 2g_\nu(p + h/\sqrt{2}) + 2g_\nu(p - h/\sqrt{2}) - g_\nu(p - 2h/\sqrt{2})}{2(h/\sqrt{2})^3} = 0 \tag{12}$$

if $p \in \Gamma_\nu$. Equation (11) is a second-order accurate finite-difference discretization of $g_\nu^{(3)}(q) = 0$. The functions $1, p$ and p^2 ($p \in \Gamma_\nu$) belong to its solution space. Since they are linearly independent, they span the solution space. Similarly, equation (12) is a second-order accurate finite-difference discretization of $g_\nu^{(3)}(p) = 0$, and the functions $1, p$ and p^2 ($p \in \Gamma_\nu$) belong to its solution space. However, the solution space for equation (12) is four-dimensional. It is not hard to see that $(-1)^{p/h_\nu}$ is an additional solution, linearly independent of $1, p$ and p^2 . This proves that the null profiles have the form

$$g_\nu(p) = \sum_{\gamma=0}^2 c_{\nu\gamma} p^\gamma \tag{13}$$

for $\nu \in \{1, 3\}$, and

$$g_\nu(p) = \sum_{\gamma=0}^2 c_{\nu\gamma} p^\gamma + d_\nu (-1)^{p/h_\nu} \tag{14}$$

for $\nu \in \{2, 4\}$.

$Ag(x_{k,l}) = 0$ is now equivalent to

$$\sum_{v=1}^4 \sum_{\gamma=0}^2 c_{v\gamma} (\omega_v^\perp \cdot x_{k,l})^\gamma + d_2 (-1)^{k+l} + d_4 (-1)^{k-l} = 0. \tag{15}$$

Since $(-1)^{k-l} = (-1)^{k+l}$, and since the grid function $u(x_{k,l}) = (-1)^{k+l}$ is not an element of the span of the grid functions $(\omega_v^\perp \cdot x_{k,l})^\gamma$, equation (15) holds if and only if

$$\sum_{v=1}^4 \sum_{\gamma=0}^2 c_{v\gamma} (\omega_v^\perp \cdot x_{k,l})^\gamma = 0 \tag{16}$$

and

$$d_2 = -d_4.$$

The proof of theorem 1 shows that equation (16) holds for all $x_{k,l} \in \Gamma$ if and only if

$$\mathcal{M}_{4k} \begin{bmatrix} c_{1k} \\ c_{2k} \\ c_{3k} \\ c_{4k} \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ 0 \end{bmatrix} \tag{17}$$

for $0 \leq k \leq 2$. The matrices $\mathcal{M}_{nk} \in \mathbb{R}^{(k+1) \times n}$ were defined in equation (4). Recall that they have full rank. Thus (17) imposes six independent conditions on the twelve coefficients $c_{v\gamma}$. Recalling the extra degree of freedom d_2 (which is the same as $-d_4$), we conclude $\dim \mathcal{N} = 12 - 6 + 1 = 7$. In summary, we have proved the following theorem.

Theorem 2. *For the fully discrete model problem without attenuation, the null profiles are given by equations (13) and (14), where $c_{v\gamma}$ are real coefficients satisfying condition (17), and d_ν are real coefficients with $d_2 = -d_4$. The dimension of \mathcal{N} is seven.*

This easily generalizes to finite grids. For example, theorem 2 remains unchanged if

$$\Gamma = \{0, h, 2h, \dots, 1 - h, 1\} \times \{0, h, 2h, \dots, 1 - h, 1\} \subseteq [0, 1]^2 \tag{18}$$

$h = 1/N, N \geq 3$ integer,

$$\Gamma_\nu = \{sh_\nu : s \in \mathbb{Z}, 0 \leq s \leq N\} \subseteq [0, 1] \tag{19}$$

for $\nu \in \{1, 3\}$, and

$$\Gamma_\nu = \{sh_\nu : s \in \mathbb{Z}, -N \leq s \leq N\} \subseteq [-1/\sqrt{2}, 1/\sqrt{2}] \tag{20}$$

for $\nu \in \{2, 4\}$. (For $N < 3$, our proof no longer makes sense, since equation (11) requires that Γ_1 and Γ_3 contain at least four points. But in fact, $\dim \mathcal{N} = 7$ even for $N = 2$.)

4. Perturbation estimates for approximate nullspaces

Here we will analyse the stability of the results of sections 2 and 3 under small perturbations of the model operators. Our analysis is based on the following observation. If \mathcal{N} is the nullspace of A , and the smallest positive singular value σ of A is of substantial size relative to the norm of A , then any \tilde{A} near A has an approximate nullspace $\tilde{\mathcal{N}}$ near \mathcal{N} . One can make this statement quantitative in several ways. First, we note that $\tilde{\mathcal{N}}$ can simply be taken to be \mathcal{N} :

Theorem 3. *Let V and W be Hilbert spaces. Let $\|\cdot\|$ denote the norms on V and W , as well as the induced norm on the space $\mathcal{B}(V, W)$ of bounded linear operators $V \rightarrow W$. Let $A \in \mathcal{B}(V, W)$, let $\mathcal{N} \subseteq V$ be the nullspace of A , and define*

$$\sigma = \inf_{v \perp \mathcal{N}, \|v\|=1} \|Av\| > 0.$$

Then for $\tilde{A} \in \mathcal{B}(V, W)$,

$$\sup_{v \in \mathcal{N}, \|v\|=1} \|\tilde{A}v\| \leq \|A - \tilde{A}\|$$

and

$$\inf_{v \in \mathcal{N}^\perp, \|v\|=1} \|\tilde{A}v\| \geq \sigma - \|A - \tilde{A}\|.$$

A second, somewhat less obvious estimate relates the nullspace of A to the span of the right singular vectors of \tilde{A} corresponding to small singular values of \tilde{A} . We present this in the finite-dimensional case only. Thus, let m and n be integers, $m \geq n \geq 1$, and let inner products on \mathbb{R}^m and \mathbb{R}^n be given, not necessarily the Euclidean ones. We denote by $\|\cdot\|$ the induced norms on \mathbb{R}^m and \mathbb{R}^n , as well as the induced matrix norm on $\mathbb{R}^{m \times n}$. Let $A \in \mathbb{R}^{m \times n}$, with

$$\begin{aligned} Av_i &= \sigma_i u_i & i &= 1, \dots, n \\ \sigma_1 &\geq \sigma_2 \geq \dots \geq \sigma_{n-k} > \sigma_{n-k+1} = \dots = \sigma_n = 0, \end{aligned}$$

$u_i \in \mathbb{R}^m$ and $v_i \in \mathbb{R}^n$ orthonormal (with respect to the given inner products on \mathbb{R}^m and \mathbb{R}^n). Similarly let $\tilde{A} \in \mathbb{R}^{m \times n}$, with

$$\begin{aligned} \tilde{A}\tilde{v}_i &= \tilde{\sigma}_i \tilde{u}_i & i &= 1, \dots, n \\ \tilde{\sigma}_1 &\geq \tilde{\sigma}_2 \geq \dots \geq \tilde{\sigma}_{n-k} \geq \tilde{\sigma}_{n-k+1} \geq \dots \geq \tilde{\sigma}_n \geq 0, \end{aligned}$$

$\tilde{u}_i \in \mathbb{R}^m$ and $\tilde{v}_i \in \mathbb{R}^n$ orthonormal. Let \mathcal{N} denote the nullspace of A , i.e. the span of v_i with $n - k + 1 \leq i \leq n$, and $\tilde{\mathcal{N}}$ the span of \tilde{v}_i with $n - k + 1 \leq i \leq n$. Note that $\tilde{\mathcal{N}}$ is not the nullspace of \tilde{A} in general. Our aim is a bound on the angle

$$\varphi = \max_{\tilde{v} \in \tilde{\mathcal{N}}, \|\tilde{v}\|=1} \min_{v \in \mathcal{N}, \|v\|=1} \cos^{-1}(v \cdot \tilde{v}) \in [0, \pi/2] \tag{21}$$

between \mathcal{N} and $\tilde{\mathcal{N}}$. (The dot in equation (21) denotes the given inner product on \mathbb{R}^n .) We will use the following well-known perturbation estimate for singular values; see for instance corollary 5.1 of [7].

Weyl's theorem.

$$|\sigma_i - \tilde{\sigma}_i| \leq \|A - \tilde{A}\|$$

for $1 \leq i \leq n$.

Any vector $\tilde{v} \in \tilde{\mathcal{N}}$ with $\|\tilde{v}\| = 1$ permits a unique decomposition

$$\tilde{v} = \sqrt{1 - \epsilon^2} v + \epsilon v^\perp$$

with $v \in \mathcal{N}$, $\|v\| = 1$, $v^\perp \in \mathcal{N}^\perp$, $\|v^\perp\| = 1$, $\epsilon \in [0, 1]$. An upper bound on ϵ valid for all \tilde{v} is an upper bound on $\sin \varphi$. We derive such a bound as follows:

$$\begin{aligned} \|A(\epsilon v^\perp)\| &= \|A\tilde{v} - \sqrt{1 - \epsilon^2}Av\| \\ &= \|A\tilde{v}\| \\ &= \|(A - \tilde{A})\tilde{v} + \tilde{A}\tilde{v}\| \\ &\leq \|A - \tilde{A}\| + \tilde{\sigma}_{n-k+1} \\ &\leq \|A - \tilde{A}\| + \sigma_{n-k+1} + \|A - \tilde{A}\| \\ &= 2\|A - \tilde{A}\|. \end{aligned}$$

The last inequality follows from Weyl's theorem. Since $\|A(\epsilon v^\perp)\| \geq \epsilon\sigma_{n-k}$, we conclude

$$\epsilon \leq \frac{2}{\sigma_{n-k}} \|A - \tilde{A}\|.$$

In summary, we have proved the following theorem.

Theorem 4. *Let $A \in \mathbb{R}^{m \times n}$ have a k -dimensional nullspace \mathcal{N} , $k \geq 0$. Let $\sigma = \sigma_{n-k}$ be the smallest positive singular value of A . Let $\tilde{A} \in \mathbb{R}^{m \times n}$, and let $\tilde{\mathcal{N}}$ denote the span of the right singular vectors of \tilde{A} associated with its k smallest singular values. Let φ denote the angle between \mathcal{N} and $\tilde{\mathcal{N}}$. Then*

$$\sin \varphi \leq \frac{2}{\sigma/\|A\|} \frac{\|A - \tilde{A}\|}{\|A\|}. \quad (22)$$

The example

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \tilde{A} = \begin{bmatrix} \delta & 0 \\ 0 & 1 - \delta \end{bmatrix}$$

with $\delta < 1/2$, $\delta \rightarrow 1/2$, $k = 1$, and using the Euclidean norm on \mathbb{R}^2 , shows that inequality (22) is sharp in the sense that 2 is the smallest uniformly valid upper bound on $(\sigma \sin \varphi)/\|A - \tilde{A}\|$. From theorem 4.4 of [16], one can derive

$$\sin \varphi \leq \left[1 - \frac{\|A - \tilde{A}\|}{\sigma} \right]^{-1} \frac{\|A - \tilde{A}\|}{\sigma}. \quad (23)$$

Note that both (22) and (23) are meaningless for $\|A - \tilde{A}\| > \sigma/2$, since the bounds are then greater than 1. For $0 < \|A - \tilde{A}\| < \sigma/2$, (23) is a slightly better estimate than (22).

5. Choice of inner products for the radiation treatment planning problem

To apply the estimates of section 4 to our problem, we must choose inner products on the spaces of treatment plans and dose distributions. These choices are essential to our results, since they determine the meaning of the word 'small' when we talk about small perturbations of treatment plans or dose distributions.

Ideally, we would like to call a perturbation of a dose distribution 'small' if and only if it has little impact on the effect of the treatment. Therefore the choice of inner product on the space of dose distribution is just as problematic as the formulation of the optimization problem itself. For lack of a better option, we will simply use the L^2 -product

$$\langle D_1, D_2 \rangle = (D_1, D_2)_{L^2(\Omega)} = \int_{\Omega} D_1(x) D_2(x) dx$$

for dose distributions D_1 and D_2 , or trapezoidal approximations of it for fully discrete model problems. The choice of the L^2 -norm is common in the literature on radiation treatment

planning. It is not fully satisfactory, and in fact no choice of norm is; see [2] for further discussion.

Once an inner product for dose distributions has been chosen, there is a natural way of deriving an inner product for treatment plans from it. We should call a perturbation of the intensity profile of a *single* broad beam ‘small’ if and only if it leads to a small perturbation of the deposited dose. Therefore we should define the inner product of two treatment plans involving a single broad beam direction to be the inner product of the dose distributions generated by the two plans. For treatment plans involving n broad beams, we then define the inner product by summing over the beams. To state this formally, let $\mathbf{g}_1 = (g_{11}, g_{21}, \dots, g_{n1})$ and $\mathbf{g}_2 = (g_{12}, g_{22}, \dots, g_{n2})$ be treatment plans, and let us define

$$\mathbf{g}_i^{(v)} = (0, \dots, 0, g_{vi}, 0 \dots 0)$$

for $i = 1, 2$, and $v = 1, \dots, n$. Then

$$\langle \mathbf{g}_1, \mathbf{g}_2 \rangle = \sum_{v=1}^n \langle A\mathbf{g}_1^{(v)}, A\mathbf{g}_2^{(v)} \rangle.$$

It is straightforward to verify that this is an inner product. In the semi-discrete case without attenuation, with

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\},$$

if the L^2 -product is used for dose distributions, the inner product of treatment plans becomes

$$\langle \mathbf{g}_1, \mathbf{g}_2 \rangle = \sum_{v=1}^n \int_{-1}^1 g_{v1}(p)g_{v2}(p)2\sqrt{1-p^2} dp. \tag{24}$$

6. The smallest positive singular value of the semi-discrete model dose operator

As shown in section 4, a gap σ between zero and the positive singular values of A implies that perturbations turn the nullspace of A into a nearby approximate nullspace of the perturbed operator \hat{A} . We therefore study the size of σ for model dose operators now.

We begin with the semi-discrete dose operator without attenuation described in section 2, with

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}.$$

In contrast with section 2, we find it convenient here to allow parallel beams in opposite directions. We define inner products as discussed in section 5. The singular value decomposition of A , with respect to these inner products, is obtained by a calculation that can in essence be found in several places in the literature, for instance in section 3 of [5], or in section IV.3 of [13]. The result is as follows.

Theorem 5. For $k = 1, 2, \dots$, let U_k be defined by

$$U_k(p) = \frac{1}{\sqrt{\pi}} \frac{\sin(k \cos^{-1} p)}{\sin(\cos^{-1} p)} \quad p \in (-1, 1)$$

and

$$C_{kn} = \left(\frac{\sin(k(\theta_v - \theta_\mu))}{k \sin(\theta_v - \theta_\mu)} \right)_{1 \leq v, \mu \leq n} \in \mathbb{R}^{n \times n}$$

with the convention

$$\frac{\sin(k(\theta_v - \theta_\mu))}{k \sin(\theta_v - \theta_\mu)} = \begin{cases} 1 & \text{if } \theta_v = \theta_\mu \\ (-1)^{k-1} & \text{if } \theta_v = \theta_\mu \pm \pi. \end{cases}$$

For $k = 1, 2, \dots$, let

$$\mathbf{a}_{k\mu} = \begin{bmatrix} a_{k\mu 1} \\ a_{k\mu 2} \\ \cdot \\ \cdot \\ a_{k\mu n} \end{bmatrix} \in \mathbb{R}^n \quad \mu = 1, \dots, n$$

be orthonormal eigenvectors of C_{kn} with eigenvalues $\lambda_{k\mu}$. Let

$$\mathbf{g}_{k\mu} = \begin{bmatrix} a_{k\mu 1} U_k(p) \\ a_{k\mu 2} U_k(p) \\ \cdot \\ \cdot \\ a_{k\mu n} U_k(p) \end{bmatrix} \quad k = 1, 2, \dots, \quad \mu = 1, \dots, n.$$

Then the $\mathbf{g}_{k\mu}$ form an orthonormal basis of the space of treatment plans. Their images

$$D_{k\mu} = A \mathbf{g}_{k\mu}$$

are orthogonal in $L^2(\Omega)$, and

$$\|D_{k\mu}\|_{L^2(\Omega)} = \sqrt{\lambda_{k\mu}}.$$

Up to normalization, the U_k are the Chebyshev polynomials of the second kind [1]. The normalization is such that U_k has norm 1 in $L^2((-1, 1); 2\sqrt{1-p^2})$, the weighted L^2 -space on the interval $(-1, 1)$ with weight $2\sqrt{1-p^2}$. The $\mathbf{g}_{k\mu}$ are the right singular vectors of A , the $D_{k\mu}$ are the left singular vectors and

$$\sigma_{k\mu} = \sqrt{\lambda_{k\mu}}$$

are the singular values. Note that

$$\lim_{k \rightarrow \infty} C_{kn}$$

exists for each fixed n . Therefore:

Corollary 1. For any choice of n and θ_v , $1 \leq v \leq n$,

$$\sigma = \inf\{\sigma_i^{(k)} : \sigma_i^{(k)} > 0\} > 0.$$

We next consider the special case of evenly spaced θ_v

$$\theta_v = v \frac{2\pi}{n} \quad v = 1, \dots, n$$

more carefully. In this case, the matrices C_{kn} are circulant, that is, the (v, μ) th entry in C_{kn} depends on $v - \mu \pmod n$ only. Their eigenvalues can therefore be calculated explicitly using discrete Fourier analysis. This calculation is completely elementary but tedious; we therefore omit it. Figures 7 and 8 show the singular values of A for $n = 19$ and $n = 20$. The explicit calculation of the eigenvalues of the matrices C_{kn} also yields an explicit expression for the size of the smallest positive singular value of A :

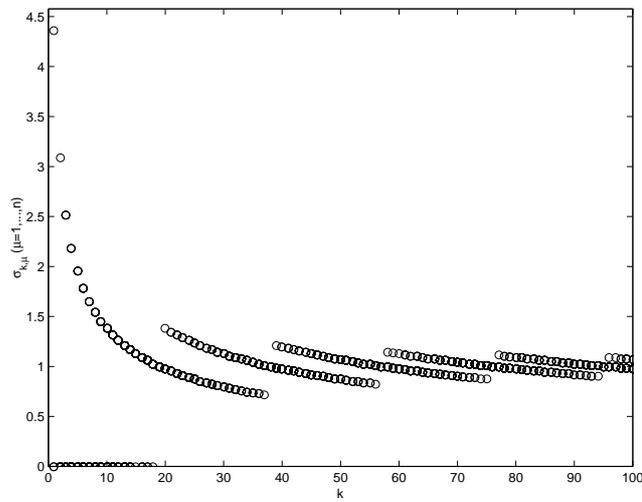


Figure 7. Singular values for 19 equispaced angles.

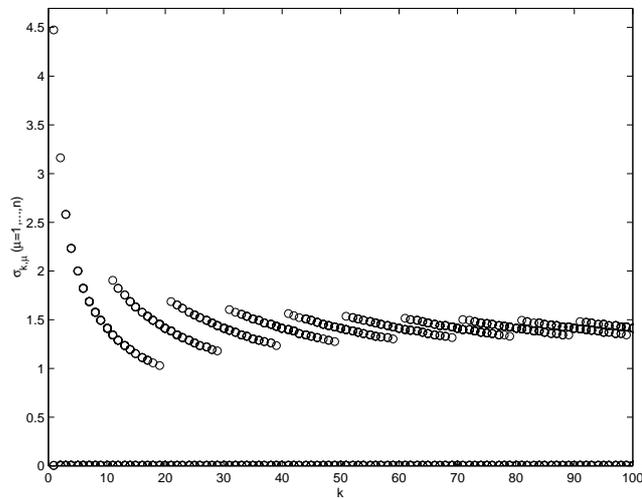


Figure 8. Singular values for 20 equispaced angles.

Theorem 6. For evenly spaced angles

$$\sigma = \sqrt{\frac{n}{2n-1}}$$

if n is odd, and

$$\sigma = \sqrt{\frac{n}{n-1}}$$

if n is even.

Thus $\sigma > 1/\sqrt{2}$ for all n . To readers familiar with the singular value decomposition of the dual Radon transform, this may seem puzzling for the following reason. It is well-known

(see for instance p 101 of [13]) that the positive singular values of the mapping

$$R^\# : L^2((-1, 1) \times (0, 2\pi), \sqrt{1 - p^2}) \rightarrow L^2(\Omega) \tag{25}$$

defined by

$$(R^\#g)(x, y) = \int_0^{2\pi} g(-\sin \theta x + \cos \theta y, \theta) d\theta \tag{26}$$

are

$$\sqrt{\frac{4\pi}{k}} \quad k = 1, 2, 3, \dots \tag{27}$$

Thus there is no gap between zero and the positive singular values. However, A is simply a discrete analogue of $R^\#$. Should not one therefore expect $\sigma \rightarrow 0$ as $n \rightarrow \infty$?

The explanation is as follows. The semi-discrete dose operator A is a discretization of $R^\#$ only *up to scaling*. First, the sum in equation (24) would be a discrete analogue of an integral with respect to θ only if it were multiplied by $\Delta\theta = 2\pi/n$. Second, equation (1) would be analogous to equation (26) only if it were multiplied by a factor of $\Delta\theta$. Third, the weight in equation (24) is $2\sqrt{1 - p^2}$, motivated by our discussion in section 5, whereas the weight in (25) is $\sqrt{1 - p^2}$, following standard notational convention. Altogether, not $\sigma_{k\mu}$ but

$$\frac{1}{\sqrt{\Delta\theta}} \Delta\theta \sqrt{2} \sigma_{k\mu} = \sqrt{2\Delta\theta} \sigma_{k\mu} \tag{28}$$

should be expected to approximate the non-zero singular values (27) of $R^\#$ as $n \rightarrow \infty$. Figures 9 and 10 confirm and illustrate this by showing, for $N = 20$ and $N = 40$, the scaled singular values $\sqrt{2\Delta\theta} \sigma_{k\mu}$ of A together with the singular values $\sqrt{4\pi/k}$ of $R^\#$.

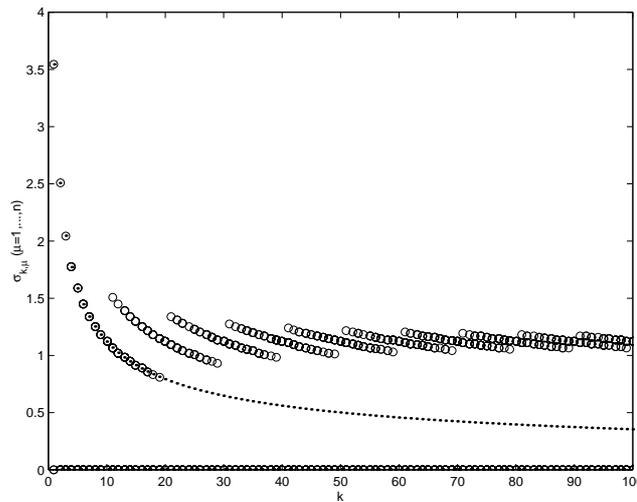


Figure 9. Circles: singular values for 20 equispaced angles, scaled as in equation (28). Dots: singular values of continuous operator.

For the fully discrete dose operator without attenuation defined in section 3, using the finite grids given by equations (18)–(20) with $N = 20$, table 1 shows, for various values of N , the smallest positive singular value σ and the norm $\|A\|$ of the dose operator (i.e. the largest singular value). The inner products used in these computations were described in section 5.

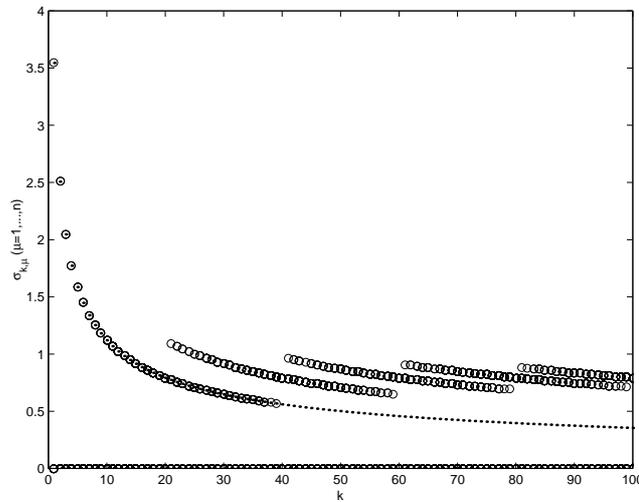


Figure 10. Circles: singular values for 40 equispaced angles, scaled as in equation (28). Dots: singular values of continuous operator.

Table 1. σ = smallest positive singular value of A , and $\|A\|$.

N	σ	$\ A\ $
8	0.72	2.0
16	0.70	2.0
24	0.69	2.0
32	0.69	2.0
40	0.69	2.0

7. A numerical study of small perturbations of the fully discrete model problem

Here we consider perturbations of the fully discrete dose operator defined in section 3. We use the finite grids (18)–(20) and assume that a piecewise continuous function $a = a(x, y) \geq 0$, $(x, y) \in \Omega = (0, 1)^2$, is given. For $\mathbf{x} = (x, y) \in (0, 1)^2$, we define

$$d_v(\mathbf{x}) = \min\{s > 0 : \mathbf{x} - s\boldsymbol{\omega}_v \notin (0, 1)^2\},$$

and

$$\tilde{A}g(\mathbf{x}) = \sum_{v=1}^4 g_v(\boldsymbol{\omega}_v^\perp \cdot \mathbf{x}) \exp\left(-\int_0^{d_v(\mathbf{x})} a(\mathbf{x} - s\boldsymbol{\omega}_v) ds\right);$$

compare this with equation (6). The exponential factor models beam attenuation. Spatially varying attenuation coefficients are important in radiation treatment planning because radiation passes through different materials, such as soft tissue, bones and air.

To simplify our numerical experiments, we assume that a is constant on

$$Q_{kl} = \{(x, y) \in (0, 1)^2 : |x - kh| < h/2, |y - lh| < h/2\}$$

for all integers k and l with $0 \leq k, l \leq N$. We consider two choices of a . The first is

$$a_1(x, y) = 0.3 \quad \text{for all } (x, y).$$

Thus, the factor by which the beam is attenuated as it penetrates the square $(0, 1)^2$ along one of its diagonals is $\exp(-0.3\sqrt{2}) \approx 0.65$, a physically realistic value. The second choice we

consider is

$$a_2(x, y) = \begin{cases} 0 & \text{if } (x, y) \in Q_{kl}, (kh - 0.3)^2 + (lh - 0.4)^2 < 0.25^2 \\ 0.3 & \text{otherwise.} \end{cases}$$

This is reminiscent of a non-attenuating air pocket in attenuating soft tissue. We denote the corresponding perturbed operators by \tilde{A}_1 and \tilde{A}_2 .

The operator A is the special case $a(x, y) = 0$. We know from section 3 that it has a seven-dimensional nullspace \mathcal{N} . For the operators \tilde{A}_1 and \tilde{A}_2 , we denote the spans of the right singular vectors associated with the seven smallest singular values by $\tilde{\mathcal{N}}_1$ and $\tilde{\mathcal{N}}_2$, respectively. (In section 9, we will show that $\tilde{\mathcal{N}}_1$ is in fact the nullspace of \tilde{A}_1 .)

Table 2 refers to the comparison between \mathcal{N} and $\tilde{\mathcal{N}}_1$. The second column shows, for various values of N , the angle φ between \mathcal{N} and $\tilde{\mathcal{N}}_1$. The third column shows the upper bound on this angle obtained from theorem 4. The fourth column shows the biggest discrepancy between a singular value of the operator with $a = 0$, and the corresponding singular value of the operator with $a = 0.3$ (the quantity that Weyl’s theorem estimates), and the fifth column shows the upper bound on the fourth column given by Weyl’s theorem. The table shows Weyl’s theorem to be almost sharp. The estimate of theorem 4 is pessimistic here. Table 3 shows similar results for $\tilde{\mathcal{N}}_2$.

Table 2. Comparison between \mathcal{N} , the nullspace of A ($a \equiv 0$), and $\tilde{\mathcal{N}}_1$, the nullspace of \tilde{A}_1 ($a \equiv 0.3$).

N	φ	Bound on φ	Max $ \sigma_i - \tilde{\sigma}_i $	Bound on max $ \sigma_i - \tilde{\sigma}_i $
8	6.9°	58°	0.25	0.30
16	6.7°	61°	0.26	0.30
24	6.7°	61°	0.26	0.30
32	6.7°	62°	0.26	0.30

Table 3. Comparison between \mathcal{N} , the nullspace of A ($a \equiv 0$), and $\tilde{\mathcal{N}}_2$, the nullspace of \tilde{A}_1 ($a \equiv 0.3$ everywhere except for a hole with $a \equiv 0$).

N	φ	Bound on φ	Max $ \sigma_i - \tilde{\sigma}_i $	Bound on max $ \sigma_i - \tilde{\sigma}_i $
8	6.2°	45°	0.19	0.25
16	6.0°	47°	0.19	0.26
24	5.9°	47°	0.19	0.26
32	5.9°	48°	0.19	0.26

8. Semi-discrete model problem with constant positive attenuation

In this and the following section, we calculate the nullspaces of model dose operators with constant positive attenuation analogous to those analysed in sections 2 and 3.

Our notation here is as in section 2, with the following modifications. We now assume Ω to be a convex and bounded open domain. We no longer rule out $\omega_\nu = -\omega_\mu$, i.e. beams in opposite directions. The dose operator A is now defined by

$$Ag(x) = \sum_{\nu=1}^n g_\nu(\omega_\nu^\perp \cdot x) \exp(-ad_\nu(x)) \tag{29}$$

for $x \in \Omega$, where

$$d_v(x) = \min\{s > 0 : x - s\omega_v \notin \Omega\}$$

and $a > 0$ is a constant. We re-write equation (29) as

$$Ag(x) = \sum_{v=1}^n g_v(\omega_v^\perp \cdot x) \exp(a(-d_v(x) + \omega_v \cdot x)) \exp(-a\omega_v \cdot x).$$

Because Ω is convex, $-d_v(x) + \omega_v \cdot x$ depends on $\omega_v^\perp \cdot x$ only:

$$-d_v(x) + \omega_v \cdot x = q_v(\omega_v^\perp \cdot x)$$

for some function q_v depending on Ω . Therefore

$$Ag(x) = \sum_{v=1}^n \widehat{g}_v(\omega_v^\perp \cdot x) \exp(-a\omega_v \cdot x) \tag{30}$$

with

$$\widehat{g}_v(p) = g_v(p) \exp(-aq_v(p)).$$

We write

$$\widehat{\mathbf{g}} = (\widehat{g}_1, \dots, \widehat{g}_n),$$

define

$$\widehat{A}\widehat{\mathbf{g}}(x) = \sum_{v=1}^n \widehat{g}_v(\omega_v^\perp \cdot x) \exp(-a\omega_v \cdot x), \tag{31}$$

and will determine

$$\widehat{\mathcal{N}} = \{\widehat{\mathbf{g}} : \widehat{A}\widehat{\mathbf{g}}(x) = 0 \text{ for all } x \in \Omega\}.$$

Equations (30) and (31) show how the nullspace \mathcal{N} of A is obtained from $\widehat{\mathcal{N}}$; in particular, \mathcal{N} has the same dimension as $\widehat{\mathcal{N}}$.

Our reasoning will closely follow that in section 2. First, we apply to the equation $\widehat{A}\widehat{\mathbf{g}} \equiv 0$ the operators $a + \omega_\mu \cdot \nabla$, $1 \leq \mu \leq n$, $\mu \neq v$, to obtain the general form of \widehat{g}_v ; see equation (34). (The analogue in section 2 was the result that \widehat{g}_v must be a polynomial of degree $\leq n - 2$.) Second, we insert (34) back into $\widehat{A}\widehat{\mathbf{g}} \equiv 0$ to determine which $\widehat{\mathbf{g}}$ of the form of (34) actually belong to $\widehat{\mathcal{N}}$.

We begin with the formula

$$\begin{aligned} \omega_\mu \cdot \nabla [\widehat{g}_v(\omega_v^\perp \cdot x) \exp(-a\omega_v \cdot x)] \\ = [-(\omega_\mu \cdot \omega_v) a \widehat{g}_v(\omega_v^\perp \cdot x) + (\omega_\mu \cdot \omega_v^\perp) \widehat{g}'_v(\omega_v^\perp \cdot x)] \exp(-a\omega_v \cdot x). \end{aligned}$$

Applying the operators $a + \omega_\mu \cdot \nabla$, $1 \leq \mu \leq n$, $\mu \neq v$, to $\widehat{A}\widehat{\mathbf{g}}(x) = 0$, we therefore find

$$\prod_{\mu \neq v} \left[a - (\omega_\mu \cdot \omega_v) a + (\omega_\mu \cdot \omega_v^\perp) \frac{\partial}{\partial p} \right] \widehat{g}_v(p) = 0,$$

or equivalently

$$\overset{\circ}{\prod}_{\mu} \left[a \lambda_{\mu,v} + \frac{\partial}{\partial p} \right] \widehat{g}_v(p) = 0, \tag{32}$$

where the circle above the product symbol indicates multiplication over all indices μ with

$$(\mu, v) \in J = \{(\mu, v) : 1 \leq v \leq n, 1 \leq \mu \leq n, \omega_\mu \neq \pm\omega_v\}, \tag{33}$$

and

$$\lambda_{\mu,v} = \frac{1 - \omega_\mu \cdot \omega_v}{\omega_\mu \cdot \omega_v^\perp}.$$

We will need the following straightforward properties of the $\lambda_{\mu,v}$.

Lemma 1. Let $(\mu, \nu) \in J$ and $(\tilde{\mu}, \tilde{\nu}) \in J$. Then

(a) $\lambda_{\mu, \nu} = -\lambda_{\nu, \mu}$,

(b) $\lambda_{\mu, \nu} \neq 0$,

(c) $\lambda_{\mu, \nu} = \text{sign}(\omega_\mu \cdot \omega_\nu^\perp) \sqrt{\frac{1 - \omega_\mu \cdot \omega_\nu}{1 + \omega_\mu \cdot \omega_\nu}}$,

(d) $\lambda_{\mu, \nu} = 1$ if and only if $\omega_\mu = \omega_\nu^\perp$, and

(e) $\lambda_{\mu, \nu} = \lambda_{\tilde{\mu}, \tilde{\nu}}$ if and only if (ω_μ, ω_ν) and $(\omega_{\tilde{\mu}}, \omega_{\tilde{\nu}})$ are equal up to (orientation-preserving) rotation.

Proof. (a) and (b) are obvious. To prove (c), we note that

$$(\omega_\mu \cdot \omega_\nu)^2 + (\omega_\mu \cdot \omega_\nu^\perp)^2 = 1.$$

This together with the definition of $\lambda_{\mu, \nu}$ implies (c), which in turn implies (d) and (e). □

From part (e) of lemma 1, $\lambda_{\mu, \nu} \neq \lambda_{\tilde{\mu}, \tilde{\nu}}$ if $(\mu, \nu) \in J$, $(\tilde{\mu}, \tilde{\nu}) \in J$ and $\mu \neq \tilde{\mu}$. Therefore, the functions

$$\exp(-a\lambda_{\mu, \nu} p)$$

with $1 \leq \mu \leq n$ and $(\mu, \nu) \in J$ form a basis of the solution space of the linear differential equation (32). Thus

$$\widehat{g}_\nu(p) = \sum_{\mu}^{\circ} c_{\mu\nu} \exp(-a\lambda_{\mu, \nu} p) \tag{34}$$

for constants $c_{\mu\nu}$, where the circle above the summation symbol indicates summation over all indices μ with $(\mu, \nu) \in J$. Inserting equation (34) into equation (31), we obtain

$$\widehat{A}\widehat{g}(x) = \sum_{\nu=1}^n \sum_{\mu}^{\circ} c_{\mu\nu} \exp(-a(\omega_\nu + \lambda_{\mu, \nu}\omega_\nu^\perp) \cdot x). \tag{35}$$

We next define

$$E = \{\xi \in \mathbb{R}^2 : \xi = \omega_\nu + \lambda_{\mu, \nu}\omega_\nu^\perp \text{ for some } (\mu, \nu) \in J\},$$

and, for $\xi \in E$,

$$J_\xi = \{(\mu, \nu) \in J : \xi = \omega_\nu + \lambda_{\mu, \nu}\omega_\nu^\perp\}.$$

Equation (35) can then be written as

$$\widehat{A}\widehat{g}(x) = \sum_{\xi \in E} \left(\sum_{(\mu, \nu) \in J_\xi} c_{\mu\nu} \right) \exp(-a\xi \cdot x).$$

Since exponentials corresponding to different elements $\xi \in E$ are linearly independent, we obtain exactly one constraint for each $\xi \in E$, namely

$$\sum_{(\mu, \nu) \in J_\xi} c_{\mu\nu} = 0.$$

These constraints are linearly independent because the J_ξ are disjoint.

Next, we determine the J_ξ . To do this, we must ask which pairs $(\mu, \nu) \in J$ and $(\tilde{\mu}, \tilde{\nu}) \in J$ satisfy

$$\omega_\nu + \lambda_{\mu, \nu}\omega_\nu^\perp = \omega_{\tilde{\nu}} + \lambda_{\tilde{\mu}, \tilde{\nu}}\omega_{\tilde{\nu}}^\perp. \tag{36}$$

Using the Pythagorean theorem, equation (36) implies $\lambda_{\mu, \nu} = \pm\lambda_{\tilde{\mu}, \tilde{\nu}}$. First, suppose

$$\lambda_{\tilde{\mu}, \tilde{\nu}} = \lambda_{\mu, \nu} = \lambda,$$

so $\omega_\nu + \lambda\omega_\nu^\perp = \omega_{\tilde{\nu}} + \lambda\omega_{\tilde{\nu}}^\perp$. This implies $\omega_\nu = \omega_{\tilde{\nu}}$, and therefore also $\omega_\mu = \omega_{\tilde{\mu}}$ by part (e) of lemma 1, so

$$\mu = \tilde{\mu} \quad \text{and} \quad \nu = \tilde{\nu}.$$

Second, suppose

$$-\lambda_{\tilde{\mu},\tilde{\nu}} = \lambda_{\mu,\nu} = \lambda, \tag{37}$$

so

$$\omega_\nu + \lambda\omega_\nu^\perp = \omega_{\tilde{\nu}} - \lambda\omega_{\tilde{\nu}}^\perp. \tag{38}$$

Then $\omega_\nu \neq -\omega_{\tilde{\nu}}$ evidently, and $\omega_\nu \neq \omega_{\tilde{\nu}}$ because $\lambda \neq 0$ by part (b) of lemma 1. Since $\omega_\nu \neq \pm\omega_{\tilde{\nu}}$, the unit vectors ω_ν and $\omega_{\tilde{\nu}}$ are linearly independent, and equation (38) is equivalent to the pair of equations obtained by taking its dot product with ω_ν and $\omega_{\tilde{\nu}}$. Both of these equations are equivalent to

$$\lambda = \lambda_{\tilde{\nu},\nu}. \tag{39}$$

Combining equations (37) and (39), we find $\lambda_{\mu,\nu} = \lambda_{\tilde{\nu},\nu}$ and $\lambda_{\tilde{\mu},\tilde{\nu}} = -\lambda_{\tilde{\nu},\nu}$. By parts (a) and (e) of lemma 1, this means

$$\mu = \tilde{\nu} \quad \text{and} \quad \nu = \tilde{\mu}.$$

Thus, J_ξ has exactly two elements; if one is (μ, ν) , the other is (ν, μ) . In summary, we have proved the following theorem.

Theorem 7. *For the semi-discrete model problem with constant positive attenuation, the null profiles are given by equation (34), $1 \leq \nu \leq n$, where the $c_{\mu\nu}$ are real coefficients with*

$$c_{\mu\nu} = -c_{\nu\mu}$$

for all $(\mu, \nu) \in J$. (The set J is defined in equation (33).)

9. Fully discrete model problem with constant positive attenuation

We repeat the analysis of section 3 for constant positive attenuation. Our notation is as in section 3, but we drop the assumption that there are no parallel beams in opposite directions:

$$\theta_\nu = (\nu - 1)\pi/4 \quad \text{for } 1 \leq \nu \leq 8.$$

The definition of \widehat{A} becomes

$$\widehat{A}\widehat{g}(\mathbf{x}) = \sum_{\nu=1}^8 \widehat{g}_\nu(\omega_\nu^\perp \cdot \mathbf{x}) \exp(-a\omega_\nu \cdot \mathbf{x}).$$

Assume that $\widehat{A}\widehat{g}(\mathbf{x}) = 0$ for all $\mathbf{x} \in \Gamma$. As before, we write $\mathbf{x}_{k,l} = (kh, lh)$ for $(k, l) \in \mathbb{Z} \times \mathbb{Z}$, and

$$D_{k,l} = \widehat{A}\widehat{g}(\mathbf{x}_{k,l}). \tag{40}$$

We define ∂_ν , $1 \leq \nu \leq 8$, to be the upstream first-order difference discretizations of the directional derivatives $\omega_\nu \cdot \nabla$. (See equations (7)–(10) for the explicit formulae for $1 \leq \nu \leq 4$.) We further define

$$L_\nu = \partial_\nu + \begin{cases} (e^{ah} - 1)/h & \text{if } \nu \text{ is odd} \\ (e^{\sqrt{2}ah} - 1)/(\sqrt{2}h) & \text{if } \nu \text{ is even} \end{cases}$$

for $1 \leq \nu \leq 8$. If we apply the finite-difference operator

$$L^{(\mu)} = \prod_{\nu \neq \mu} L_\nu$$

to the equation $D_{k,l} = 0$, all terms in the sum in equation (40) except for the one with $\nu = \mu$ drop out, and we obtain a finite-difference equation

$$\sum_{\alpha_\mu \leq j \leq \beta_\mu} w_{\mu,j} \widehat{g}_\mu(p + jh_\mu) = 0 \tag{41}$$

for $p \in \Gamma_\mu$, with integers α_μ and β_μ , $\alpha_\mu < \beta_\mu$, and

$$w_{\mu,\alpha_\mu} \neq 0 \quad \text{and} \quad w_{\mu,\beta_\mu} \neq 0.$$

It is easy to check that

$$\alpha_\mu = -3 \quad \text{and} \quad \beta_\mu = 3 \quad \text{if } \mu \text{ is odd}$$

and

$$\alpha_\mu = -4 \quad \text{and} \quad \beta_\mu = 4 \quad \text{if } \mu \text{ is even.}$$

Therefore the solution space of equation (41) is six-dimensional for odd μ and eight-dimensional for even μ . However, the six functions

$$\widehat{g}_\nu(p) = \exp(-a\lambda_{\mu,\nu} p)$$

with $(\mu, \nu) \in J$ are independent elements of the solution space. They therefore span it for odd μ . For even μ , it is easy to construct two further independent solutions:

$$\widehat{g}(p) = \begin{cases} \exp(ap) & \text{if } p/h_\nu \text{ even,} \\ -\exp(ap) & \text{if } p/h_\nu \text{ odd} \end{cases}$$

and

$$\widehat{g}(p) = \begin{cases} \exp(-ap) & \text{if } p/h_\nu \text{ even} \\ -\exp(-ap) & \text{if } p/h_\nu \text{ odd.} \end{cases}$$

This proves that null profiles have the form

$$\widehat{g}_\nu(p) = \sum_{\mu} c_{\mu\nu} \exp(-a\lambda_{\mu,\nu} p) + \begin{cases} (-1)^{p/h_\nu} (d_{\nu,+} \exp(ap) + d_{\nu,-} \exp(-ap)) & \text{if } \nu \text{ is even} \\ 0 & \text{if } \nu \text{ is odd} \end{cases} \tag{42}$$

$p \in \Gamma_\nu$, $1 \leq \nu \leq 8$. By inserting (42) into $\widehat{A}\widehat{g}(x_{k,l}) = 0$, as in previous sections, we easily obtain the conditions on the coefficients $c_{\mu\nu}$, $d_{\nu,+}$ and $d_{\nu,-}$ that make (42) a null profile. We summarize these results in the following theorem.

Theorem 8. *For the fully discrete model problem with constant positive attenuation, the null profiles are given by equation (42), where $c_{\mu\nu}$ are real coefficients with*

$$c_{\mu\nu} = -c_{\nu\mu}$$

for all $(\mu, \nu) \in J$, and $d_{\nu,+}$ and $d_{\nu,-}$ are real coefficients with

$$d_{2,-} = d_{4,+} \quad d_{4,-} = d_{6,+} \quad d_{6,-} = d_{8,+} \quad d_{8,-} = d_{2,+}.$$

The dimension of $\widehat{\mathcal{N}}$ is 28.

10. An optimization experiment

In this section, we give the details of the optimization experiment referred to in the introduction. Four beams are aimed at

$$\Omega = (0, 1)^2$$

with the aim of depositing a high dose in the disk

$$T = \{(x, y) \in \Omega : [(x - 0.5)^2 + (y - 0.5)^2]^{1/2} < 0.25\}$$

and a low dose everywhere else; see figure 1. We use the fully discrete model of dose deposition without beam attenuation defined in section 3. The grids used are given by equations (18)–(20), with $N = 25$, i.e. $h = 1/25$. We define the ‘ideal dose distribution’

$$D_i(x, y) = \begin{cases} 1 & \text{if } (x, y) \in T \\ 0 & \text{otherwise.} \end{cases}$$

Our aim is to determine a treatment plan that results in a dose distribution D minimizing

$$F(D) = \|D - D_i\|_{L^2}^2$$

where $\|\cdot\|_{L^2}$ denotes the discrete L^2 -norm, i.e. the trapezoidal approximation of the continuous L^2 -norm on the square $(0, 1)^2$. We denote by \mathcal{D} the set of all dose distributions D that are images under the dose operator A of treatment plans with non-negative intensities. \mathcal{D} is a convex set and F is a strictly convex function of D . Therefore, there exists a unique $D = D_{\text{opt}} \in \mathcal{D}$ that minimizes F . In this example, it is very easy to prove that the treatment plan g_{opt} with non-negative intensities such that $A g_{\text{opt}} = D_{\text{opt}}$ is unique as well. The optimal beam profiles, computed using the projected steepest-descent method, are shown in figure 2, and the optimal dose distribution D_{opt} is shown in figure 3. Notice that the unique optimal treatment plan g_{opt} contains oscillations on the scale of the grid. These oscillations can be removed by locally averaging the intensities of beams 2 and 4 (compare figure 1), for instance replacing $g_\nu(p)$, $p \in \Gamma_\nu$, $\nu \in \{2, 4\}$, by

$$\bar{g}_\nu(p) = (2g_\nu(p) + g_\nu(p - h_\nu) + g_\nu(p + h_\nu))/4.$$

(Although there is no compelling reason to use precisely this local averaging scheme, it does have the desirable property of completely removing zig-zag oscillations.) One obtains a smoothed treatment plan \bar{g}_{opt} (figure 4) generating a smoothed dose distribution \bar{D}_{opt} (figure 5) that is only very slightly suboptimal:

$$\|D_{\text{opt}} - D_i\|_{L^2} = 0.2672 \quad \text{and} \quad \|\bar{D}_{\text{opt}} - D_i\|_{L^2} = 0.2678.$$

The calculations were carried out in double precision arithmetic; we are confident that the digits given are accurate.

The oscillations shown in figures 2 and 4 stem from the fact that the discretizations of the beam and of the region Ω are aligned in a very special way. In realistic problems, such alignments could occur accidentally and give rise to similar oscillations in the optimal beam profiles. However, our point is more general. Certain substantial changes in beam profiles lead to small changes in dose distributions only, and this may offer an opportunity to compute substantially simplified yet still nearly optimal beam profiles.

We have neglected scattering of radiation in this paper, although it is crucially important in realistic problems. However, we believe that in fact scattering strengthens our point, since it dampens the effect of highly oscillatory perturbations in the beam profiles, and therefore little damage is done if highly oscillatory components are removed from optimal beam profiles.

We note that in [4], an algorithm was proposed that requires knowledge of the null profiles. This idea was explored further in [8, 10, 15]. However, our study of the null profiles has not been motivated by algorithmic ideas, but simply by the goal of more clearly understanding the nature of the optimization problem. The explicit computation of approximate null profiles is likely to be much too expensive to be practical for realistic problems. Even if approximate null profiles were known explicitly for a realistic problem, they would be useful only if they could be added to a computed near-optimal or optimal profile without introducing any negative intensities. For these reasons, it is probably best not to use approximate null profiles for plan simplification directly, but to modify the objective function in such a way that it rewards simplicity. We intend to explore this in future work.

Acknowledgments

We thank James Satterthwaite (ADAC Laboratories and Tufts Medical School) for many helpful discussions, and Michael Goitein and Andrzej Niemierko (Harvard Medical School) for a clarifying conversation at the beginning of this work. C Börgers was partially supported by NSF grant DMS-9626696. E T Quinto was partially supported by NSF grant DMS-9622947.

References

- [1] Arfken G 1985 *Mathematical Methods for Physicists* (New York: Academic)
- [2] Börgers C 1999 The radiation therapy planning problem *Computational Radiology and Imaging* ed C Börgers and F Natterer (*IMA Volumes in Mathematics and its Applications vol 110*) (New York: Springer)
- [3] Börgers C, Quinto E T and Satterthwaite J 1999 Small singular values of the mapping from beam intensities to dose distributions in radiation therapy planning, in preparation
- [4] Cormack A M and Quinto E T 1990 The mathematics and physics of radiation dose planning *Contemp. Math.* **113** 41–55
- [5] Davison M E and Grünbaum F A 1981 Tomographic reconstruction with arbitrary directions *Commun. Pure Appl. Math.* **34** 77–119
- [6] Deasy J O 1997 Multiple local minima in radiotherapy optimization problems with dose-volume constraints *Med. Phys.* **24** 1157–61
- [7] Demmel J W 1997 *Applied Numerical Linear Algebra* (Philadelphia, PA: SIAM)
- [8] Gregerson G 1992 The null space and singular value decomposition of the spherical radon transform as a method in radiation dose planning *MSc Thesis* Tufts University
- [9] Jackson A, Wang X H and Mohan R 1994 Optimization of conformal treatment planning and quadratic dose objectives (abstract) *Med. Phys.* **21** 1006
- [10] Lissiano S 1996 Some mathematical problems of tomography and radiation treatment *PhD Thesis* Wichita State University
- [11] Ludwig D 1966 The Radon transform on Euclidean space *Commun. Pure Appl. Math.* **19** 49–81
- [12] Madych W R and Nelson S A 1984 Polynomial based algorithms for computed tomography II *SIAM J. Appl. Math.* **44** 193–208
- [13] Natterer F 1986 *The Mathematics of Computerized Tomography* (New York: Wiley)
- [14] Niemierko A 1996 Optimization of intensity modulated beams: local or global optimum? (abstract) *Med. Phys.* **23** 1072
- [15] Ponomaryov I 1997 Numerical analysis of problems of tomography, radiotherapy and photonic crystal theory *PhD Thesis* Wichita State University
- [16] Stewart G W and Sun J-G 1990 *Matrix Perturbation Analysis* (New York: Academic)