

Iterative Prescription Refinement in Fully-Discretized Inverse Problems of Radiation Therapy Planning

Yair Censor¹ and Lei Xing²

¹Department of Mathematics, University of Haifa
Mt. Carmel, Haifa 31905, Israel

²Department of Radiation Oncology
Stanford University
875 Blake Wilbur Drive
Stanford, CA 94305-5847, USA

November 25, 2010. Revised: April 24, 2011.

Abstract

We formulate the method of *Iterative Prescription Refinement* (IPR) for inverse planning in any fully-discretized model of radiation therapy. The method starts out from an *ideal dose prescription* and repeatedly refines it into a *refined dose prescription*. This is done computationally without human interaction until a prespecified stopping rule is met, at which point the refined dose vector and the accompanying beamlet intensities vector are evaluated and presented to the planner. The algorithmic regime is general enough to encompass various physical models that may use different particles (photons, protons, etc.) It is formulated for a general *inversion operator* thus different objective functions or approaches to the optimization problem (such as DVH, gEUD, or TCP and NTCP cost functions) may all be applied. Although not limited to this model, we demonstrate that the approach at all works on two exemplary cases from photon intensity-modulated radiation therapy (IMRT).

1 Introduction and Problem Definition

Radiation therapy treatment planning (RTTP) is basically an *inverse problem* but of a nature that sets it apart from other inverse problems in science and technology. Many inverse problems are *data-inversion* problems. By this we mean that there is a relation of the form $Qx = b$ where b is a vector of *measured data*, which cannot be changed once the measurements have been performed, the vector x is the object that needs to be recovered, and Q is an operator that represents our knowledge of the physics of the situation that relates the measurements to the object that was measured. There are some fundamental deviations from this in the inverse problem of RTTP. First, in RTTP the right-hand side b is not fixed throughout like in a data-inversion problem but is a prescription vector that might be modified if the planner does not like to accept, for some reason, the recovered object x . This property justifies us in dubbing the inverse problem of RTTP as a *prescription-inversion* problem. Secondly, the operator Q cannot be inverted mathematically, without major compromise on its realism, therefore, a fully-discretized approach must be taken and not a continuous model of the physical situation. For a discussion of this matter, see, e.g., [3, 4]. General review and tutorial papers on RTTP from various aspects are abound. Some of the more mathematically algorithmically oriented reviews are [19, 17, 9, 2].

There is yet another important aspect to the prescription-inversion problem in RTTP that sets it apart from conventional data-inversion inverse problems. The *inversion method* itself requires *information-input* from the planner, which will, in turn, affect the resulting recovered object x . Let us agree, for the purpose of this paper, to use the term “inversion method” as an overall term that consists of the union of the following ingredients: (1) the chosen mathematical model, (2) the chosen objective (cost) function (if an optimization model is used), (3) the choice of algorithm employed, and (4) the specification of all parameters necessary to run the algorithm on the chosen cost function for the chosen model. Now, if the “inversion method” is applied to some prescribed right-hand side vector b and if the resulting recovered object x (e.g., beamlets intensities in IMRT) does not satisfy the planners goals then the planner is confronted with a dilemma. He can either “revise” his original prescribed right-hand side vector b or insist on that b as originally prescribed. In the latter case some changes must be made to the information-input for the “inversion method” itself before running it again, because otherwise (i.e., same b and no changes to the “inversion method”)

there will be no new alternative recovered object x obtained for the planner's assessment.

Just to make the above more concrete, think – by way of example only – about an inversion method in which a weighted sum of several cost functions, associated with specific organs, has to be minimized via some optimization algorithm or package. An example of changing the information-input to the “inversion method” might be changes to the coefficients (weights of importance) given in the weighted sum cost function. The difficulty is manifold: It is not simple or clear how such changes should be made, they might be done repeatedly several times, and, most importantly, they require the planner's human involvement.

Our goal in this paper is propose a methodology to handle the above situation. The proposed approach and procedure are not limited to IMRT or to the specific objective function used in the examples, but is applicable to ANY fully-discretized model of prescription-inversion. The mathematical background and connection with another method from medical imaging, offered here, puts the whole approach on more secure foundations (in spite of the admitted, still remaining, questions even in the math of the procedure). The case examples are presented to only demonstrate that the approach and method “work” at all. More efforts might be needed in any specific area where this will be employed to fine-tune the method to reach computational advantages. To these ends we formulate a method, that we call *Iterative Prescription Refinement* (IPR), for inverse planning in any model of radiation therapy treatment planning that is fully-discretized. IPR is not limited to IMRT or to a specific choice of objective functions for optimization. It is a general principle that is applicable to any fully-discretized model of an inversion problem that might occur in radiation therapy treatment planning (RTTP), or even in prescription-inversion problems outside RTTP. It puts forward and identifies an approach, not considered hitherto, that replaces the currently commonly used sequence of steps in RTTP by a different one which attacks the problem from a different angle. The method should start out from an *ideal dose prescription* (ideal in the sense that it is most probably not deliverable) and repeatedly refines this prescription into a *refined dose prescription*. This is done computationally without human interaction until a pre-specified stopping rule is met, at which point the refined dose vector is ready to be evaluated by the planner. The purpose is to circumvent in this way the inevitable “imperfections” of the inversion operator that is employed in the solution process of the inverse problem. This will be further clarified

below.

The algorithmic regime is general enough to encompass various physical models that may use different particles (photons, protons, etc.) and/or model the problem mathematically in one of several possible different ways.

1.1 The fully-discretized model

We consider the following fully-discretized model in RTTP which has its early roots in [3]. There are J voxels in total in the cross-section. A dose vector is $d = (d_j)_{j=1}^J$ where d_j is dose in voxel j . (observe the “notational rule” to use superscripts to count vectors and subscript to count vector components). There are N recognized organs (counting both targets and organs at risk (OARs)) in the cross-section, denoted by their voxel sets O_1, O_2, \dots, O_N . A dose vector d is partitioned accordingly into subvectors

$$d = \begin{pmatrix} d^1 \\ d^2 \\ \vdots \\ d^N \end{pmatrix} \quad (1)$$

where $d^i = (d_j)_{j \in O_i}$ is a subvector of d that contains the doses in only the voxels of organ O_i .

There are T external radiation beamlets, counted by $t = 1, 2, \dots, T$, in total, and $x = (x_t)_{t=1}^T$ is a vector of beamlet intensities such that x_t is the intensity of beamlet t . We partition the $J \times T$ dose matrix A accordingly, so that,

$$Ax = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_N \end{pmatrix} x = \begin{pmatrix} d^1 \\ d^2 \\ \vdots \\ d^N \end{pmatrix}, \quad (2)$$

namely, $A_i x = d^i$ for all $i = 1, 2, \dots, N$. The dimensionality of the submatrix A_i is $|O_i| \times T$ where $|O_i|$ is the cardinality of O_i , i.e., the number of voxels in organ O_i . The, now classical, see, e.g., [4], inverse problem of radiation therapy treatment planning is to find an intensities vector x that will deliver to the patient’s cross-section a prescribed dose vector d , for a given patient data and for a given particle model and available treatment machine setup.

1.2 The Problem

We assume that \mathcal{R} is an operator that represents some given “inversion method” as an overall term that consists of the union of the following ingredients: (1) the chosen mathematical model, (2) the chosen objective (cost) function (if an optimization model is used), (3) the choice of algorithm employed, and (4) the specification of all parameters necessary to run the algorithm on the chosen cost function for the chosen model, for the inverse problem of a fully-discretized radiation therapy treatment planning problem, such as, but not necessarily only, photon IMRT. By this we mean that $\mathcal{R} : R^J \rightarrow R^T$ (i.e., it maps the J -dimensional Euclidean space R^J into R^T). This operator takes a dose vector $d \in R^J$ and outputs an intensity vector $x = \mathcal{R}(d)$ that supposedly will create the dose distribution d when applied in the clinic to the treatment case whose parameters were used in the inversion with \mathcal{R} . Therefore, we call \mathcal{R} an *inversion operator*.

We further assume that $\mathcal{D} : R^T \rightarrow R^J$ is a “state-of-the-art” forward dose calculation package that is available to us. It takes any intensity vector $x \in R^T$ and outputs a dose vector $d = \mathcal{D}(x)$ that supposedly will be created by that x , when applied in the clinic to the treatment case whose parameters were used in the inversion with \mathcal{R} . Therefore, we call \mathcal{D} a *dose operator*. The difficulties associated with both the forward and inverse problems in IMRT stem from the fact that to this date there exists no, realistically adequate, closed-form analytic representation of the dose operator \mathcal{D} . Although the interaction between radiation and tissue is measured and understood at the atomic level, the situation is so complex that, to solve the forward problem in practice, a state-of-the-art computer program (i.e., a sufficiently accurate dose calculation engine), which represents a *computational approximation* of the operator \mathcal{D} , must be used.

Denote by d^{pres} a vector of prescribed doses to voxels, for a given treatment case. If \mathcal{R} and \mathcal{D} were perfect (ideal), i.e., representing the model perfectly, applying the inversion method without any errors (numerical or others), and performing a perfect forward dose calculation, respectively, then the composition of the ideal inversion operator \mathcal{R} and the ideal dose operator \mathcal{D} should return the input, i.e.,

$$\mathcal{D}\mathcal{R}(d^{pres}) = d^{pres} . \quad (3)$$

In this case $\tilde{x} = \mathcal{R}(d^{pres})$ would be an intensity vector that will deposit precisely d^{pres} as prescribed, see Box “A” in Figure 1. Let us call such \mathcal{R}

and \mathcal{D} *ideal operators* and note that if (3) holds for all d^{pres} then $\mathcal{D}\mathcal{R} = \mathcal{I}$, where \mathcal{I} stands for the *identity operator* that leaves any vector to which it is applied unchanged.

In the real world we have at our disposal only approximations of the ideal operators \mathcal{R} and \mathcal{D} , let's denote them by $\hat{\mathcal{R}}$ and $\hat{\mathcal{D}}$ and call them *actual inversion* and *actual dose* operators, respectively. So, in the real world, applying the composition of operators $\hat{\mathcal{D}}\hat{\mathcal{R}}$ to a dose vector does not return the vector to which it is applied, namely,

$$\hat{\mathcal{D}}\hat{\mathcal{R}}(d^{pres}) = \hat{d} , \quad (4)$$

meaning that the dose vector \hat{d} deposited by the intensity vector $\hat{x} = \hat{\mathcal{R}}(d^{pres})$ will differ from the prescribed dose vector d^{pres} .

It is our purpose in this study to address only imperfections of the actual inversion operator $\hat{\mathcal{R}}$ (and not of the actual dose operator $\hat{\mathcal{D}}$). Therefore, we assume from this point onward that $\hat{\mathcal{D}} = \mathcal{D}$ is an ideal dose operator which is also actual, thus available to us.

1.3 The usual approach

The usual approach to cope with the discrepancy between the prescribed dose d^{pres} and the actually deliverable dose \hat{d} , obtained from (4), consists basically of the following actions performed in the given order: (1) perform the inverse planning with the actual inversion operator $\hat{\mathcal{R}}$ (according to some given model and solution method), (2) apply the dose operator and show the resulting dose vector and its various derivatives (e.g., isodose map, dose-volume histogram (DVH), etc.) to the planner, (3) if the results do not satisfy the planner's clinical goals, change parameters in the actual inversion operator $\hat{\mathcal{R}}$ and repeat the process (i.e., go back to (1) above). For example, if $\hat{\mathcal{R}}$ represents a minimization of a linear combination of some individual cost functions for various organs, then a commonly used "change of parameters" is to change the coefficients in the linear combination of individual cost functions that is being minimized. These coefficients reflect the relative importance attributed by the planner to minimizing the individual cost functions, each of which is usually related to one organ (target or OAR). Then the, so revised, actual inversion operator $\hat{\mathcal{R}}$ is applied again and the new resulting actually deliverable dose \hat{d} is evaluated for the case at hand, say, via observation of the dose-volume histograms. This process demands continuous and repeated involvement of the treatment planner and it is stopped

only when the latter accepts a clinically acceptable treatment plan presented to him, even if it sometimes deviates from the original prescription that he gave initially.

The main difficulty with this approach seems to lie not so much in the repetitious application of the actual inversion operator $\hat{\mathcal{R}}$. This repetitious application is a technical problem that can be addressed by hardware acceleration via more powerful computer and/or parallel computations such as on general purpose graphics processing units (GP-GPUs). The main difficulties with this approach are twofold: (i) the need to repeatedly and manually evaluate the resulting actual dose \hat{d} and, even more fundamentally, (ii) since the overall process is a “search process” and the options for changing parameters in the actual inversion operator $\hat{\mathcal{R}}$ are numerous, it is far from clear or simple how to change the parameters each time the resulting dose distribution \hat{d} does not yet meet the clinical requirements.

2 The Iterative Prescription Refinement (IPR) Method

To circumvent the difficulties mentioned above, we examine here, instead of the traditional approach, an iterative process by which the final refined dose \hat{d} is reached by iteratively refining and steering a given ideal dose prescription vector. Repeated applications of the operator $\hat{\mathcal{R}}$ are still used but instead of doing parameter changes inside $\hat{\mathcal{R}}$ that require manual intervention and are not always known how to perform, $\hat{\mathcal{R}}$ is used without changes and the method itself keeps reducing the discrepancy between the current refinement of the dose vector and the original ideal prescription.

2.1 The method

The method, which we call *Iterative Prescription Refinement* (IPR), iterates on dose vectors. The procedure is outlined in Figure 1. Box “A” depicts the not achievable situation of (3). If we had (i.e., if we knew and could implement) the ideal inversion operator \mathcal{R} we could have calculated the desired intensity vector \tilde{x} . This not being the case, we propose to resort to the IPR procedure depicted on the right-hand side of Figure 1. The aim is to generate a sequence of intensity vectors $\{\hat{x}^k\}_{k=0}^{\infty}$ such that the dose vectors deposited

by them $\hat{d}^k = \mathcal{D}(\hat{x}^k)$ get closer, or as close as possible, to d^{pres} . As we show below, this aim can be achieved by the IPR procedure.

No work needs to be done to implement what is in Box “B” because, by $\mathcal{DR} = \mathcal{I}$, its input and output are identical. The right-hand side within Box “C” uses the actual inversion operator $\hat{\mathcal{R}}$ to generate $\hat{x}^k = \hat{\mathcal{R}}(d^k)$, from which \hat{d}^k is obtained by $\hat{d}^k = \mathcal{D}(\hat{x}^k)$. The discrepancy vector $\Delta^k = d^k - \mu\hat{d}^k$ is then added to μd^{pres} , where μ is a user-chosen *relaxation parameter*, to generate the next refined dose vector d^{k+1} and so on, until some pre-defined stopping rule applies, see the algorithm below.

Algorithm 1 *Iterative Prescription Refinement (IPR)*

Initialization: Choose d^0 , a very reasonable choice is $d^0 = d^{pres}$ and choose a relaxation parameter real number μ .

Iterative Step: Given the current prescription dose vector iterate d^k , $k = 0, 1, 2, \dots$,

(1) Calculate the k -th intensity vector \hat{x}^k by applying the actual inversion operator

$$\hat{x}^k = \hat{\mathcal{R}}(d^k), \quad (5)$$

(2) Calculate the dose vector \hat{d}^k that will be obtained from \hat{x}^k by applying the dose operator \mathcal{D}

$$\hat{d}^k = \mathcal{D}(\hat{x}^k), \quad (6)$$

(3) Calculate the dose discrepancy vector Δ^k

$$\Delta^k = d^k - \mu\hat{d}^k, \quad (7)$$

(5) Calculate the next prescription dose vector iterate d^{k+1} by

$$d^{k+1} = \mu d^{pres} + \Delta^k. \quad (8)$$

Stopping Rule: Stop the process by applying some stopping rule.

2.2 Analysis of the IPR procedure

The IPR method can be mathematically identified (with appropriate nonessential changes of notations dictated by the problem at hand here) with the *Iterative Data Refinement* (IDR) method of [5], see also [6, Section 10.5] and [11, pp. 90 and 99]. This method has been developed in medical imaging to cope with difficulties stemming from discrepancies between *idealized* and

actual measuring devices in models of image reconstruction from projections. It has been found useful for *beam hardening correction* in computerized tomography, for *attenuation correction* in single photon emission computed tomography (SPECT) and in *image reconstruction with incomplete data*, see [5, Section 2]. It was further used for *magnetic resonance imaging*, see [12, 18]. Recently it has been used for *transfer function restoration in 3D electron microscopy* by Sorzano et al. [20], and in *positron emission tomography (PET)* by Crepaldi and De Pierro [8].

Comparing IDR with the IPR outline proposed here the necessary adjustments to the IDR which yield the IPR and the identification of quantities can be observed as follows. The “idealized measuring device” and the “actual measuring device” that appear in the IDR scheme are of no use here because what is there the measurement vector y is here a prescription vector d that is not at all measured, but prescribed. IPR aims to overcome imperfections of the inversion operator (not of the measuring device, as in the IDR). The correspondence between quantities in each of the two procedures is given in the next table.

<i>Iterative Data Refinement</i> (IDR) notations in [5]	<i>Iterative Prescription Refinement</i> (IPR) notations in this paper
data vectors $y^{(k)}$	dose vectors d^k
relaxation parameter μ	relaxation parameter μ
actual \tilde{y} ideal \hat{y}	assumed $\tilde{y} = \hat{y} = d^{pres}$
the product operator AS	the identity operator \mathcal{I}
the product operator BS	the product operator $\mathcal{D}\tilde{R}$
$U = (A - \mu B)S$	$\mathcal{U} = \mathcal{I} - \mu\mathcal{D}\tilde{R}$

Exploiting this homomorphism enables us to “translate” some statements and thus shed light on the algorithmic behavior of IPR. This can be done in the following manner.

Definition 2 *Iterative Prescription Refinement (IPR) associates with every 4-tuple $(d^{pres}, d^0, \mathcal{U}, \mu)$ where $\mathcal{U} : R^T \rightarrow R^T$ is an operator that maps dose vectors onto dose vectors, μ is a real number, and d^{pres} and d^0 are vectors in R^T , a sequence $\{d^k\}_{k=0}^\infty$ generated by*

$$d^{k+1} = \mu d^{pres} + \mathcal{U}(d^k). \quad (9)$$

The quantities and notations identifications in the table show that IPR is indeed a special case of IDR, and, in particular, that our (9) coincides with (1) of [5]. Specifically, focusing on the IPR process depicted in Figure 1 and described by Algorithm 1, we actually use the operator

$$\mathcal{U} = \mathcal{I} - \mu \mathcal{D} \hat{\mathcal{R}} \quad (10)$$

thus, $\mathcal{U}(d^k) = \Delta^k$. Now that the connection between IPR and IDR is established, the results of the mathematical analysis of [5] are applicable. The interested reader is referred to [5] for this mathematical contents. In particular, Proposition 8 in [5] (consult also [6, Section 10.5]) translated to the IPR procedure, guarantees, generally speaking, that under the conditions set forth there, one of the following two possibilities must happen:

- (i) Asymptotic convergence of the sequence of refined dose vectors d^k to the prescribed dose vector d^{pres} , as k goes to infinity, i.e., $\lim_{k \rightarrow \infty} d^k = d^{pres}$, or
- (ii) There exists an index k_0 such that $d^0, d^1, d^2, \dots, d^{k_0}$ are all in a certain set (denoted G_β in [5]), and

$$\|d^{k_0+1} - d^{pres}\| < dist_\beta^{d^0}, \quad (11)$$

where $dist_\beta^{d^0}$ is the distance from d^{pres} of the point in $G_\beta^{d^0}$ which is nearest to d^{pres} . $G_\beta^{d^0}$ is the component of G_β which contains d^0 . For practical reasons (explained in [5, p. 116]) it seems that the convergence option (i) is unlikely and the most that we could hope for is (ii). The meaning is that, provided that $d^0 \in G_\beta$, improvement is guaranteed by (11) all along and up to a certain iteration index k_0 . Not having a constructive way of finding this index ahead of time, one must rely on experiments to demonstrate that this initial improvement actually occurs. The situation (ii) is not uncommon for ill-posed problems. The phenomenon is often referred to as *semi-convergence*, see, e.g., [10].

3 A Demonstration of IPR in Photon IMRT

To illustrate the IPR method, we have generated IMRT treatment plans for two clinical head and neck cases in the following manner. We emphasize that the case examples are presented to only demonstrate that the approach

and method “work” at all. More details about the experimental work that leads to the following can be found in [15]. Further efforts might be needed in any specific area where this will be employed to fine-tune the method to reach computational advantages. First, we have calculated the dose matrix A from a patient’s CT scan using CERR (Computational Environment for Radiotherapy Research) [7] and the VMC++ Monte Carlo package optimized for radiation treatment planning [21] in MATLAB on a PC with 2.66 Core 2 Duo CPU and 3.25 GB of RAM. To reduce computational time we have down sampled the original CT voxel size by a factor of 4 in the xy plane. No downsampling in the z direction was used. Next, we have obtained an optimal fluency map \hat{x} as a solution of the inversion problem $\hat{x} = \hat{\mathcal{R}}(d^{pres})$. This was done in MATLAB using MOSEK [1] optimization software [16]. The MOSEK Optimization Software is designed to solve large-scale mathematical optimization problems. It provides specialized solvers for linear problems, conic quadratic problems, convex quadratic problems, general convex problems, and mixed integer problems. No specific regularization is used in the MOSEK package. However, it may happen that some optimization problem with regularization can be reformulated as a quadratic programming problem (see, e.g., Appendix A of Zhu et al. [23]) or the problem mentioned above. The reformulated problem may then be efficiently solved using MOSEK. The resultant fluency map was converted into a 3D dose distribution using the dose matrix A . To characterize and compare different dose distributions we have calculated corresponding dose-volume histograms (DVHs) and isodose curves.

To demonstrate that the IPR method works we have generated two plans for each clinical case. The first plan is a standard plan, without iterative prescription refinement, with the prescription set to its ideal values – 100% to the target and 0% to organs at risk. The second plan is a realization of the proposed iterative prescription refinement method with the number of model iterations set to 5. The plans were compared using DVHs and 3D dose distributions plotted in MATLAB using CERR package.

The iterative calculation has two parts here: First, for a given set of prescription, the optimization finds the corresponding solution via an iterative process and this typically takes 20~30 iterations. Secondly, after the above iterative calculation is done, we modify the prescription in such a way that the dose inhomogeneity in the PTV and the high doses in the sensitive structures can be reduced. After the prescription is changed, the optimization described above (inner-loop optimization) is performed again. We found

that after 4~5 changes of the prescription, the final solution saturates and any further change of the prescription does not improve the quality of the treatment plan any more. Considering the inner-loop optimization, the total number of iteration for 5 sets of prescription is roughly 100~150. The computational time for 150 iterations on a PC with 3GHz CPU is roughly 10~15 min.

In this work, the beamlet kernels for unit beamlet intensity were pre-computed by using the Voxel-based Monte Carlo (VMC) algorithm (see Kawrakow [14]) after the treatment beam configuration was determined. The total dose at a point is a sum of all the contributing beamlets weighted by appropriate intensities resulted from the optimization.

3.1 Case I

For the first clinical case we used treatment geometry with 7 beams at 0, 40, 80, 90, 120, 180 and 270 degree angles, represented by a total of 2,384 beamlets. The total number of voxels in all structures including all targets is $256 \times 256 \times 224$. We compare treatment plans by plotting corresponding dose-volume histograms (DVHs). Our figure of merit is a plan with the fixed ideal prescription. In Figure 2 we compare DVHs for the fixed and iterative prescription refinement (after 5 iterations) plans for different structures. We see that, while the target coverage is comparable for both plans, there is a reduction in the delivered dose to the mandible and pharynx for the iterative prescription plan. To run a more detailed comparison of the plans we have plotted correspondent 3D dose distributions for a fixed slice of the patient's anatomy. The resultant color dose maps and organ outlines are depicted in Figures 3 and 4.

3.2 Case II

For the second case we chose a slightly different beam geometry with 7 beams at 30, 55, 95, 130, 150, 170 and 210 degrees, represented by a total of 3,490 beamlets. The total number of voxels in all structures is 169,050. As in the previous case we compare a treatment plan generated with the iterative prescription refinement method to a standard fixed prescription plan. Figure 5 shows the comparison for different structures. We observe that both plans are identical in the PTV dose coverage for both the primary PTV and the secondary PTV (50 Gy). PTV stands for Planning Target Volume, which

is defined to select appropriate beam sizes and beam arrangements, taking into consideration the net effect of all the possible geometrical variations and inaccuracies in order to ensure that the prescribed dose is actually absorbed in the Clinical Target Volume (CTV). Its size and shape depend on the CTV but also on the treatment technique used, to compensate for the effects of organ and patient movement, and inaccuracies in beam and patient setup, see [13]. The cord, the pharynx and the oral cavity enjoy a slight dose reduction in the case of the iterative prescription refinement plan.

4 Concluding Comments

The basic idea of *Iterative Prescription Refinement* (IPR) admittedly comes from the earlier works [15, 22] with the current formulation (1) giving this idea a new motivation, (2) putting it on sound theoretical ground, and (3) being more general than in those papers which were limited to IMRT. Any inverse problem, in RTTP or in another physics problem, that can be identified as a fully-discretized prescription-inversion problem can adopt and adapt the IPR methodology. Experimental work in IMRT that implemented the IPR methodology is reported in our recent [15] and in [22]. The methodology of IPR is not specific to IMRT or to any specific inverse planning method within IMRT. Any inverse planning situation modeled by full-discretization can benefit from IPR. In this paper we identified the algorithmic nature of IPR and validated it mathematically by showing that it is a special case of the general iterative data refinement (IDR) paradigm. The mathematics of IDR itself is limited, thus so is our current understanding of the foundation of IPR. Therefore IPR needs to be further studied experimentally and mathematically, raising a host of questions. Is the inverse problem at hand ill-posed, e.g., is the matrix A in (2) ill-conditioned (or is the underlying continuous operator compact in some reasonable topology?) It would be interesting to see a plot of the singular values of A even for a small model problem. Even for noise-free problems possible ill-posedness must be taken into account. Otherwise, the computed solution can be corrupted by the noise necessarily introduced in the computational procedure itself. How should a stopping criterion be constructed in the IPR procedure? How should the relaxation parameter μ be chosen?

Acknowledgments. We heartily thank Dr. Pavel Lougovski for his

help in all phases of the work on this paper. We thank three anonymous referees whose constructive comments helped us to improve an earlier version of this work. This work was partially supported by Award Number R01HL070472 from the National Heart, Lung, And Blood Institute at the National Institutes of Health (NIH), and by United States-Israel Binational Science Foundation (BSF) grant No. 2009012.

References

- [1] E.D. Andersen and K.D. Andersen, The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm, in: Frenk, H., Roos, K., Terlaky, T. and Zhang, S. (eds.), *High Performance Optimization*, pp. 197–232, Kluwer, Boston (2000).
- [2] T. Bortfeld, IMRT: a review and preview, *Physics in Medicine and Biology*, **51** (2006), R363.
- [3] Y. Censor, M.D. Altschuler and W.D. Powlis, On the use of Cimmino’s simultaneous projections method for computing a solution of the inverse problem in radiation therapy treatment planning, *Inverse Problems*, **4** (1988), 607–623.
- [4] Y. Censor and J. Unkelbach, From analytic inversion to contemporary IMRT optimization: Radiation therapy planning revisited from a mathematical perspective, *Physica Medica: European Journal of Medical Physics*, accepted for publication.
- [5] Y. Censor, T. Elfving, and G.T. Herman, A method of iterative data refinement and its applications, *Mathematical Methods in the Applied Sciences*, **7** (1985), 108–123.
- [6] Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, New York, NY, USA, 1997.
- [7] Computational Environment for Radiotherapy Research (CERR), a software platform for developing and sharing research results using radiation therapy treatment planning information, available at: <http://radium.wustl.edu/CERR/about.php>

- [8] F. Crepaldi and A.R. De Pierro, Activity and attenuation reconstruction for positron emission tomography using emission data only via maximum likelihood and iterative data refinement, *IEEE Transactions on Nuclear Science*, **54** (2007), 100–106.
- [9] M. Ehrgott, Ç. Güler, H.W. Hamacher and L. Shao, Mathematical optimization in intensity modulated radiation therapy, *4OR*, **6** (2008), 199–262.
- [10] T. Elfving, T. Nikazad, and P.C. Hansen, Semi-convergence and relaxation parameters for a class of SIRT algorithms, *Electronic Transactions on Numerical Analysis (ETNA)*, **37** (2010), 321–336.
- [11] G.T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, 2nd ed., Springer, 2009.
- [12] G.T. Herman and D.W. Ro, Image recovery using iterative data refinement with relaxation, *Optical Engineering*, **29** (1990), 513–523.
- [13] International Commission on Radiation Units (ICRU) Report 50 (1993): “Prescribing, Recording, and Reporting Photon Beam Therapy”. International Commission on Radiation Units (ICRU), Washington, DC, USA. See: http://www.icru.org/index.php?option=com_content&task=view&id=72.
- [14] I. Kawrakow, Improved modeling of multiple scattering in the voxel Monte Carlo model, *Medical Physics*, **24** (1997), 505–517.
- [15] P. Lougovski, J. LeNoach, L. Zhu, Y. Ma, Y. Censor and L. Xing, Toward truly optimal IMRT dose distribution: Inverse planning with voxel-specific penalty, *Technology in Cancer Research and Treatment*, **9** (2010), 629–636.
- [16] MOSEK ApS, Denmark, see: <http://www.mosek.com/>.
- [17] R. Reemtsen and M. Alber, Continuous optimization of beamlet intensities for intensity modulated photon and proton radiotherapy, in: Pardalos, P.M. and Romeijn, H.E. (eds.) *Handbook of Optimization in Medicine*, Springer Optimization and Its Applications, 2009, Volume **26**, pp. 1–40, Springer.

- [18] D.W. Ro, G.T. Herman and P.M. Joseph, Resolution enhancement of magnetic resonance images using an iterative data refinement technique, in: *Visual Communications and Image Processing IV*, W.A. Pearlman, (ed.), SPIE Proceedings Vol. 1199 (1989), pp. 952–962.
- [19] H.E. Romeijn and J.F. Dempsey, Intensity modulated radiation therapy treatment plan optimization, *TOP*, **16** (2008), 215–243.
- [20] C.O.S. Sorzano, R. Marabini, G.T. Herman, Y. Censor and J.M. Carazo, Transfer function restoration in 3D electron microscopy via iterative data refinement, *Physics in Medicine and Biology*, **49** (2004), 509–522.
- [21] VMC++ Monte Carlo package optimized for radiation treatment planning, see: <http://www.nrc-cnrc.gc.ca/eng/projects/inms/vmc.html>
- [22] Y. Yang and L. Xing, Inverse treatment planning with adaptively evolving voxel-dependent penalty scheme, *Medical Physics*, **31** (2004), 2839–2844.
- [23] L. Zhu, L. Lee, Y. Ma, Y. Ye, R. Mazzeo and L. Xing, Using total-variation regularization for intensity modulated radiation therapy inverse planning with field-specific numbers of segments, *Physics in Medicine and Biology* **53** (2008), 6653–6672.

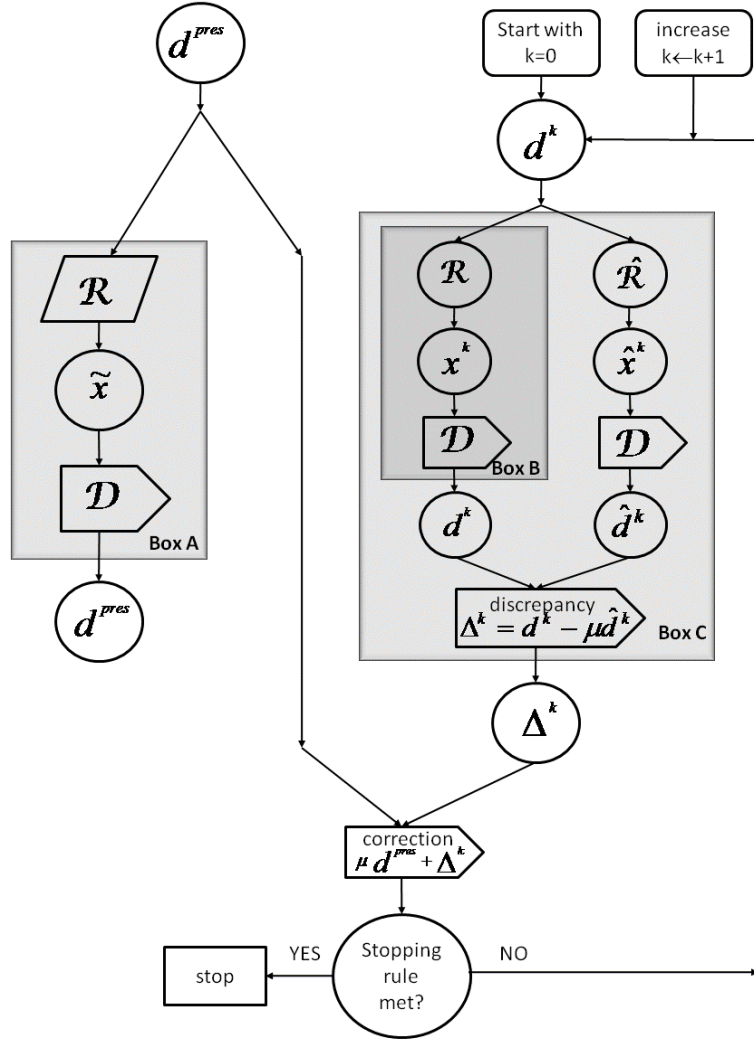


Figure 1: Outline of the IPR procedure.

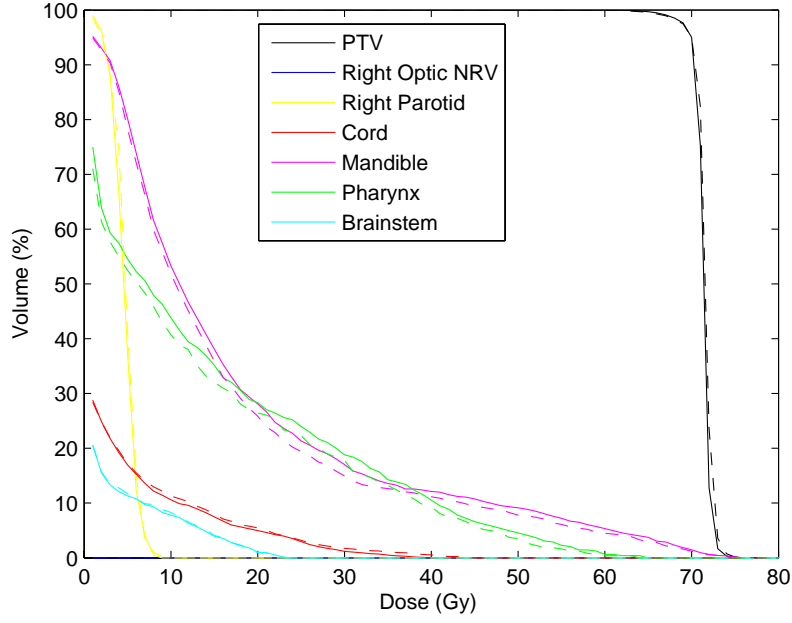


Figure 2: DVH comparison for the fixed ideal prescription and iterative prescription refinement plans. Solid lines represent the fixed prescription dose plan. Dashed lines correspond to the iterative prescription refinement plan. The plans are normalized such that 95% of the PTV receives 100% of the prescription dose (70 Gy)

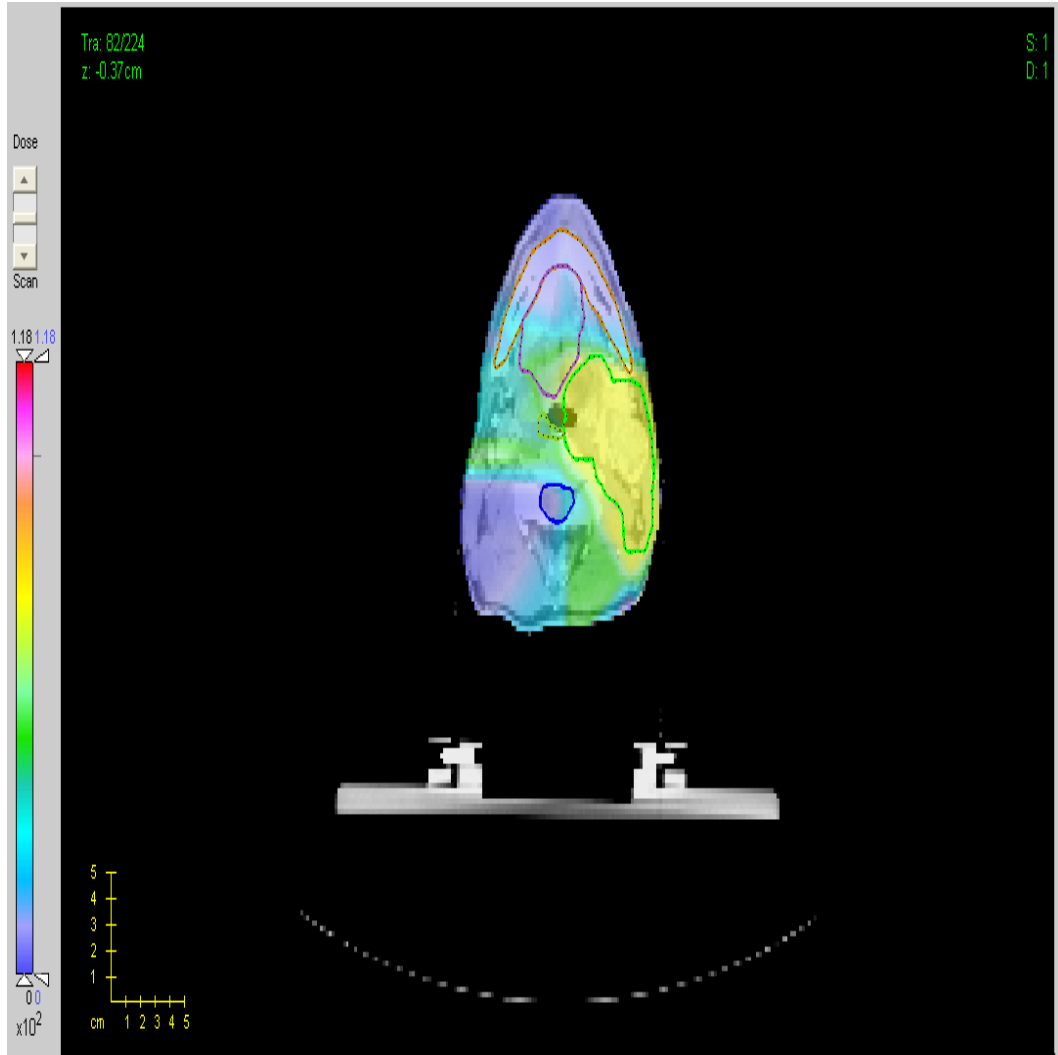


Figure 3: Dose map for the fixed prescription plan. Colors correspond to different doses (left slider). The PTV is outlined with the green dotted line. The cord, mandible, pharynx and oral cavity are outlined with the blue, orange, light green and purple lines

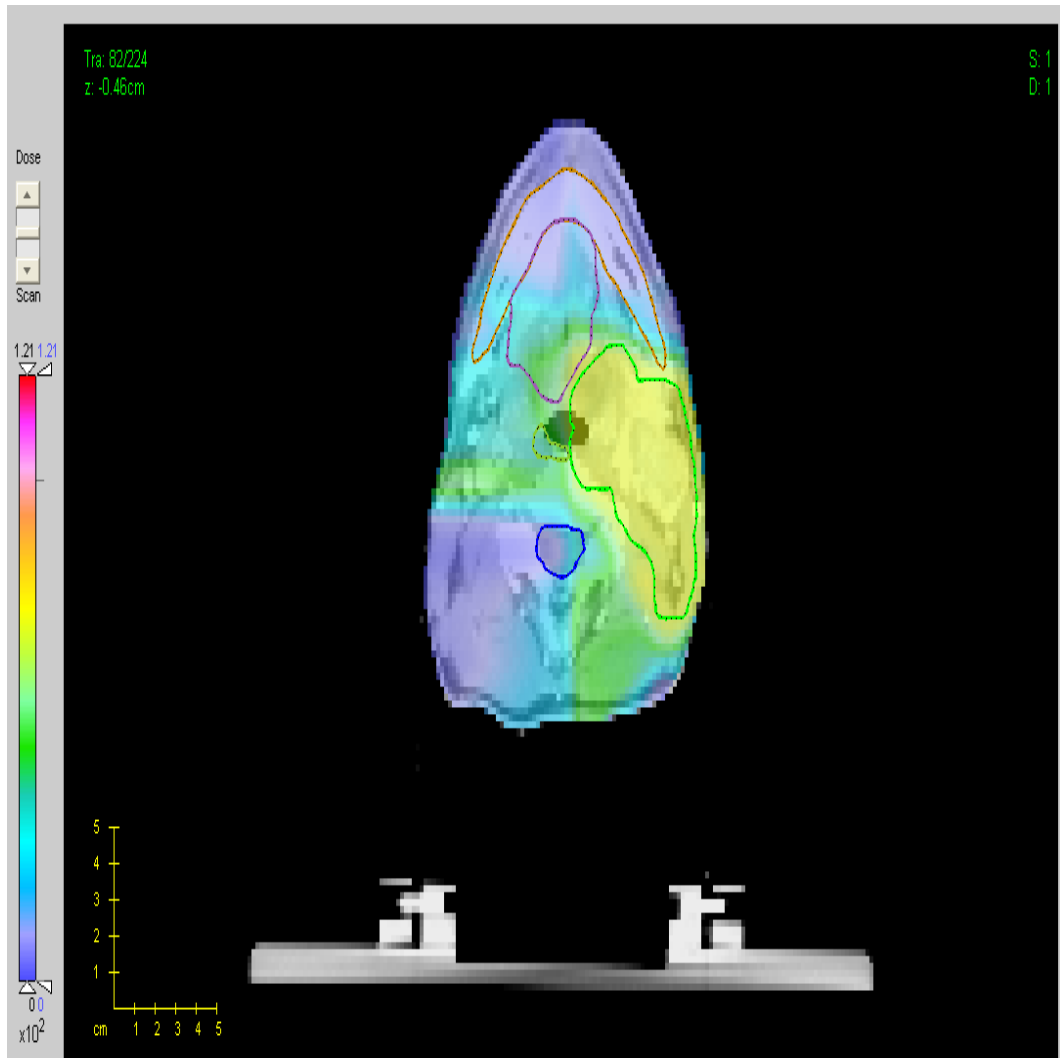


Figure 4: Dose map for the iterative prescription refinement plan. Colors correspond to different doses (left slider). The PTV is outlined with the green dotted line. The cord, mandible, pharynx and oral cavity are outlined with the blue, orange, light green and purple lines

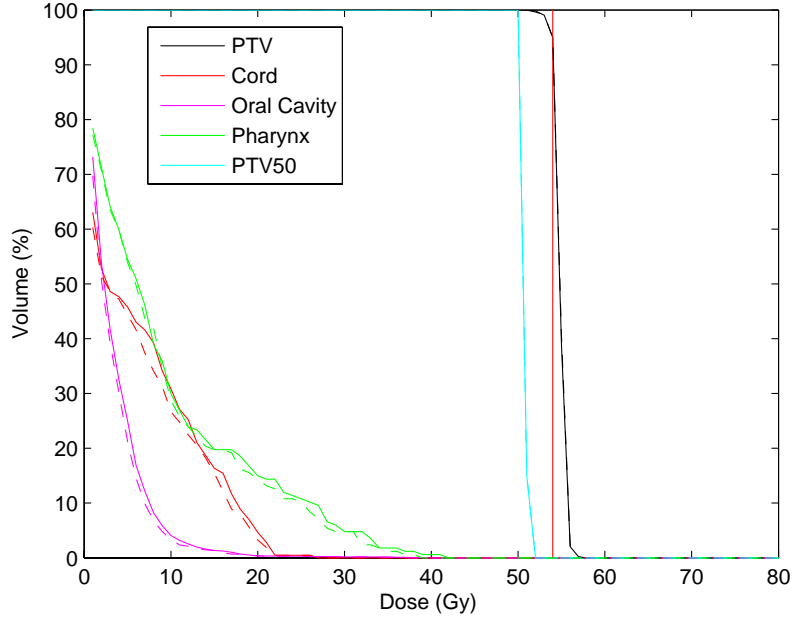


Figure 5: DVH comparison for the fixed ideal prescription and iterative prescription refinement plans. Solid lines represent the fixed prescription dose plan. Dashed lines correspond to the iterative prescription refinement plan. The plans are normalized such that 95% of the PTV receives 100% of the prescription dose (54 Gy)