Online Supplement for "Adaptive and robust radiation therapy optimization for lung cancer"

Timothy C. Y. Chan * Velibor V. Mišić †

May 21, 2013

Organization of Online Supplement

Section A provides further background on the two real patient PMF sequences used in the study. Section B provides results for the second PMF sequence. Section C provides results for a pathological PMF sequence. Section D provides the theoretical development necessary to prove Theorems 1 and 2 in Section 6 of the main body of the paper.

A Further background on PMF sequences

The two patient PMF sequences used in this paper were calculated using the method of Lujan et al. (1999) from breathing data obtained using Varian's real-time position management (RPM) system (Varian Medical Systems, Inc., Palo Alto, CA). In each imaging session, the RPM system was used to monitor the displacement over time of an external marker on the patient's abdomen. The full range of this displacement was divided into subintervals corresponding to the motion states obtained from the pre-treatment 4D-CT images. The proportion of the total time that the marker spent in each of the subintervals was then calculated, leading to a breathing motion PMF for that session. While the RPM system measures the external motion of a marker, which is not necessarily the same as the internal tumor motion, studies have shown that they are generally highly correlated (see, for example, Tsunashima et al. (2004) and Gierga et al. (2005)). We emphasize here that our adaptive robust method is not predicated on the use of the RPM system; any method is acceptable so long as it can be used to measure the patient's breathing motion and ultimately estimate the patient's breathing motion PMF. An imaging modality that can be used is 4D cone-beam CT (4D-CBCT), first introduced in Sonke et al. (2005). Using daily 4D-CBCT imaging, the treatment planner can directly measure the internal motion of the tumor and obtain accurate PMFs after each fraction.

B Results for the second PMF sequence

Table B.1 displays the dose statistics for the second PMF sequence. Figure B.1 plots the minimum tumor dose as a percentage of 72Gy versus the mean left lung dose as a percentage of the static margin mean left lung dose for each method applied to the second PMF sequence. Figure B.2 zooms in on the part of Figure B.1 containing the adaptive and prescient treatments.

*Department of Mechanical and Industrial Engineering, University of Toronto; Toronto ON, M5S 3G8, Canada; tcychan@mie.utoronto.ca

†Operations Research Center, Massachusetts Institute of Technology; 77 Massachusetts Avenue, Building E40, Cambridge, MA 02139, United States of America; vvmisic@mit.edu

Table B.1: Dose statistics for the second PMF sequence.

Implementation[1]	Min. tumor dose		Mean lung dose		Mean n. tissue dose		Mean lung tumor dose		Lung V20 tumor dose	
	Gy	% [2]	Gy	% [3]	Gy	% [4]	Gy	% [5]	Gy	% [6]
(S,N)	64.37	89.40	17.18	85.11	8.99	88.92	75.62	104.96	78.67	109.20
(ES(0.1),N)	70.46	97.87	17.51	86.78	9.00	88.99	81.20	112.70	88.44	122.75
(ES(0.5),N)	71.61	99.46	17.55	86.96	8.98	88.87	82.35	114.29	90.94	126.22
(ES(0.9),N)	71.69	99.56	17.54	86.92	8.98	88.85	82.48	114.47	91.23	126.62
(ES(1),N)	71.68	99.56	17.54	86.90	8.98	88.84	82.49	114.49	91.26	126.66
(RA,N)	71.49	99.30	17.55	86.97	8.99	88.90	82.20	114.09	90.87	126.13
(S,R)	71.99	99.98	17.99	89.12	9.39	92.90	80.77	112.11	84.60	117.43
(ES(0.1),R)	71.98	99.98	17.68	87.59	9.12	90.20	82.18	114.06	88.15	122.35
(ES(0.5),R)	72.00	100.00	17.59	87.15	9.01	89.16	82.61	114.66	91.17	126.54
(ES(0.9),R)	71.96	99.94	17.58	87.08	9.00	89.03	82.63	114.69	91.42	126.88
(ES(1),R)	71.94	99.92	17.57	87.04	8.99	88.97	82.65	114.72	91.60	127.14
(RA,R)	71.96	99.95	17.59	87.16	9.04	89.40	82.57	114.60	90.66	125.83
(S,M)	72.05	100.07	20.18	100.00	10.11	100.00	72.05	100.00	72.05	100.00
(ES(0.1),M)	72.05	100.07	18.07	89.54	9.36	92.62	80.47	111.69	82.73	114.82
(ES(0.5),M)	72.01	100.01	17.69	87.67	9.06	89.65	82.14	114.00	89.93	124.82
(ES(0.9),M)	71.98	99.97	17.65	87.43	9.02	89.27	82.33	114.27	90.85	126.10
(ES(1),M)	71.96	99.95	17.64	87.40	9.02	89.21	82.34	114.28	91.05	126.37
(RA,M)	72.02	100.03	17.75	87.94	9.15	90.48	81.90	113.67	87.59	121.57
(DLYP)	72.00	100.00	17.55	86.97	8.98	88.84	82.78	114.90	91.72	127.30
(AVGP)	72.00	100.00	17.57	87.06	8.98	88.85	82.70	114.78	91.79	127.40

[1] Under "Implementation", the first term indicates the type of method: static (S), exponential smoothing with parameter α (ES(α)), running average (RA), daily prescient (DLYP) and average prescient (AVGP). For the non-prescient methods, the second term indicates which of the uncertainty sets described in Section 5.1 was used as the initial uncertainty set: nominal (N), robust (R) and margin (M).

[2] The percentage under "Min. tumor dose" is the minimum tumor dose as a percentage of the prescription dose (72Gy).

[3] The percentage under "Mean lung dose" is the mean left lung dose as a percentage of the mean left lung dose delivered in the static margin treatment, indicated as implementation (S,M).

[4] The percentage under "Mean n. tissue dose" is defined analogously to the percentage under "Mean lung dose".

[5] The percentage under "MLLD-scaled tumor dose" is the minimum tumor dose as a percentage of the static margin minimum tumor dose when the treatment in question is scaled to have the same mean left lung dose as the static margin treatment.

[6] The percentage under "LLV20-scaled tumor dose" is the minimum tumor dose as a percentage of the static margin minimum tumor dose when the treatment in question is scaled to have the same left lung V20 as the static margin treatment.

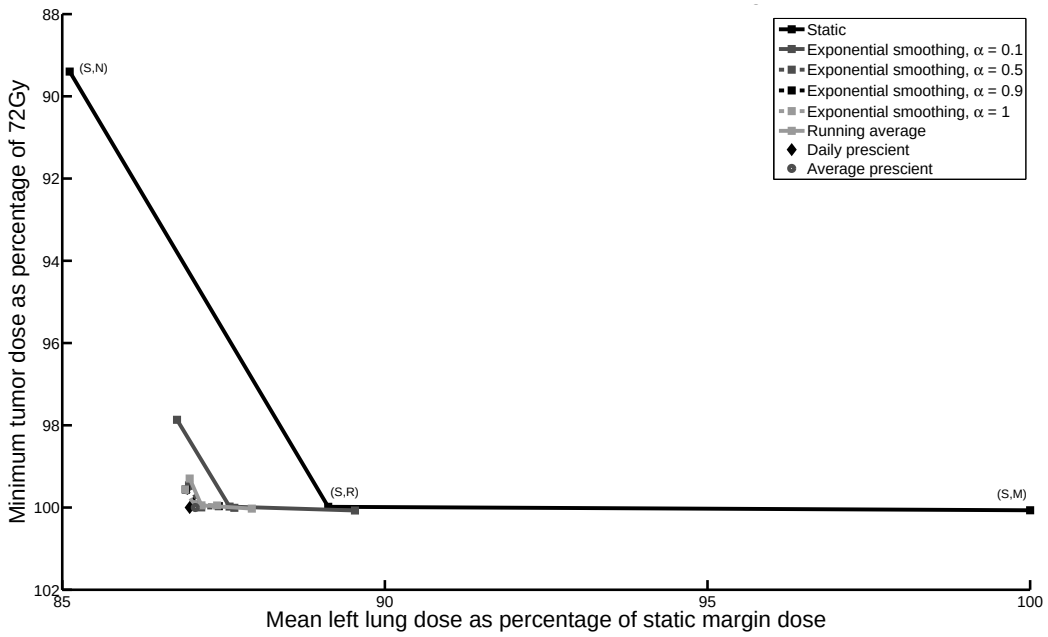

Figure B.1: Plot of minimum tumor dose versus mean left lung dose for the different implementations applied to the second PMF sequence (full). The points in the bottom left part of the plot are the adaptive robust and prescient treatments. Point labels follow the notation of Table B.1.

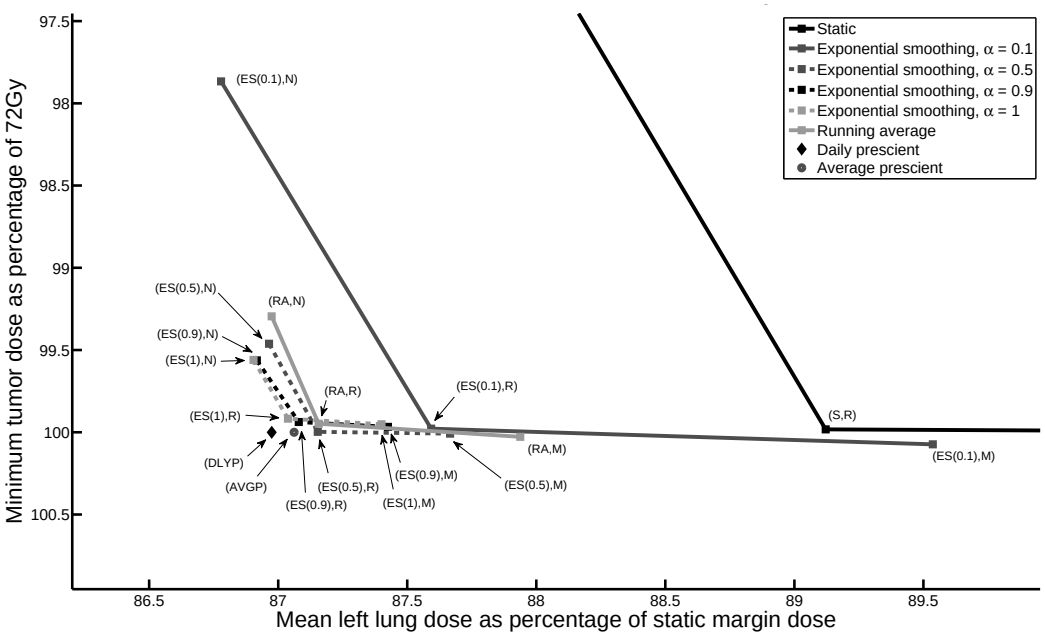

Figure B.2: Plot of minimum tumor dose versus mean left lung dose for the different implementations applied to the second PMF sequence (zoomed-in). Point labels follow the notation of Table B.1.

C A pathological PMF sequence example

Consider a sequence of 30 PMFs where

$$\mathbf{p}^{2k+1} = (0.05, 0.20, 0.50, 0.20, 0.05), \quad k = 0, \ldots, 14,$$
$$\mathbf{p}^{2k+2} = (0.30, 0.15, 0.10, 0.15, 0.30), \quad k = 0, \ldots, 14.$$

This PMF sequence oscillates between two PMFs over all 30 fractions, and these PMFs differ quite significantly from each other. Due to the oscillatory nature of this PMF sequence, we would expect that the lower and upper bounds from exponential smoothing would be constantly "out-of-sync" with the PMF that is realized on each day, resulting in tumor underdose by the end of the treatment.

We can verify this by plotting the realized PMFs and the lower and upper bound vectors from three exponential smoothing implementations for the last two fractions: Figure C.1 shows fraction 29, while Figure C.2 shows fraction 30. (For the $\alpha = 0.5$ and $\alpha = 0.9$ methods, the lower and upper bound vectors are virtually identical by the last fraction, which is why there is only one line for each of these methods on each of the plots.) We can very clearly see that in both fractions, the realized PMF is not in any of the uncertainty sets.

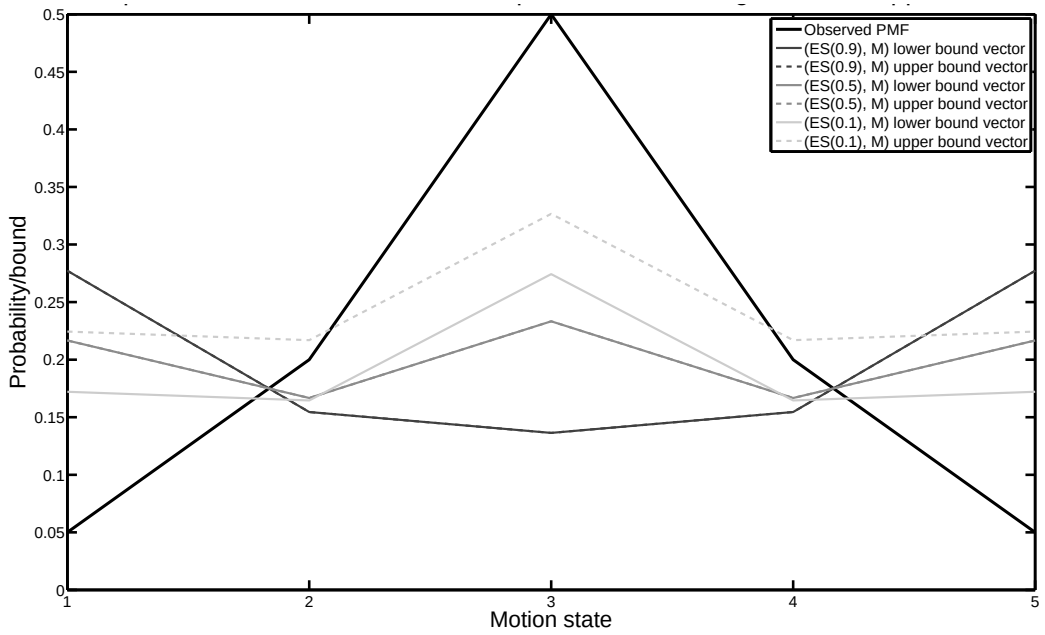

Figure C.1: Plot of observed PMF in fraction 29 versus lower/upper bound vectors for fraction 29 for three different exponential smoothing treatments.

The consequences of trying to track the oscillating PMFs are visible in the dose statistics that are shown in Table C.1. In Figures C.3 and C.4 we also plot the minimum tumor dose as a percentage of 72Gy versus mean left lung dose as a percentage of the static margin dose for each of the treatments for this PMF sequence. The same implementations considered for the first PMF sequence in Section 5 were considered for this sequence. The same patient anatomy and robust optimization parameters used for the real patient sequences were also used here. As we can see from Table C.1 and Figure C.3, the exponential smoothing update algorithm exhibits poorer performance with increasing α under this sequence. When $\alpha = 0$, the exponential smoothing and the static solutions are identical, so as α approaches 0, the exponential smoothing treatment will approach the static treatment with the same initial uncertainty set. For this particular sequence, it happens that the static nominal and robust treatments actually lead to high levels of tumor coverage, due to the fact that the nominal and robust uncertainty sets are very close to the average PMF of this sequence.

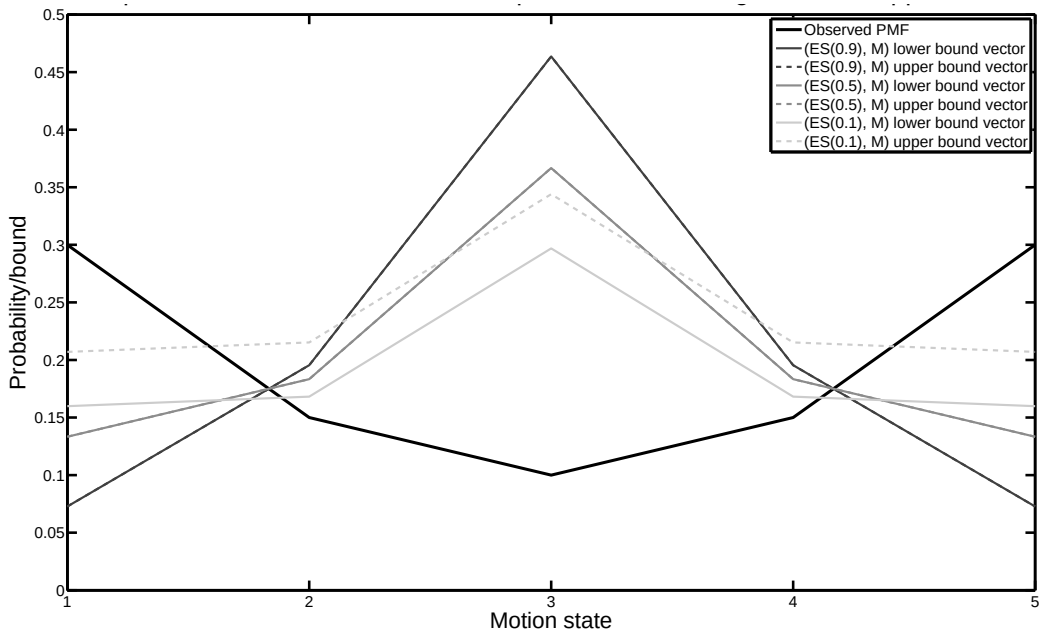

Figure C.2: Plot of observed PMF in fraction 30 versus lower/upper bound vectors for fraction 30 for three different exponential smoothing treatments.

In contrast to exponential smoothing, the running average method is relatively insensitive to the oscillatory nature of this PMF sequence. The reason for this is that this particular PMF sequence has a well-behaved average. When the running average update algorithm is applied to this sequence, the lower and upper bound vectors tend to the average of the sequence, and the intensity vector delivered in each fraction converges to the intensity vector that protects against the average PMF. As we noted in our discussion of prescient solutions in Section 4.2, delivering the intensity vector that protects against the average PMF in every fraction results in no tumor underdose, and the running average treatment, for a large number of fractions, is delivering a very similar intensity vector to the average PMF intensity vector, which is why it is still able to ensure a high level of tumor coverage.

These results are not too concerning from a clinical standpoint, because a patient PMF sequence that changes so drastically from fraction to fraction is very unlikely to be observed in practice. In the context of the patient anatomy that we have used for this sequence and for the real patient sequences, a patient breathing according to this sequence would alternate between spending 10% of the time in full inhale or exhale (motion states 1 and 5 respectively) during one treatment session, and spending 60% of the time in those same breathing phases during the next treatment session. As long as the uncertainty set update algorithm retains enough memory to avoid simply tracking the latest PMF, these pathological cases will not significantly affect the quality of the final adaptive robust treatment.

Table C.1: Dose statistics for the pathological PMF sequence.

Implementation [1]	Min. tumor dose		Mean lung dose		Mean n. tissue dose		Mean lung tumor dose		Lung V20 tumor dose	
	Gy	% [2]	Gy	% [3]	Gy	% [4]	Gy	% [5]	Gy	% [6]
(S,N)	71.25	98.96	17.39	85.31	9.01	89.00	83.52	115.90	87.17	120.97
(ES(0.1),N)	70.58	98.02	17.37	85.22	9.01	88.96	82.82	114.93	86.32	119.79
(ES(0.5),N)	63.46	88.14	17.41	85.39	9.00	88.82	74.32	103.14	77.90	108.11
(ES(0.9),N)	54.02	75.03	17.37	85.21	8.94	88.25	63.40	87.98	67.45	93.60
(RA,N)	70.76	98.28	17.37	85.24	9.01	88.94	83.02	115.21	86.45	119.97
(S,R)	71.97	99.96	18.19	89.23	9.41	92.92	80.66	111.93	84.57	117.36
(ES(0.1),R)	71.86	99.80	17.76	87.11	9.16	90.48	82.49	114.47	86.37	119.85
(ES(0.5),R)	64.27	89.26	17.49	85.81	9.03	89.20	74.90	103.94	78.12	108.41
(ES(0.9),R)	53.89	74.85	17.40	85.37	8.96	88.43	63.13	87.60	67.12	93.15
(RA,R)	71.50	99.30	17.57	86.17	9.08	89.64	82.97	115.14	86.39	119.89
(S,M)	72.06	100.08	20.38	100.00	10.13	100.00	72.06	100.00	72.06	100.00
(ES(0.1),M)	72.04	100.06	18.34	89.97	9.42	92.97	80.07	111.12	82.13	113.98
(ES(0.5),M)	64.76	89.95	17.62	86.43	9.09	89.71	74.93	103.99	77.77	107.93
(ES(0.9),M)	54.29	75.40	17.50	85.83	8.99	88.75	63.25	87.77	66.93	92.87
(RA,M)	72.03	100.04	17.89	87.76	9.20	90.86	82.07	113.89	85.14	118.15
(DLYP)	72.00	100.00	17.39	85.31	8.91	88.01	84.40	117.13	90.89	126.13
(AVGP)	72.00	100.00	17.39	85.30	9.01	88.98	84.41	117.14	88.17	122.35

[1] Under "Implementation", the first term indicates the type of method: static (S), exponential smoothing with parameter α (ES(α)), running average (RA), daily prescient (DLYP) and average prescient (AVGP). For the non-prescient methods, the second term indicates which of the uncertainty sets described in Section 5.1 was used as the initial uncertainty set: nominal (N), robust (R) and margin (M).

[2] The percentage under "Min. tumor dose" is the minimum tumor dose as a percentage of the prescription dose (72Gy).

[3] The percentage under "Mean lung dose" is the mean left lung dose as a percentage of the mean left lung dose delivered in the static margin treatment, indicated as implementation (S,M).

[4] The percentage under "Mean n. tissue dose" is defined analogously to the percentage under "Mean lung dose".

[5] The percentage under "MLLD-scaled tumor dose" is the minimum tumor dose as a percentage of the static margin minimum tumor dose when the treatment in question is scaled to have the same mean left lung dose as the static margin treatment.

[6] The percentage under "LLV20-scaled tumor dose" is the minimum tumor dose as a percentage of the static margin minimum tumor dose when the treatment in question is scaled to have the same left lung V20 as the static margin treatment.

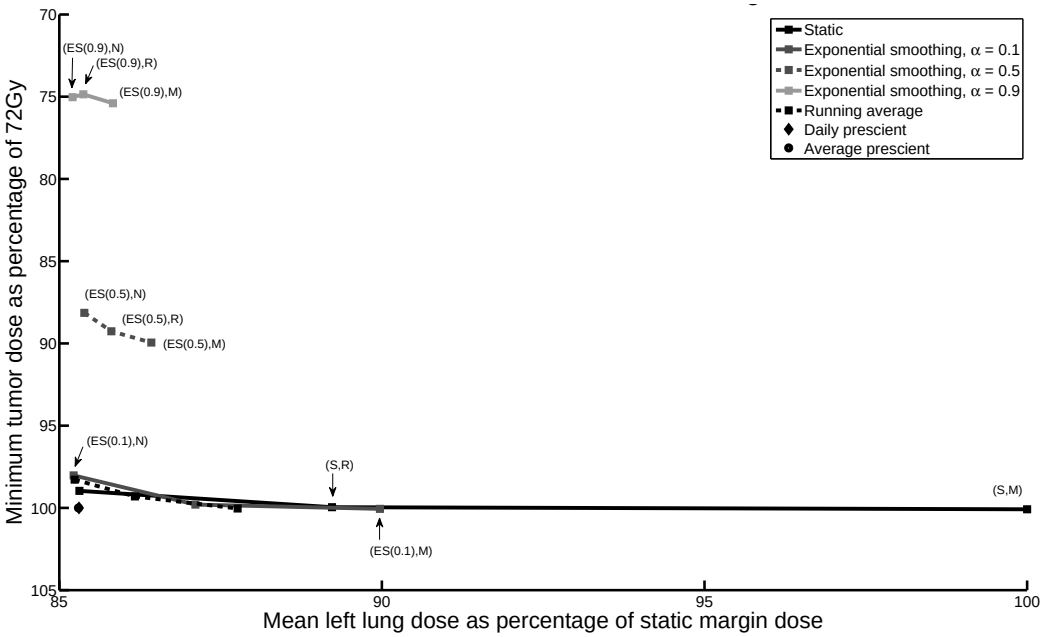

Figure C.3: Plot of minimum tumor dose versus mean left lung dose for the different implementations applied to the pathological PMF sequence (full).

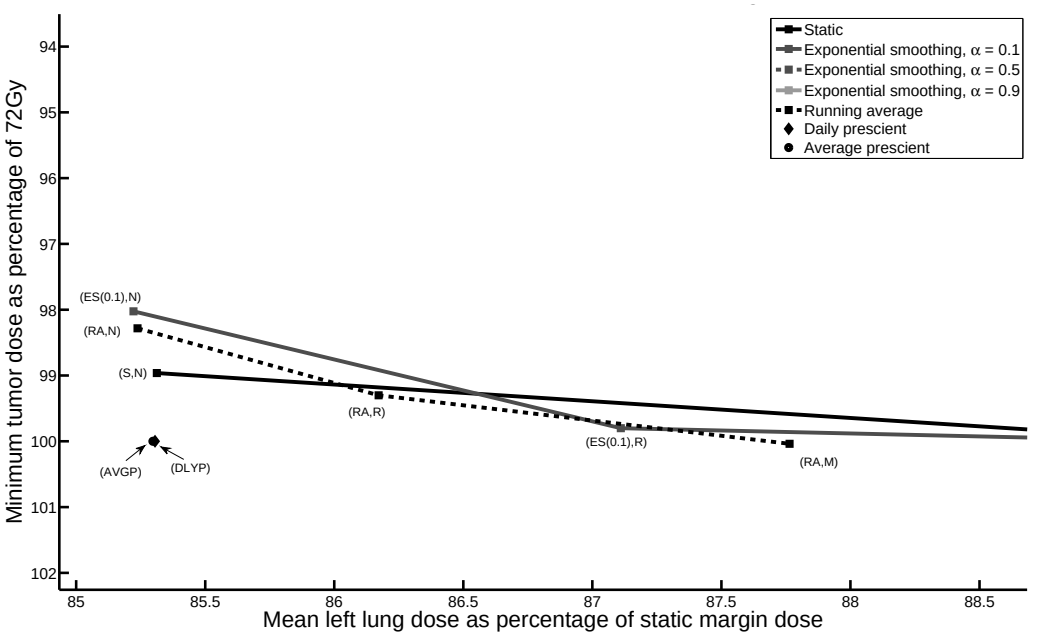

Figure C.4: Plot of minimum tumor dose versus mean left lung dose for the different implementations applied to the pathological PMF sequence (zoomed-in).

D Proofs

In this section, we prove Theorems 1 and 2 in Section 6. The arc of reasoning we will take to prove these results is as follows. After stating our assumptions in Section D.1, we begin in Section D.2 by defining an alternative robust problem, where instead of enforcing the minimum and maximum tumor dose constraints for a (possibly infinite) set of PMFs P, we enforce those constraints at finitely many PMFs $\mathbf{q}^1, \ldots, \mathbf{q}^M$. This problem is equivalent to the original robust problem under a simple condition. Next, we show that if the defining PMFs of the alternative robust problem, $\mathbf{q}^1, \ldots, \mathbf{q}^M$, all converge to a single PMF $\mathbf{p}^*$ in the usual mathematical sense, then the feasible regions of the alternative robust problems along the way converge to the feasible region of the alternative robust problem with all M defining PMFs set to $\mathbf{p}^*$. We then build on this result to show that the corresponding sets of optimal solutions also converge.

Subsequently, in Section D.3, we show that if the sequence of breathing motion PMFs $(\mathbf{p}^i)_{i=1}^{\infty}$ converges to a PMF $\mathbf{p}^*$ and the sequences of lower and upper bound vectors $(\boldsymbol{\ell}^i)_{i=1}^{\infty}$ and $(\mathbf{u}^i)_{i=1}^{\infty}$ are obtained by any convex-convergent update algorithm, then the optimal solution sets of the original robust problems enter the epsilon neighborhood of the set of optimal solutions to the nominal problem with respect to $\mathbf{p}^*$. Finally, in Section D.4, we prove our core result of convergence of the dose distribution, where we also establish the same convergence result for the dose distributions obtained from the daily and average prescient methods.

We begin by describing in Section D.1 the assumptions that were made in the analysis.

D.1 Assumptions

Assumption 1 *For every voxel $v \in \mathcal{T}$ and motion state $x \in X$, there exists a beamlet $b \in \mathcal{B}$ such that $\Delta_{v,x,b} > 0$.*

This assumption is reasonable because it simply asks for the number of beamlets to not be too small. If this assumption does not hold, then the margin problem is infeasible.

Let γ^* be defined as the optimal objective value of the following linear program:

$$\begin{aligned} \text{minimize } & \gamma \\ \text{subject to } & \sum_{b \in \mathcal{B}} \Delta_{v,x,b} w_b \geq \theta_v, \quad \forall v \in \mathcal{T},\ x \in X, \\ & \sum_{b \in \mathcal{B}} \Delta_{v,x,b} w_b \leq \gamma \theta_v, \quad \forall v \in \mathcal{T},\ x \in X, \\ & \gamma \geq 1, \\ & w_b \geq 0, \quad \forall b \in \mathcal{B}. \end{aligned} \tag{1}$$

It is the lowest value of the parameter γ of problem (1) under which the margin problem remains feasible. Problem (1) is bounded because $\gamma \geq 1$, and feasible because of Assumption 1. Our next assumption is on how the maximum dose parameter γ is selected:

Assumption 2 *The parameter γ of problem (1) satisfies $\gamma > \gamma^*$.*

If γ is strictly less than γ^*, then the margin problem is infeasible. If γ is exactly equal to γ^*, then every feasible solution of the margin is such that there is some $v \in \mathcal{T}$ and some $x \in X$ such that $\sum_{b \in \mathcal{B}} \Delta_{v,x,b} w_b = \gamma \theta_v$. By setting γ to be slightly higher than γ^*, we allow for a slightly higher amount of dose in the tumor. The importance of the maximum dose constraint in the formulation of the robust problem is to control the homogeneity of the tumor dose (see Bortfeld et al. (2008)); the homogeneity we lose by increasing γ by a small amount above γ^* is negligible. This assumption is satisfied for the data used in Section 5.

An immediate consequence of Assumptions 1 and 2 is the following proposition.

Proposition 1 *There exists a beamlet intensity vector $\mathbf{w}$ such that for every $b \in \mathcal{B}$, $w_b > 0$, and for every $v \in \mathcal{T}$ and every $x \in X$, $\theta_v < \sum_{b \in \mathcal{B}} \Delta_{v,x,b} w_b < \gamma \theta_v$.*

Proof: Since γ^* is the minimum of problem (1), there must be an intensity vector $\mathbf{w}$ that is feasible for γ^*, which means that for every $v \in \mathcal{T}$ and every $x \in X$, $\theta_v \leq \sum_{b \in \mathcal{B}} \Delta_{v,x,b} w_b \leq \gamma^* \theta_v$ and for every $b \in \mathcal{B}$, $w_b \geq 0$.

Since $\gamma > \gamma^*$ by Assumption 2, this means that for every $v \in \mathcal{T}$ and $x \in X$, $\theta_v \leq \sum_{b\in\mathcal{B}} \Delta_{v,x,b} w_b < \gamma\theta_v$. Assumption 1 ensures that by infinitesimally increasing certain beamlets, we can modify $\mathbf{w}$ to obtain a $\mathbf{w}'$ for which $\theta_v < \sum_{b\in\mathcal{B}} \Delta_{v,x,b} w'_b$ for every $v \in \mathcal{T}$ and $x \in X$. Then, by infinitesimally increasing the beamlets that are zero, we can modify $\mathbf{w}'$ to obtain a $\bar{\mathbf{w}}$ that additionally satisfies $\bar{w}_b > 0$ for each $b \in \mathcal{B}$. ■

Assumption 3 *For every beamlet $b \in \mathcal{B}$ and every phase $x \in X$, there exists a voxel $v \in \mathcal{V}$ such that $\Delta_{v,x,b} > 0$.*

This assumption is reasonable, because it simply asks that the patient geometry accounts for a sufficiently large amount of the healthy tissue around the tumor. This condition is satisfied for the data used in Section 5.

The importance of Assumption 3 is that it implies that the robust problem has a bounded optimal solution set; the proof of this is straightforward and is omitted.

Proposition 2 *The set of optimal solutions to the robust problem* (1) *for any uncertainty set P is a bounded set.*

D.2 An alternative robust problem

Given M PMFs $\mathbf{q}^1, \ldots, \mathbf{q}^M$, the alternative robust problem is defined as

$$\begin{aligned}
\text{minimize} \quad & \sum_{v\in\mathcal{V}} \sum_{x\in X} \sum_{b\in\mathcal{B}} \Delta_{v,x,b} \bar{p}(x) w_b \\
\text{subject to} \quad & \sum_{x\in X} \sum_{b\in\mathcal{B}} \Delta_{v,x,b} q^j(x) w_b \geq \theta_v, \quad \forall v \in \mathcal{T},\ j \in \{1,\ldots,M\}, \\
& \sum_{x\in X} \sum_{b\in\mathcal{B}} \Delta_{v,x,b} q^j(x) w_b \leq \gamma\theta_v, \quad \forall v \in \mathcal{T},\ j \in \{1,\ldots,M\}, \\
& w_b \geq 0, \quad \forall b \in \mathcal{B}.
\end{aligned} \tag{2}$$

We will refer to the M PMFs $\mathbf{q}^1, \ldots, \mathbf{q}^M$ as the *defining PMFs* of the corresponding problem (2). The convex hull of these PMFs is denoted as $\text{conv}\,\{\mathbf{q}^1, \ldots, \mathbf{q}^M\}$. Under a very simple condition, the alternative robust problem is equivalent to the original robust problem.

Lemma 1 *If $P = \text{conv}\,\{\mathbf{q}^1, \ldots, \mathbf{q}^M\}$, then problems* (1) *and* (2) *are equivalent.*

Proof: Problem (2) is clearly equivalent to problem (1) where the uncertainty set is the finite set $P' = \{\mathbf{q}^1, \ldots, \mathbf{q}^M\}$. The equivalence of this new problem to problem (1) with P follows from the fact that a robust linear program remains unchanged when the uncertainty set is replaced by its convex hull (see, for instance, Ben-Tal & Nemirovski (1999)). ■

Now, let $H(\mathbf{q}^1, \ldots, \mathbf{q}^M)$ denote the feasible region of the alternative robust problem (2) with the M defining PMFs $\mathbf{q}^1, \ldots, \mathbf{q}^M$.

Proposition 3 *Suppose that the ith instance of the alternative robust problem* (2) *is defined by the M defining PMFs $\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i}$, for all $i \in \mathbb{Z}_+$. If $\mathbf{q}^{j,i} \to \mathbf{p}^*$ as $i \to \infty$ for all $j \in \{1,\ldots,M\}$, then*

$$\lim_{i\to\infty} H(\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i}) = H(\mathbf{p}^*, \ldots, \mathbf{p}^*),$$

where the notion of set convergence is that of Dantzig et al. (1967).

In other words, if the defining PMFs of (2) all converge to some PMF $\mathbf{p}^*$, then the feasible region of (2) with those defining PMFs converges to the feasible region of (2) with all defining PMFs set to $\mathbf{p}^*$. Note that when all M defining PMFs are set to $\mathbf{p}^*$, problem (2) is equivalent to the nominal problem with respect to $\mathbf{p}^*$.

Before we prove Proposition 3, it is helpful to introduce some additional notation. Let the matrix $\mathbf{A}(\mathbf{q}^1, \ldots, \mathbf{q}^M)$ be defined as

$$\mathbf{A}(\mathbf{q}^1, \ldots, \mathbf{q}^M) = \begin{bmatrix} \mathbf{T}(\mathbf{q}^1, \ldots, \mathbf{q}^M) \\ -\mathbf{T}(\mathbf{q}^1, \ldots, \mathbf{q}^M) \\ -\mathbf{I} \\ \mathbf{0}^T \end{bmatrix},$$

where $\mathbf{I}$ is the $|\mathcal{B}|$-by-$|\mathcal{B}|$ identity matrix, $\mathbf{0}^T$ is a row vector of zeroes of length $|\mathcal{B}|$ and $\mathbf{T}(\mathbf{q}^1, \ldots, \mathbf{q}^M)$ is a matrix defined (component-wise) as

$$T_{(v,j),b}(\mathbf{q}^1, \ldots, \mathbf{q}^M) = \sum_{x \in X} \Delta_{v,x,b} q^j(x).$$

The rows of $\mathbf{T}(\mathbf{q}^1, \ldots, \mathbf{q}^M)$ correspond to (v, j) pairs ($v \in \mathcal{T}$, $j \in \{1, \ldots, M\}$) and the columns correspond to beamlets $b \in \mathcal{B}$.

Let the column vector $\mathbf{c}$ be defined as

$$\mathbf{c} = \begin{bmatrix} -\gamma \mathbf{z} \\ \mathbf{z} \\ \mathbf{0} \\ 0 \end{bmatrix},$$

where $\mathbf{0}$ is a $|\mathcal{B}|$-component column vector of zeroes, 0 is the scalar zero and $\mathbf{z}$ is a column vector defined (component-wise) as

$$z_{(v,j)} = \theta_v.$$

To understand why $\mathbf{A}$ and $\mathbf{c}$ are defined in this way, consider the set

$$\mathcal{W} = \left\{ \mathbf{w} \in \mathbb{R}^{|\mathcal{B}|} \;\middle|\; \mathbf{A}(\mathbf{q}^1, \ldots, \mathbf{q}^M)\mathbf{w} + \mathbf{c} \leq \mathbf{0} \right\},$$

where $\mathbf{0}$ is a $2M|\mathcal{T}| + |\mathcal{B}| + 1$-component column vector of zeroes and the inequality is taken component-wise. The set $\mathcal{W}$ is the set of all $\mathbf{w} \in \mathbb{R}^{|\mathcal{B}|}$ such that

$$\begin{aligned} &\sum_{x \in X} \Delta_{v,x,b} q^j(x) w_b - \gamma \theta_v \leq 0, \quad \forall j \in \{1, \ldots, M\}, v \in \mathcal{T}, \\ &- \sum_{x \in X} \Delta_{v,x,b} q^j(x) w_b + \theta_v \leq 0, \quad \forall j \in \{1, \ldots, M\}, v \in \mathcal{T}, \\ &- w_b \leq 0, \quad \forall b \in \mathcal{B}, \end{aligned}$$

which is precisely the set of feasible solutions to problem (2) with the defining PMFs $\mathbf{q}^1, \ldots, \mathbf{q}^M$. In other words, the set of feasible $\mathbf{w}$ can be written as a set of the form $\left\{ \mathbf{w} \in \mathbb{R}^{|\mathcal{B}|} \;\middle|\; g(\mathbf{w}) \leq 0 \right\}$, where g is an affine function.

We define the rank of an affine function $g(\mathbf{w}) = \mathbf{A}\mathbf{w} + \mathbf{c}$ as the rank of the matrix $\mathbf{A}$. If $g : \mathbb{R}^n \to \mathbb{R}^m$ and $I \subset \{1, \ldots, m\}$, then the function g_I is the function consisting only of those coordinates of g corresponding to indices in I, i.e. $g_I(x) = \mathbf{A}_I x + \mathbf{c}_I$ where $\mathbf{A}_I$ ($\mathbf{c}_I$) consists of the rows of $\mathbf{A}$ (components of $\mathbf{c}$) corresponding to indices in I.

We can now prove Proposition 3.

Proof of Proposition 3: Let $(f^i)_{i=1}^{\infty}$ be a sequence of affine functions, where $f^i : \mathbb{R}^{|\mathcal{B}|} \to \mathbb{R}^{2M|\mathcal{T}|+|\mathcal{B}|+1}$, defined as

$$f^i(\mathbf{w}) = \mathbf{A}(\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i})\mathbf{w} + \mathbf{c},$$

and let $f : \mathbb{R}^{|\mathcal{B}|} \to \mathbb{R}^{2M|\mathcal{T}|+|\mathcal{B}|+1}$ be an affine function defined as

$$f(\mathbf{w}) = \mathbf{A}(\mathbf{p}^*, \ldots, \mathbf{p}^*)\mathbf{w} + \mathbf{c}.$$

By the definition of the $\mathbf{A}$ matrices, it is clear that if $\mathbf{q}^{j,i} \to \mathbf{p}^*$ as $i \to \infty$ for each $j \in \{1, \ldots, M\}$, then $f^i \to f$ pointwise as $i \to \infty$.

Let $H(f^i) = \left\{\mathbf{w} \in \mathbb{R}^{|\mathcal{B}|} \,\middle|\, f^i(\mathbf{w}) \leq \mathbf{0}\right\}$ and $H(f) = \left\{\mathbf{w} \in \mathbb{R}^{|\mathcal{B}|} \,\middle|\, f(\mathbf{w}) \leq \mathbf{0}\right\}$, where $\mathbf{0}$ is a $2M|\mathcal{T}| + |\mathcal{B}| + 1$-component column vector of zeroes and the inequality is taken component-wise. We can immediately see that

$$H(f^i) = H(\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i})$$

and

$$H(f) = H(\mathbf{p}^*, \ldots, \mathbf{p}^*).$$

The margin problem uncertainty set contains the singleton set $\{\mathbf{p}^*\}$, so the feasible region of the margin problem is contained in the feasible region of the nominal problem with respect to $\mathbf{p}^*$. The nominal problem with respect to $\mathbf{p}^*$ is equivalent to the alternative robust problem (2) with all M defining PMFs set to $\mathbf{p}^*$. By Proposition 1, the margin problem has an interior solution, so the feasible region of the margin problem is non-empty. Therefore, $H(f)$ is also non-empty.

Now, define I as

$$I = \{k \mid f_k(\mathbf{w}) = 0 \text{ for all } \mathbf{w} \in H(f)\},$$

with k ranging over all the coordinates of f (row indices of the matrix $\mathbf{A}(\mathbf{p}^*, \ldots, \mathbf{p}^*)$). The set I consists of only the index corresponding to the last row of the matrix $\mathbf{A}(\mathbf{p}^*, \ldots, \mathbf{p}^*)$. This is because the $\mathbf{w}$ of Proposition 1 satisfies $-\mathbf{I}\mathbf{w} < \mathbf{0}$ (from the condition on the w_b's being strictly positive), $-\mathbf{T}(\mathbf{p}^*, \ldots, \mathbf{p}^*)\mathbf{w} + \mathbf{z} < 0$ (because $\mathbf{w}$ satisfies $\sum_{b\in\mathcal{B}} \Delta_{v,x,b} w_b > \theta_v$ for all $v \in \mathcal{T}$, $x \in X$) and $\mathbf{T}(\mathbf{p}^*, \ldots, \mathbf{p}^*)\mathbf{w} - \gamma\mathbf{z} < 0$ (because $\mathbf{w}$ satisfies $\sum_{b\in\mathcal{B}} \Delta_{v,x,b} w_b < \gamma\theta_v$ for all $v \in \mathcal{T}$, $x \in X$). In contrast, the last coordinate of the affine function f is

$$\mathbf{0}^T\mathbf{w} + 0 = 0$$

regardless of $\mathbf{w}$. Therefore, the set I consists of only the index of this last coordinate.

This means that $\operatorname{rank} f_I = \operatorname{rank}\left[\mathbf{0}^T\right] = 0$ and for every i, $\operatorname{rank} f_I^i = \operatorname{rank}\left[\mathbf{0}^T\right] = 0$, and in particular that

$$\limsup_{i\to\infty} \left(\operatorname{rank} f_I^i\right) \leq \operatorname{rank} f_I. \tag{3}$$

Invoking now Theorem II.2.2 of Dantzig et al. (1967) which requires (3), we have that either $\lim_{i\to\infty} H(f^i) = H(f)$ or $H(f^i)$ is empty for infinitely many i. The latter case is not possible by Proposition 1 (the $\mathbf{w}$ of Proposition 1 belongs to every set $H(f^i)$). Therefore, we must have $\lim_{i\to\infty} H(f^i) = H(f)$ or equivalently, $\lim_{i\to\infty} H(\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i}) = H(\mathbf{p}^*, \ldots, \mathbf{p}^*)$. ■

Next, we show that the sets of optimal solutions converge. Define the function $Z : \mathbb{R}^{|\mathcal{B}|} \to \mathbb{R}$ as

$$Z(\mathbf{w}) = \sum_{v\in\mathcal{V}} \sum_{x\in X} \sum_{b\in\mathcal{B}} \Delta_{v,x,b} \bar{p}(x) w_b;$$

Z is the objective function of both problems (1) and (2). Suppose that V is a subset of $\mathbb{R}^n$ and $\phi : \mathbb{R}^n \to \mathbb{R}$ is a function. Following Dantzig et al. (1967), we define the set $M(\phi \mid V)$ as

$$M(\phi \mid V) = \{x \in V \mid \phi(x) = \inf\{\phi(y) \mid y \in V\}\}.$$

When the infimum can be attained, $M(\phi \mid V)$ is the set of optimal solutions to the constrained minimization problem $\min_{x\in V} \phi(x)$.

The next result states that, if the defining PMFs of problem (2) all converge to a PMF $\mathbf{p}^*$, then the corresponding optimal solution sets eventually enter the epsilon neighborhood of the optimal solution set of problem (2) with all of the defining PMFs set to $\mathbf{p}^*$ (which, by our note above, is the optimal solution set of the nominal problem with respect to $\mathbf{p}^*$).

Proposition 4 *Suppose that for every $j \in \{1, \ldots, M\}$, $\mathbf{q}^{j,i} \to \mathbf{p}^*$ as $i \to \infty$. Then for every $\epsilon > 0$, there exists an $N \in \mathbb{Z}_+$ such that for $i > N$, $M(Z \mid H(\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i})) \subseteq U(M(Z \mid H(\mathbf{p}^*, \ldots, \mathbf{p}^*)), \epsilon)$.*

Proof of Proposition 4: By Proposition 3, whenever $(\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i})$ converges to $(\mathbf{p}^*, \ldots, \mathbf{p}^*)$ as $i \to \infty$, $H(\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i})$ converges to $H(\mathbf{p}^*, \ldots, \mathbf{p}^*)$. For any $(\mathbf{q}^1, \ldots, \mathbf{q}^M)$-tuple, $H(\mathbf{q}^1, \ldots, \mathbf{q}^M)$ is non-empty because it contains the $\mathbf{w}$ of Proposition 1. The set $H(\mathbf{q}^1, \ldots, \mathbf{q}^M)$ is also a closed and convex set (it is a polyhedron). The function Z is a linear function on $\mathbb{R}^{|\mathcal{B}|}$, and therefore continuous and quasiconvex. The

set $M(Z \mid H(\mathbf{p}^*, \ldots, \mathbf{p}^*))$ is non-empty because by the constraints in $H(\mathbf{p}^*, \ldots, \mathbf{p}^*)$, each beamlet's intensity w_b is non-negative, so Z is always bounded from below by 0, and

$$\begin{aligned} &\text{minimize } Z(\mathbf{w}) \\ &\text{subject to } \mathbf{w} \in H(\mathbf{p}^*, \ldots, \mathbf{p}^*) \end{aligned}$$

is a linear optimization problem. The set $M(Z \mid H(\mathbf{p}^*, \ldots, \mathbf{p}^*))$ is a bounded set, because it is the set of optimal solutions of the nominal problem with respect to $\mathbf{p}^*$, which is bounded by Proposition 2. The set $U(M(Z \mid H(\mathbf{p}^*, \ldots, \mathbf{p}^*)), \epsilon)$ is therefore the epsilon neighborhood of a bounded set, and so it must also be bounded. Since $U(M(Z \mid H(\mathbf{p}^*, \ldots, \mathbf{p}^*)), \epsilon)$ is the union of open sets, it is open.

From these observations, the hypotheses of Theorem I.3.3 of Dantzig et al. (1967) are satisfied. Invoking it furnishes a neighborhood of $(\mathbf{p}^*, \ldots, \mathbf{p}^*)$ such that whenever $(\mathbf{q}^1, \ldots, \mathbf{q}^M)$ is inside that neighborhood, $M(Z \mid H(\mathbf{q}^1, \ldots, \mathbf{q}^M)) \subseteq U(M(Z \mid H(\mathbf{p}^*, \ldots, \mathbf{p}^*)), \epsilon)$. Since each $(\mathbf{q}^{j,i})_{i=1}^{\infty}$ sequence converges to $\mathbf{p}^*$, there must be an $N \in \mathbb{Z}_+$ such that for $i > N$, $(\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i})$ is within that neighborhood, and thus $M(Z \mid H(\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i})) \subseteq U(M(Z \mid H(\mathbf{p}^*, \ldots, \mathbf{p}^*)), \epsilon)$. ■

D.3 Convergence of optimal solutions sets under convex-convergent update algorithms

Having developed convergence results for the alternative robust problem (2), we now connect these results to the robust problem (1). The next result states that $\mathbf{w}^*(\boldsymbol{\ell}^i, \mathbf{u}^i)$ eventually enters the epsilon neighborhood of $\mathbf{w}^*(\mathbf{p}^*)$.

Proposition 5 *Let $(\mathbf{p}^i)_{i=1}^{\infty}$ be a sequence of PMFs that converges to $\mathbf{p}^*$. Let $(\boldsymbol{\ell}^i)_{i=1}^{\infty}$ and $(\mathbf{u}^i)_{i=1}^{\infty}$ be lower and upper bound sequences generated from $(\mathbf{p}^i)_{i=1}^{\infty}$ by any convex-convergent update algorithm. Then for every $\epsilon > 0$, there exists an $N \in \mathbb{Z}_+$ such that for all $i > N$, $\mathbf{w}^*(\boldsymbol{\ell}^i, \mathbf{u}^i) \subseteq U(\mathbf{w}^*(\mathbf{p}^*), \epsilon)$.*

Before proving Proposition 5, we need to establish an auxilliary lemma. This lemma states that if we modify an uncertainty set P by taking the convex combination of $\boldsymbol{\ell}$ and $\mathbf{u}$ with an observed PMF $\tilde{\mathbf{p}}$, then the resulting uncertainty set is the same as one obtained by applying the same convex combination to the M defining PMFs $\mathbf{q}^1, \ldots, \mathbf{q}^M$ whose convex hull is P.

Lemma 2 *Let $\alpha \in [0, 1]$. Let P be an uncertainty set defined by $\boldsymbol{\ell}$ and $\mathbf{u}$, and let $\mathbf{q}^1, \ldots, \mathbf{q}^M$ be M PMFs such that $P = \operatorname{conv}\{\mathbf{q}^1, \ldots, \mathbf{q}^M\}$. Let $\tilde{\mathbf{p}} \in \mathcal{P}$, and let $\tilde{P}$ be the uncertainty set defined by*

$$\begin{aligned} \tilde{\boldsymbol{\ell}} &= (1-\alpha)\boldsymbol{\ell} + \alpha\tilde{\mathbf{p}}, \\ \tilde{\mathbf{u}} &= (1-\alpha)\mathbf{u} + \alpha\tilde{\mathbf{p}}. \end{aligned}$$

Then $\tilde{P} = \operatorname{conv}\{\tilde{\mathbf{q}}^1, \ldots, \tilde{\mathbf{q}}^M\}$, where

$$\tilde{\mathbf{q}}^j = (1-\alpha)\mathbf{q}^j + \alpha\tilde{\mathbf{p}},$$

for every $j \in \{1, \ldots, M\}$.

Proof: If $\alpha = 1$, then $\tilde{\boldsymbol{\ell}} = \tilde{\mathbf{u}} = \tilde{\mathbf{p}}$, so $P = \{\tilde{\mathbf{p}}\}$. At the same time, $\tilde{\mathbf{q}}^j = \tilde{\mathbf{p}}$ for every j, so $\operatorname{conv}\{\tilde{\mathbf{q}}^1, \ldots, \tilde{\mathbf{q}}^M\} = \{\tilde{\mathbf{p}}\}$. Now consider $\alpha < 1$. We will first show that $\tilde{P} \subseteq \operatorname{conv}\{\tilde{\mathbf{q}}^1, \ldots, \tilde{\mathbf{q}}^M\}$. Let $\mathbf{p}^2 \in \tilde{P}$, and define $\mathbf{p}^1$ as

$$\mathbf{p}^1 = \frac{1}{1-\alpha}\mathbf{p}^2 - \frac{\alpha}{1-\alpha}\tilde{\mathbf{p}}.$$

We claim that $\mathbf{p}^1 \in P$. For any $x \in X$,

$$\begin{aligned} p^1(x) &= \frac{1}{1-\alpha}p^2(x) - \frac{\alpha}{1-\alpha}\tilde{p}(x) \\ &\geq \frac{1}{1-\alpha}\tilde{\ell}(x) - \frac{\alpha}{1-\alpha}\tilde{p}(x) \\ &= \frac{1}{1-\alpha}\left[(1-\alpha)\ell(x) + \alpha\tilde{p}(x)\right] - \frac{\alpha}{1-\alpha}\tilde{p}(x) \\ &= \ell(x). \end{aligned}$$

A similar argument can be used to establish that $p^1(x) \leq u(x)$ for each $x \in X$.

Since $\mathbf{p}^1 \in P$ and $P = \operatorname{conv}\{\mathbf{q}^1, \ldots, \mathbf{q}^M\}$, there exist $\lambda_1, \ldots, \lambda_M \geq 0$ such that $\sum_{j=1}^M \lambda_j = 1$ and

$$\mathbf{p}^1 = \sum_{j=1}^M \lambda_j \mathbf{q}^j.$$

If we now form the sum $\sum_{j=1}^M \lambda_j \tilde{\mathbf{q}}^j$, we see that

$$\begin{aligned}
\sum_{j=1}^M \lambda_j \tilde{\mathbf{q}}^j &= \sum_{j=1}^M \lambda_j \left[(1-\alpha)\mathbf{q}^j + \alpha\tilde{\mathbf{p}}\right] \\
&= (1-\alpha)\mathbf{p}^1 + \alpha\tilde{\mathbf{p}} \\
&= (1-\alpha)\left[\frac{1}{1-\alpha}\mathbf{p}^2 - \frac{\alpha}{1-\alpha}\tilde{\mathbf{p}}\right] + \alpha\tilde{\mathbf{p}} \\
&= \mathbf{p}^2;
\end{aligned}$$

in other words, $\mathbf{p}^2 \in \operatorname{conv}\{\tilde{\mathbf{q}}^1, \ldots, \tilde{\mathbf{q}}^M\}$.

To prove that $\operatorname{conv}\{\tilde{\mathbf{q}}^1, \ldots, \tilde{\mathbf{q}}^M\} \subseteq \tilde{P}$, note that since $P = \operatorname{conv}\{\mathbf{q}^1, \ldots, \mathbf{q}^M\}$, each $\mathbf{q}^j \in P$. Then, for any $j \in \{1, \ldots, M\}$ and any component $x \in X$,

$$\tilde{q}^j(x) = (1-\alpha)q^j(x) + \alpha\tilde{p}(x) \geq (1-\alpha)\ell(x) + \alpha\tilde{p}(x) = \tilde{\ell}(x)$$

and

$$\tilde{q}^j(x) = (1-\alpha)q^j(x) + \alpha\tilde{p}(x) \leq (1-\alpha)u(x) + \alpha\tilde{p}(x) = \tilde{u}(x).$$

As a result, each $\tilde{\mathbf{q}}^j$ belongs to $\tilde{P}$. Since $\tilde{P}$ is a convex set, it will contain any convex combination of $\tilde{\mathbf{q}}^1, \ldots, \tilde{\mathbf{q}}^M$. It therefore follows that $\operatorname{conv}\{\tilde{\mathbf{q}}^1, \ldots, \tilde{\mathbf{q}}^M\} \subseteq \tilde{P}$, and thus that $\tilde{P} = \operatorname{conv}\{\tilde{\mathbf{q}}^1, \ldots, \tilde{\mathbf{q}}^M\}$. ■

We use Lemma 2 in the proof of Proposition 5.

Proof of Proposition 5: The first lower and upper bound vectors $\boldsymbol{\ell}^1$ and $\mathbf{u}^1$ correspond to the uncertainty set P^1, which is a bounded polyhedron. As such, it can be written as the convex hull of finitely many PMFs $\mathbf{q}^{1,1}, \ldots, \mathbf{q}^{M,1}$. Since the update algorithm is convex-convergent, for every $i \in \mathbb{Z}_+$ there is an α_i such that

$$\begin{aligned}
\boldsymbol{\ell}^{i+1} &= (1-\alpha_i)\boldsymbol{\ell}^i + \alpha_i\mathbf{p}^i, \\
\mathbf{u}^{i+1} &= (1-\alpha_i)\mathbf{u}^i + \alpha_i\mathbf{p}^i.
\end{aligned}$$

For each $j \in \{1, \ldots, M\}$ and each $i \in \mathbb{Z}_+$, define $\mathbf{q}^{j,i+1}$ recursively as

$$\mathbf{q}^{j,i+1} = (1-\alpha_i)\mathbf{q}^{j,i} + \alpha_i\mathbf{p}^i.$$

Let P^i denote the uncertainty set specified by $\boldsymbol{\ell}^i$ and $\mathbf{u}^i$. By repeatedly applying Lemma 2, we can see that $P^i = \operatorname{conv}\{\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i}\}$. By Lemma 1, problem (1) with the uncertainty set P^i and problem (2) with defining PMFs $\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i}$ are equivalent and have the same feasible region. Therefore, it follows that

$$\mathbf{w}^*(\boldsymbol{\ell}^i, \mathbf{u}^i) = M(Z \mid H(\mathbf{q}^{1,i}, \ldots, \mathbf{q}^{M,i})).$$

Also, as we noted immediately after Proposition 3, the nominal problem with respect to $\mathbf{p}^*$ is equivalent to the alternative robust problem (2) with all M defining PMFs set to $\mathbf{p}^*$; we therefore have that

$$\mathbf{w}^*(\mathbf{p}^*) = M(Z \mid H(\mathbf{p}^*, \ldots, \mathbf{p}^*)).$$

By the definition of the sequences $(\mathbf{q}^{j,i})_{i=1}^\infty$ for each $j \in \{1, \ldots, M\}$ and by the convergence property of convex-convergent update algorithms, $\mathbf{q}^{j,i} \to \mathbf{p}^*$ for each $j \in \{1, \ldots, M\}$ as $i \to \infty$ (each $(\mathbf{q}^{j,i})_{i=1}^\infty$ sequence can be treated as a lower or upper bound vector sequence in the definition of a convex-convergent update algorithm). Applying Proposition 4 with any $\epsilon > 0$, we obtain an $N \in \mathbb{Z}_+$ such that for every $i > N$,

$$\mathbf{w}^*(\boldsymbol{\ell}^i, \mathbf{u}^i) \subseteq U(\mathbf{w}^*(\mathbf{p}^*), \epsilon),$$

as required. ■

D.4 Convergence of dose distributions under convex-convergent algorithms and for prescient solutions

Before we can prove Theorem 1 (the dose distribution from the adaptive robust method approaches the set **D**), we need two additional lemmata. The proofs of these lemmata involve straightforward algebraic manipulation and are omitted.

Lemma 3 *Let $\mathbf{p}^i \in \mathcal{P}$ and $\mathbf{w}, \mathbf{y} \in \mathbb{R}^{|\mathcal{B}|}$. Then*

$$\|\Delta\mathbf{p}^i(\mathbf{w}-\mathbf{y})\| \leq |\mathcal{V}| \cdot \max \Delta_{v',x',b'} \cdot \|\mathbf{w}-\mathbf{y}\|,$$

where the maximum is taken over all voxels $v' \in \mathcal{V}$, motion states $x' \in X$ and beamlets $b' \in \mathcal{B}$.

Lemma 4 *Let $\mathbf{p}, \mathbf{q} \in \mathcal{P}$ and $\mathbf{w} \in \mathbb{R}^{|\mathcal{B}|}$, with each $w_b \geq 0$. Then*

$$\|\Delta(\mathbf{p}-\mathbf{q})\mathbf{w}\| \leq \max_{x' \in X} \sum_{v \in \mathcal{V}} \sum_{b \in \mathcal{B}} \Delta_{v,x',b} w_b \cdot \|\mathbf{p}-\mathbf{q}\|.$$

We now prove Theorem 1.

Proof of Theorem 1: Let $\epsilon > 0$ be given. The set of optimal solutions of problem (1) with $P = \{\mathbf{p}^*\}$, denoted $\mathbf{w}^*(\mathbf{p}^*)$, is closed (it is a polyhedron) and is bounded by Proposition 2. Since $\mathbf{w}^*(\mathbf{p}^*) \subseteq \mathbb{R}^{|\mathcal{B}|}$, $\mathbf{w}^*(\mathbf{p}^*)$ is compact.

The function $Y : \mathbb{R}^{|\mathcal{B}|} \to \mathbb{R}$ defined as

$$Y(\mathbf{w}) = \max_{x' \in X} \sum_{v \in \mathcal{V}} \sum_{b \in \mathcal{B}} \Delta_{v,x',b} w_b$$

is a continuous function, so by the extreme value theorem, it attains a maximum on $\mathbf{w}^*(\mathbf{p}^*)$. Let $\bar{Y}$ denote that maximum value.

Since $\mathbf{p}^i \to \mathbf{p}^*$ as $i \to \infty$, there must exist an $N_1 \in \mathbb{Z}_+$ such that, for $i > N_1$,

$$\|\mathbf{p}^i - \mathbf{p}^*\| < \frac{\epsilon}{3 \cdot \bar{Y}}.$$

By Proposition 5, there must also exist an $N_2 \in \mathbb{Z}_+$ such that, for $i > N_2$,

$$\mathbf{w}^*(\boldsymbol{\ell}^i, \mathbf{u}^i) \subseteq U\left(\mathbf{w}^*(\mathbf{p}^*), \frac{\epsilon}{3 \cdot |\mathcal{V}| \cdot \max \Delta_{v',x',b'}}\right).$$

Set $N_3 = \max\{N_1, N_2\}$.

Define the functions $\Omega_i : \mathbb{R}^{|\mathcal{B}|} \to \mathbb{R}$ for each $i \leq N_3$ as

$$\Omega_i(\mathbf{w}) = \|\Delta\mathbf{p}^i\mathbf{w}^i - \Delta\mathbf{p}^*\mathbf{w}\|.$$

Recall that for each i, $\mathbf{w}^i \in \mathbf{w}^*(\boldsymbol{\ell}^i, \mathbf{u}^i)$ is the optimized intensity vector determined by the adaptive robust method for fraction i and $\mathbf{w}^i/n$ is delivered to the patient in fraction i. Like Y, each Ω_i is also a continuous function, and since the set $\mathbf{w}^*(\mathbf{p}^*)$ is compact, each Ω_i attains a maximum on $\mathbf{w}^*(\mathbf{p}^*)$. Let $\bar{\Omega}_i$ denote each of those maximum values. Since $1/n$ is a decreasing function on $\mathbb{Z}_+$, there must be an $N_4 \in \mathbb{Z}_+$, $N_4 > N_3$, such that, for $n > N_4$,

$$\frac{1}{n}\sum_{i=1}^{N_3} \bar{\Omega}_i < \frac{\epsilon}{3}.$$

Now, let $n > N_4$. To prove our result, we will construct a dose vector inside **D** whose epsilon neighborhood contains $1/n \cdot \sum_{i=1}^{n} \Delta\mathbf{p}^i\mathbf{w}^i$.

For each $i > N_3$, select $\mathbf{y}^i \in \mathbf{w}^*(\mathbf{p}^*)$ so that

$$\|\mathbf{w}^i - \mathbf{y}^i\| < \frac{\epsilon}{3 \cdot |\mathcal{V}| \cdot \max \Delta_{v',x',b'}}.$$

Such a $\mathbf{y}^i$ exists because $\mathbf{w}^i$ is contained inside the union of all open balls of size $\epsilon/\left(3 \cdot |\mathcal{V}| \cdot \max \Delta_{v',x',b'}\right)$ centered at points in $\mathbf{w}^*(\mathbf{p}^*)$, so there must be at least one $\mathbf{y}^i \in \mathbf{w}^*(\mathbf{p}^*)$ such that

$$\mathbf{w}^i \in B\left(\mathbf{y}^i, \frac{\epsilon}{3 \cdot |\mathcal{V}| \cdot \max \Delta_{v',x',b'}}\right).$$

For $i \leq N_3$, select any $\mathbf{y}^i \in \mathbf{w}^*(\mathbf{p}^*)$.

Let $\mathbf{y} = 1/n \cdot \sum_{i=1}^{n} \mathbf{y}^i$. The beamlet intensity vector $\mathbf{y}$ belongs to $\mathbf{w}^*(\mathbf{p}^*)$ because it is the convex combination of the beamlet intensity vectors $\mathbf{y}^1, \ldots, \mathbf{y}^n$, which are all elements of the convex set $\mathbf{w}^*(\mathbf{p}^*)$. Since $\mathbf{y} \in \mathbf{w}^*(\mathbf{p}^*)$, the dose distribution $\mathbf{d}$ defined as

$$\mathbf{d} = \Delta\mathbf{p}^*\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} \Delta\mathbf{p}^*\mathbf{y}^i$$

belongs to $\mathbf{D}$. We will now verify that the epsilon ball of $\mathbf{d}$ indeed contains $1/n \cdot \sum_{i=1}^{n} \Delta\mathbf{p}^i\mathbf{w}^i$. We have

$$\begin{aligned}
\left\| \frac{1}{n}\sum_{i=1}^{n} \Delta\mathbf{p}^i\mathbf{w}^i - \mathbf{d} \right\| &= \left\| \frac{1}{n}\sum_{i=1}^{n} \Delta\mathbf{p}^i\mathbf{w}^i - \frac{1}{n}\sum_{i=1}^{n} \Delta\mathbf{p}^*\mathbf{y}^i \right\| \\
&\leq \frac{1}{n}\sum_{i=1}^{N_3} \left\| \Delta\mathbf{p}^i\mathbf{w}^i - \Delta\mathbf{p}^*\mathbf{y}^i \right\| + \frac{1}{n}\sum_{i=N_3+1}^{n} \left\| \Delta\mathbf{p}^i\mathbf{w}^i - \Delta\mathbf{p}^i\mathbf{y}^i \right\| \\
&\quad + \frac{1}{n}\sum_{i=N_3+1}^{n} \left\| \Delta\mathbf{p}^i\mathbf{y}^i - \Delta\mathbf{p}^*\mathbf{y}^i \right\|,
\end{aligned}$$

where the inequality follows by the triangle inequality. By Lemmata 3 and 4 and by our definitions of $\bar{Y}$ and $\bar{\Omega}_i$, we have

$$\begin{aligned}
\left\| \frac{1}{n}\sum_{i=1}^{n} \Delta\mathbf{p}^i\mathbf{w}^i - \mathbf{d} \right\| &\leq \frac{1}{n}\sum_{i=1}^{N_3} \bar{\Omega}_i + \frac{1}{n}\sum_{i=N_3+1}^{n} |\mathcal{V}| \cdot \max \Delta_{v',x',b'} \cdot \|\mathbf{w}^i - \mathbf{y}^i\| \\
&\quad + \frac{1}{n}\sum_{i=N_3+1}^{n} \max_{x' \in X} \sum_{v \in \mathcal{V}} \sum_{b \in \mathcal{B}} \Delta_{v,x',b} y_b^i \cdot \|\mathbf{p}^i - \mathbf{p}^*\| \\
&\leq \frac{1}{n}\sum_{i=1}^{N_3} \bar{\Omega}_i + \frac{1}{n}\sum_{i=N_3+1}^{n} |\mathcal{V}| \cdot \max \Delta_{v',x',b'} \cdot \|\mathbf{w}^i - \mathbf{y}^i\| + \frac{1}{n}\sum_{i=N_3+1}^{n} \bar{Y} \cdot \|\mathbf{p}^i - \mathbf{p}^*\|.
\end{aligned}$$

Since $n > N_4$, we now have that

$$\begin{aligned}
\left\| \frac{1}{n}\sum_{i=1}^{n} \Delta\mathbf{p}^i\mathbf{w}^i - \mathbf{d} \right\| &\leq \frac{1}{n}\sum_{i=1}^{N_3} \bar{\Omega}_i + \frac{1}{n}\sum_{i=N_3+1}^{n} |\mathcal{V}| \cdot \max \Delta_{v',x',b'} \cdot \|\mathbf{w}^i - \mathbf{y}^i\| + \frac{1}{n}\sum_{i=N_3+1}^{n} \bar{Y} \cdot \|\mathbf{p}^i - \mathbf{p}^*\| \\
&< \frac{\epsilon}{3} + \frac{(n - N_3)}{n} \cdot \frac{\epsilon}{3} + \frac{(n - N_3)}{n} \cdot \frac{\epsilon}{3} \\
&\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\
&= \epsilon.
\end{aligned}$$

We have just shown that $1/n \cdot \sum_{i=1}^{n} \Delta\mathbf{p}^i\mathbf{w}^i \in B(\mathbf{d}, \epsilon)$. Since $B(\mathbf{d}, \epsilon) \subseteq U(\mathbf{D}, \epsilon)$, we have that

$$\frac{1}{n}\sum_{i=1}^{n} \Delta\mathbf{p}^i\mathbf{w}^i \in U(\mathbf{D}, \epsilon),$$

which proves the theorem. ■

Lastly, we prove the dose distribution convergence results for the daily and average prescient methods. For the purposes of this proof, we use $H(\mathbf{q})$ to denote the feasible region of the alternative robust problem (2) with only one defining PMF $\mathbf{q}$. In this case, $H(\mathbf{q})$ corresponds to the feasible region of the nominal problem with respect to $\mathbf{q}$.

Proof of Theorem 2: (a) (Daily prescient dose convergence.) First, observe that the robust problem (1) with $\boldsymbol{\ell} = \mathbf{p}^i$ and $\mathbf{u} = \mathbf{p}^i$ contains only one PMF: $\mathbf{p}^i$. Therefore, the robust problem (1) with $\boldsymbol{\ell} = \mathbf{u} = \mathbf{p}^i$ is equivalent to the alternative robust problem (2) with $\mathbf{p}^i$ as its one and only defining PMF. For every $i \in \mathbb{Z}_+$, it then follows that

$$\mathbf{w}^*(\mathbf{p}^i, \mathbf{p}^i) = M(Z \,|\, H(\mathbf{p}^i)).$$

The feasible region of the nominal problem with respect to $\mathbf{p}^*$ is equal to $H(\mathbf{p}^*)$, and we have that

$$\mathbf{w}^*(\mathbf{p}^*) = M(Z \,|\, H(\mathbf{p}^*)).$$

The PMF sequence $(\mathbf{p}^i)_{i=1}^{\infty}$ converges to $\mathbf{p}^*$, so we can invoke Proposition 4. Doing so we obtain that for every $\epsilon > 0$, there exists an $N \in \mathbb{Z}_+$ such that for all $i > N$,

$$M(Z \,|\, H(\mathbf{p}^i)) \subseteq U\left(M(Z \,|\, H(\mathbf{p}^*)), \epsilon\right),$$

or equivalently,

$$\mathbf{w}^*(\mathbf{p}^i, \mathbf{p}^i) \subseteq U\left(\mathbf{w}^*(\mathbf{p}^*), \epsilon\right).$$

From this point forward, the proof proceeds in exactly the same way as the proof of Theorem 1.

(b) (Average prescient dose convergence.) Just as in the proof of part (a), the robust problem (1) with $\boldsymbol{\ell} = \mathbf{u} = 1/n \cdot \sum_{i=1}^{n} \mathbf{p}^i$ is equal to the alternative robust problem with $1/n \cdot \sum_{i=1}^{n} \mathbf{p}^i$ as its one and only defining PMF. Analogously to that proof, we have that

$$\mathbf{w}^*\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{p}^i, \frac{1}{n}\sum_{i=1}^{n}\mathbf{p}^i\right) = M\left(Z \,\middle|\, H\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{p}^i\right)\right),$$

and that

$$\mathbf{w}^*(\mathbf{p}^*) = M(Z \,|\, H(\mathbf{p}^*)).$$

The sequence of averages $(1/n \cdot \sum_{i=1}^{n} \mathbf{p}^i)_{n=1}^{\infty}$ converges to $\mathbf{p}^*$, which shows that $1/n \cdot \sum_{i=1}^{n} p^i(x) \to p^*(x)$ as $n \to \infty$ if $p^i(x) \to p^*(x)$ as $i \to \infty$ for every $x \in X$).

Since $1/n \cdot \sum_{i=1}^{n} \mathbf{p}^i \to \mathbf{p}^*$ as $n \to \infty$, we can invoke Proposition 4. Doing so we obtain that for every $\epsilon > 0$, there exists an $N \in \mathbb{Z}_+$ such that for all $n > N$,

$$M\left(Z \,\middle|\, H\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{p}^i\right)\right) \subseteq U\left(M(Z \,|\, H(\mathbf{p}^*)), \epsilon\right),$$

or equivalently,

$$\mathbf{w}^*\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{p}^i, \frac{1}{n}\sum_{i=1}^{n}\mathbf{p}^i\right) \subseteq U\left(\mathbf{w}^*(\mathbf{p}^*), \epsilon\right).$$

From this point forward, the proof proceeds in the same way as the proof of Theorem 1 with suitable modifications. ■

References

Ben-Tal, A., & Nemirovski, A. (1999). Robust solutions of uncertain linear programs. *Oper. Res. Lett.*, *25*, 1–14.

Bortfeld, T., Chan, T., Trofimov, A., & Tsitsiklis, J. (2008). Robust management of motion uncertainty in intensity-modulated radiation therapy. *Oper. Res.*, *56*, 1461–1473.

Dantzig, G., Folkman, J., & Shapiro, N. (1967). On the continuity of the minimum set of a continuous function. *J. Math. Anal. Appl.*, *17*, 519–548.

Gierga, D., Brewer, J., Sharp, G., Betke, M., Willett, C., & Chen, G. (2005). The correlation between internal and external markers for abdominal tumors: implications for respiratory gating. *Int. J. Radiat. Oncol. Biol. Phys.*, *61*, 1551–1558.

Lujan, A., Larsen, E., Balter, J., & Ten Haken, R. (1999). A method for incorporating organ motion due to breathing into 3D dose calculations. *Med. Phys.*, *26*, 715–720.

Sonke, J., Zijp, L., Remeijer, P., & van Herk, M. (2005). Respiratory correlated cone beam CT. *Med. Phys.*, *32*, 1176–1186.

Tsunashima, Y., Sakae, T., Shioyama, Y., Kagei, K., Terunuma, T., Nohtomi, A., & Akine, Y. (2004). Correlation between the respiratory waveform measured using a respiratory sensor and 3d tumor motion in gated radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, *60*, 951–958.