Minimizing beam-on time during radiotherapy treatment by total-variation regularization

Xing Lab, Department of Radiation Oncology, Stanford.


Rasmus Bokrantz
Outline

1. The trade-off between beam-on time and plan complexity.
2. Approaches for handling constraints on deliverability.
3. Treatment planning viewed as a signal processing problem.
4. Recovery of sparse signals.
5. Comparison between regularization approaches for a patient case.
Plan complexity vs. beam-on time

- Increased number of segments allows greater flexibility in shaping the dose distribution.
- Few and regular segments promotes robustness with respect to geometric uncertainties (e.g. organ motion).
- A decrease in number of segments reduces the risk of radiation-induced secondary cancers.

**Overall planning goal**
Minimize the number of segments while satisfying the clinical goals.
The treatment planning problem

Perform a least squares fit to prescription dose

\[ \text{minimize} \quad \| Px - \hat{d} \|_2^2 \]
\[ \text{subject to} \quad x \geq 0, \]
\[ \quad x \in \{ \text{deliverable apertures} \} \]

where
\[ x \quad \text{beamlet weights} \]
\[ P \quad \text{beamlet kernel matrix} \]
\[ \hat{d} \quad \text{prescription dose} \]
Approaches for handling the aperture constraint

Beamlet weight optimization

- The aperture constraint is ignored.
  + Convex quadratic program.
  - Regularization is needed (e.g. quadratic smoothing).
  - Requires conversion into deliverable machine settings.

Direct machine parameter optimization

- The aperture constraint is formulated explicitly.
  + The optimized plan is directly deliverable.
  - The aperture constraint is highly nonlinear and nonconvex.
Novel approach suggested by Zhu et al.

View the treatment planning as a signal processing problem

- Optimal beam profiles are viewed as a “signal” to be detected.
- Deliverable fluence maps are piecewise constant.
- The gradient of a piecewise constant fluence map is sparse.
- Fluence maps can be constructed using *compressed sensing*: a technique for recovery of sparse signals.
A signal is said to be sparse if a large number of coefficient are close or equal to zero when represented in some domain.

\[ f = \sum_{i=1}^{n} x_i \psi_i, \quad \text{where} \quad f \in \mathbb{R}^n, \quad (\text{signal}) \]

\[ \Psi = (\psi_1, \psi_2, \ldots, \psi_n). \quad (\text{ON basis}) \]
Nyquist-Shannon sampling theorem
Perfect signal reconstruction requires a sampling frequency at least twice the maximum frequency present in the signal.
Compressed sensing
A sparse signal can be reconstructed at a significantly lower sampling rate than the Nyquist rate.

Algorithm: recover the signal by $\ell_1$-norm minimization
An $n$-dimensional signal $f$ expressed in the representation basis $\Psi$ can be reconstructed using $m$ measurements $y_k = \langle \varphi_k, \Psi x \rangle$ in the sensing basis $\Phi$ by solving the convex optimization problem

$$\min_{x \in \mathbb{R}^n} \|x\|_{\ell_1},$$
subject to $y_k = \langle \varphi_k, \Psi x \rangle$, $k = 1, \ldots, m$. 
If $f$ is sufficiently sparse, recovery via $\ell_1$ minimization is exact with overwhelming probability.

**Theorem:** Candes, E. & Romberg, J. (2007) Inverse Prob. 23(3).

Let $f \in \mathbb{R}^n$ and suppose the coefficient sequence $x$ of $f$ in the representation basis $\Psi$ is $S$-sparse. Select $m$ measures in the sensing domain $\Phi$ uniformly at random. Then if

$$m \geq C\mu^2(\Phi, \Psi)S\log(n/\delta),$$

where $C > 0$, and $\mu(\Phi, \Psi)$ is the mutual coherence between $\Psi$ and $\Phi$, then, the probability of perfect reconstruction exceeds $1 - \delta$.

$$\mu(\Phi, \Psi) \triangleq \sqrt{n} \max_{1 \leq k, j \leq n} |\langle \varphi_k, \psi_j \rangle|.$$
Noise contaminated signal

$\ell_1$-norm regularization promotes sparsity

$\ell_2$-norm regularization promotes smoothness

Reformulating the treatment planning problem

Initial problem formulation

\[
\min_{x \in \mathbb{R}^n} \frac{1}{2} ||Px - \hat{d}||^2_2
\]
subject to
\[x \geq 0,\]
\[x \in \{ \text{deliverable apertures} \}\]

Include a total-variation regularization term and drop the aperture constraint

\[
\min_{x \in \mathbb{R}^n} \frac{1}{2} ||Px - \hat{d}||^2_2 + \beta ||\nabla x||_1
\]
subject to
\[x \geq 0\]

This problem can be posed as a convex quadratic program using help variables.
Comparison between regularization approaches

Total-variation regularization

\[
\min_{x \in \mathbb{R}^n} \| Px - \hat{d} \|_2^2 + \beta \| \nabla x \|_1
\]
subject to \( x \geq 0 \)

Quadratic smoothing

\[
\min_{x \in \mathbb{R}^n} \| Px - \hat{d} \|_2^2 + \beta \| \nabla x \|_2^2
\]
subject to \( x \geq 0 \)
Without regularization

Highly irregular fluence profile which is ill-suited for conversion into deliverable machine settings.
Quadratic smoothing

Smooth fluence profile which requires a relatively large number of segments for conversion.
Total-variation regularization

Pieciewise continous fluence profile which can be converted into a small number of segments.
Finding the plan with minimal number of segments

- Re-solving at various regularization parameter values gives points on a two-dimensional Pareto surface in objective space.
- A plan can be identified as clinically satisfactory if it meets with a set of acceptance criteria.
- The authors selects the plan with smallest number of segments which also meets the acceptance criteria as the final solution.

<table>
<thead>
<tr>
<th>Regions</th>
<th>Acceptance criteria</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTV</td>
<td>%vol &gt; 66 Gy ≥ 95</td>
<td>%vol &gt; 66 Gy = 95.0</td>
</tr>
<tr>
<td>Brainstem</td>
<td>Maximum &lt; 54 Gy</td>
<td>Maximum = 46.4 Gy</td>
</tr>
<tr>
<td>Larynx</td>
<td>Maximum &lt; 70 Gy</td>
<td>Maximum = 5.60 Gy</td>
</tr>
<tr>
<td></td>
<td>Mean &lt; 26 Gy</td>
<td>Mean = 2.42 Gy</td>
</tr>
<tr>
<td>Optic chiasm/nerves</td>
<td>Maximum &lt; 54 Gy</td>
<td>Maximum = 8.90 Gy</td>
</tr>
<tr>
<td>Lens</td>
<td>Maximum &lt; 12 Gy</td>
<td>Maximum = 8.80 Gy</td>
</tr>
<tr>
<td>Left parotid</td>
<td>Maximum &lt; 70 Gy</td>
<td>Maximum = 29.1 Gy</td>
</tr>
<tr>
<td></td>
<td>Mean &lt; 26 Gy</td>
<td>Mean = 5.80 Gy</td>
</tr>
<tr>
<td>Spinal cord</td>
<td>Maximum &lt; 45 Gy</td>
<td>Maximum = 14.6 Gy</td>
</tr>
<tr>
<td>Body</td>
<td>Maximum &lt; 75.9 Gy (115%)</td>
<td>Maximum= 74.0 Gy (112.0%)</td>
</tr>
</tbody>
</table>

%vol > x Gy: percentage of the volume that receives more than x Gy dose.
Comparison to quadratic smoothing

- At a fixed number of segments, total-variation regularization yields clinically superior plans.
- At a fixed “dose distribution performance”, total-variation regularization results in a significant decrease in number of segments.

Clinical performance is assessed using dose-volume histograms.
Comparison to direct aperture optimization

- No numerical comparison is performed due to difficulties in implementing direct machine parameter optimization.
- The authors claim that the suggested method is superior since global optimality is guaranteed and the treatment planning problem can be solved more efficiently.
- Direct aperture optimization requires prefixing the number of segments before optimization.
Summary

Novel approach for beamlet weight optimization

- A sparsity objective is included in the problem formulation.
- Plan degeneracy is exploited for the benefit of reducing beam-on time.

Conclusion

- Total-variation regularization allows for a reduction in beam-on time without compromising plan quality.