Improving the Speed of Optimizing Intensity Modulated Proton Therapy Treatment Plans by Means of Redesigning the Fluence Map Optimization Based In-House Proton Therapy Center Optimizer at The University of Texas MD Anderson Cancer Center

A. Adair[1], L. Liao[1], X. Zhang[1], X. R. Zhu[1], P. Yepes[2], N. Sahoo[1], B. Gunn[1], S. Frank[1]

1. The University of Texas M. D. Anderson Cancer, Houston, Texas, USA
2. Rice University, Houston, Texas, USA

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center Proton Therapy

Making Cancer History®

JULY 12–16 2020 VIRTUAL JOINT AAPM | COMP MEETING EASTERN TIME [GMT-4]

BACKGROUND INFORMATION

The higher distal end dose and the lower entrance dose make proton therapy (PT) preferable for sparing tissue surrounding tumors than standard photon therapies. This work focuses on a PT modality called intensity modulated proton therapy (IMPT). In IMPT the proton beams are delivered with a large number of narrow scanning pencil beams (beamlets). The target volume (TV) is partitioned into various scanning spots. IMPT “paints” over each spot by controlling the energy level/layer (depth of spot to patient surface) and the intensity of the scanning magnets (lateral position of a spot). Beamlet intensity (weight) is controlled by exposure time. The longer the exposure time for a specific spot, the higher radiation dose deposited to that spot. The intensity map from all beamlets is known as the fluence map.

Radiation therapy aims to treat the TV while sparing the surrounding organs at risk (OARs). However, as TV and OARs can be overlapping or near each other, these goals may be mutually opposed. The balancing act between these objectives turns treatment planning into an optimization problem. In particular this work focuses on the fluence map optimization (FMO) problem; adjusting the beamlets intensity to ensure the treatment plan meets TV and OAR objectives. This is a non-linear problem with the amount of decision variables corresponding to the amount of beamlets (typically thousand+). The in-house Proton Therapy Center (PTC) IMPT Optimizer (OPT) at MD Anderson Cancer Center currently uses a Limited-memory BFGS (L-BFGS) (Liu and Nocedal 1989) based optimizer from the ALGLIB numerical analysis and data processing library (Sergey Bochkanov 2019) to solve this problem. The dose-based objective encoding the deviation from the TV prescription dose or OAR tolerance dose (Liao 2016). In this study we obtain our dose measurements for the OPT from the fast dose calculator (FDC) (Yepes 2016). Additionally, we include comparisons from the Varian eclipse optimizer (Eclipse 2019).

METHODOLOGY

- **Areas of focus**: CPU profiling (timing analysis), Cache profiling (fine timing analysis), and Data structures optimization (memory analysis)
- **Common methods**: Order of complexity reductions, Factor optimizations, Optimize Data structures, Fine-tune Multithreading, Aligning data to cache
- **Consistency**: Seeking consistent metrics (ex DVH (Drzymala *et al* 1991)) between OPT versions and Varian Eclipse (TPS)
- **OPT versions**: latest version (v4) is called (+):
  - V0: Original Version (some minimal modifications)
  - V1: Pre-keyed Voxel lookup in Dose Calculation and Gradient loops
  - V2: V1 + Pre-Keyed Beamlet lookup
  - V3: V2 + Eliminated Sparse Voxel Storage
  - V4: Sparse objects → Parallel dense arrays. “Cross” between Coordinate list format and Compressed Sparse Row (Buluc *et al* 2009)
- **Patients**: We present data involving two patients
  - Each case has 1 uniform (uni) aim (TV) and 1 upper (upp) aim
  - Prostate (P1) – 96780569 records, 2036 Total Beamlets
  - Lung (P2) – 359093706 records, 11023 Total Beamlets
- **CPU Architecture**:
  - OPT: 64bit, 48CPU, Intel Xeon(R) CPU E5-2690 v3 @ 2.60GHz
  - TPS: 64bit, 32CPU, Intel CPU E5-2690 v2 @ 3.00GHz

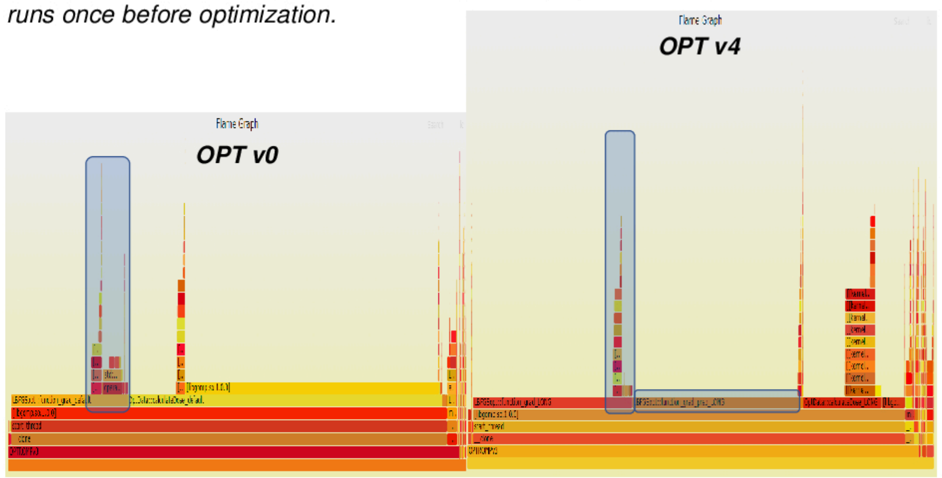
GOAL

We present work on improving the in-house Proton Therapy Center IMPT Optimizer (OPT) at MD Anderson Cancer Center. The two goals being speed improvement and improving the memory management. By decreasing the run time as well as the memory footprint, we aim to make the IMPT optimization process more conducive to real-time clinical use.

EXAMPLES AND RESULTS

Software Profiling:

The software profiling tools we used include perf, perf-c2c, and valgrind. The following CPU flame-plots (Gregg 2017) illustrate activity of the code-paths for OPT v0 and v4. Example: Time complexity was reduced by writing a gradient preparation function that runs once before optimization.



Cache analysis and false sharing detection (HITM = hit modified in a cache-line) was performed for each new OPT version using perf cache-2-cache. The current Last Level Cache (LLC) utilization for v4 is given, false sharing is not a huge problem.

	Load LLC Misses	LLC Misses to Local DRAM	LLC Misses to Remote DRAM	LLC Misses to Remote cache (HIT)	LLC Misses to Remote cache (HITM)
P1	16268	28.90%	43.10%	16.70%	11.40%
P2	94485	31.40%	44.40%	17.10%	7.10%

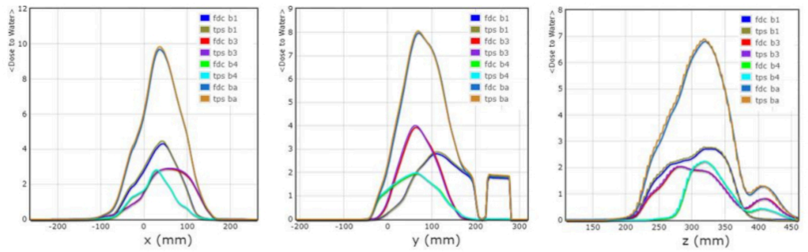
Key Measures:

A summary of key optimizer measures. OPT v4 out performs all others. This is largely do to eliminating all sparse objects.

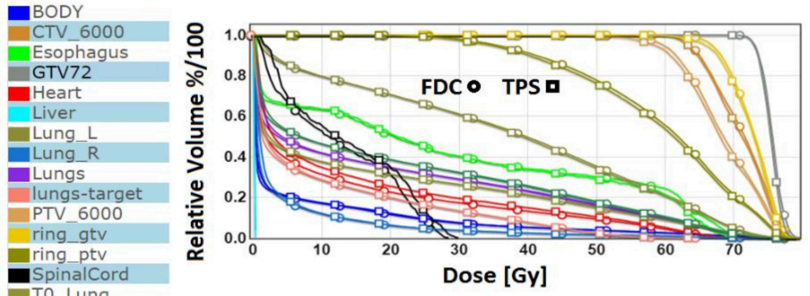
P1	Mem (GB)	Opt Time (s)	Total Time (s)	AVG uni (ms)	AVG upp (ms)
OPT v0	2.60	225.83	241.70	156.75	158.99
OPT v1	3.00	35.70	49.96	56.53	40.68
OPT v2	3.32	30.90	46.39	56.24	41.17
OPT v3	5.19	22.70	39.43	37.22	28.01
OPT v4	3.13	18.13	34.27	26.57	14.87
TPS			84.3		
P2	**Mem (GB)**	**Opt Time (m)**	**Time (m)**	**AVG uni (ms)**	**AVG upp (ms)**
OPT v0	22.67	9.46	10.41	65.01	2771.60
OPT v1	23.81	3.70	4.67	23.76	518.13
OPT v2	24.50	2.85	3.91	24.41	520.23
OPT v3	46.75	2.09	3.49	19.81	348.70
OPT v4	12.88	1.80	2.83	14.10	215.62
TPS			28.4		

Consistency Checks:

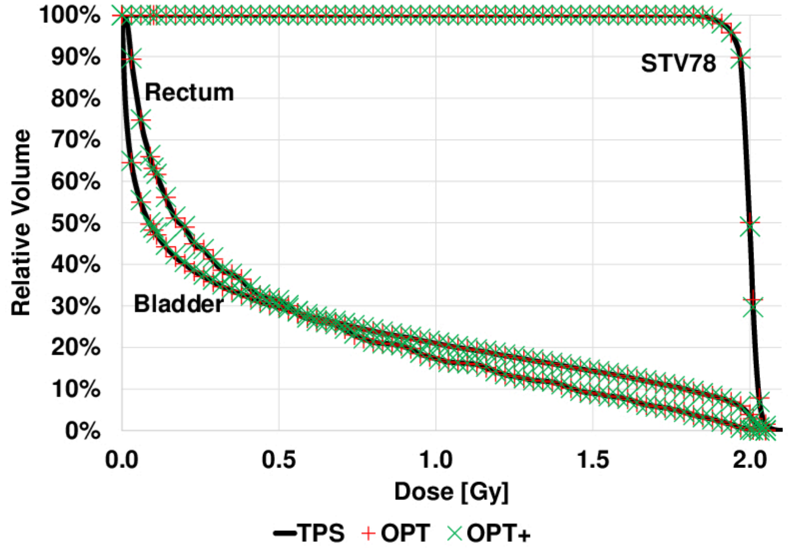
Pre-Optimization: The dose-to-water distribution is checked between the TPS and FDC MC. Example shown here is for the prostate patient.



Pre-Optimization: The differences in the corresponding DVHs are checked



Post-Optimization: Verify that versions of OPT (OPT = v0, OPT+ = v4) and Eclipse (TPS) produce consistent results. The figure below is an example of such check by means of DVH histogram comparison (P1).



CONCLUSIONS

- OPT+ (v4) was able to optimize a smaller plan P1 (2036 beamlets) in less than a minute and a larger plan P2 (11023 beamlets) in less than 3 minutes
- OPT+ processed P1 (P2) at a rate of 59.4 (64.9) beamlets per minute
- OPT+ was 2.5 (P1) to 10 (P2) times faster than eclipse
- The OPT+ false sharing rate goes from 11.4% (P1) to 7.1% (P2) decreasing with beamlet increase. This is notable but not the biggest concern.
- OPT+ is better in every key measure than all past versions. This is primarily due to redefinition of the data structures holding the voxel-beamlets information
- Consistency checks pre-optimization (Dose and DVH) and post-optimization are all positive
- OPT+ can provide fast real-time optimization, overcoming the time bottleneck provided by TPS systems like Eclipse
- A GPU version of OPT+ is currently under development
- We estimate that with a GPU version, Robust LET and RBE optimization for even the most complicated cases could be done in the near future within five minutes

REFERENCES

- Liu D. and Nocedal J., 1989, *“On the limited memory bfgs method for large scale Optimization”*, Mathematical programming, 45(1-3):503–528
- Sergey Bochkanov, 2019, *“ALGLIB (www.alglib.net)”*
- Liao L., 2016, *“Intensity Modulated Proton Therapy Optimization Under Uncertainty: Field Misalignment and Internal Organ Motion”*, PhD Dissertation, University of Houston, Houston
- Yepes P., Eley J., Liu A., Mirkovic D., Randeniya S., Titt U., and Mohan R., 2016, *“Validation of a track repeating algorithm for intensity modulated proton therapy: clinical cases study”*, Phys. Med. Bio. 61 2633-45
- Eclipse, 2019, *“Eclipse treatment planning system (TPS)”*, Varian Medical Systems, Palo Alto, CA
- Buluc A., Fineman J.T., Frigo M., Gilbert J. R., Leiserson, C. E. (2009). “*Parallel sparse matrix-vector and matrix-transpose-vector multiplication using compressed sparse blocks”,* ACM Symp. on Parallelism in Algorithms and Architectures.
- Drzymala R.E., Mohan R., Brewster L., Chu J., Goitein M., Harms W., and M. Urie M., 1991, *“Dose-volume histograms”*, International Journal of Radiation Oncology* Biology* Physics, 21(1):71–78
- Gregg B., 2017, *“Visualizing Performance with Flame Graphs”*, USENIX Association

ACKNOWLEDGEMENTS

This work is supported by the Proton Therapy Center at the University of Texas MD Anderson Cancer Center

CONTACT INFORMATION

Antony Adair, ahadair@mdanderson.org