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

Comparison of Pareto Surface Interpolations for Multi-Criterial Optimization Algorithm Evaluation

P.J. Jensen [1,2], J. Zhang [1], and Q.J. Wu[1,2]
[1]Duke Cancer Institute, Durham, NC
[2]Duke University, Durham, NC

DukeHealth

INTRODUCTION

Multi-criteria optimization (MCO) is a useful tool which assists the treatment planning process in radiation therapy. MCO allows the treatment planner to explore a given patient's feasible dosimetric trade-offs. MCO has been studied extensively and many methods for exactly generating the range of dosimetric trade-offs (the Pareto surface) have been implemented for radiation therapy. Since most MCO algorithms have historically operated by iteratively, inversely optimizing the underlying MCO problem to exactly sample the Pareto surface without error, there has been little research into Pareto surface similarity metrics. Instead, most previous Pareto surface comparison research has focused on evaluating which of the Pareto surfaces is superior [1-9]. This type of evaluation is not as useful for determining the similarity between the Pareto surfaces, which is more relevant when comparing fundamentally different MCO generation algorithms, such as emerging deep learning prediction methods.

AIM

To develop and compare several Pareto surface interpolation similarity metrics using theoretical Pareto surface examples as well as a Pareto surface generated from a previously developed prostate MCO dose prediction model.

METHODS

For all metrics studied here, the Pareto surfaces for comparison are treated as simplicial complexes, with each vertex of each simplex defined by the values of each dosimetric objective within the vertex's corresponding plan. To improve metric robustness, each simplex is upsampled in its barycentric coordinates, where the number of samples taken per coordinate needs to be determined based on experimental convergence rates. Three Pareto surface metrics are considered in this study:

1. Root-mean-square error (RMSE):
   - $RMSE(X,Y) = \sqrt{\frac{\sum_{i=1}^{s}\left\|\vec{x}_i-\vec{y}_i\right\|_2^2}{s}}$
2. Average projected distance (APD):
   - $APD(X,Y) = \frac{\sum_{i=1}^{s}\left((\vec{x}_i-\vec{y}_i)\cdot\overrightarrow{\mu_{XY}}\right)}{s\left\|\overrightarrow{\mu_{XY}}\right\|_2}$
3. Average nearest point distance (ANPD):
   - $ANPD(X,Y) = avg\left\{\underset{\vec{y}\epsilon S(Y)}{\text{avg}}\ \underset{\vec{x}\epsilon S(X)}{\inf}\left\|\vec{x}-\vec{y}\right\|_2, \underset{\vec{x}\epsilon S(X)}{\text{avg}}\ \underset{\vec{y}\epsilon S(Y)}{\inf}\left\|\vec{x}-\vec{y}\right\|_2\right\}$

where $\vec{x}_i$ is a point on Pareto surface X, $\overrightarrow{\mu_{XY}}$ is the mean displacement between surfaces X and Y, s is the number of samples taken per barycentric coordinate, and S(X) is the simplicial complex spanned by linearly interpolating the points sampling surface X. For theoretical analysis, 10,000 two-dimensional simplex pairs were generated to test many different combinations of relative size and orientation such that none of the pairs are geometrically equivalent. Additionally, to estimate the applicability to real cases, one three-dimensional case from prostate MCO dose predictions is included, with PTV $D_{95\%}$, bladder $D_{25\%}$, and rectum $D_{25\%}$ representing the MCO objective dimensions.

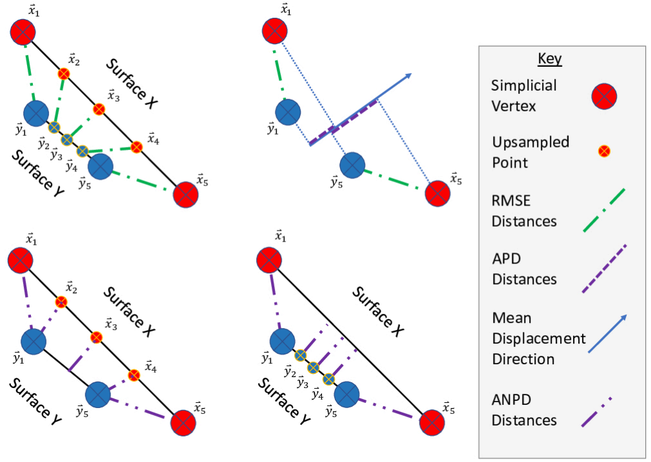

Figure 1: Depiction of the various distances which contribute to the RMSE (top left), APD (top right), and both averages in the ANPD (bottom left and bottom right), after acquiring three samples per simplex.

RESULTS

Convergence within 1% of each metric's asymptotic value is typically achieved at approximately 50 and 80 samples per barycentric dimension for the RMSE and the ANPD, respectively. Calculation requires approximately 50 milliseconds and 3 seconds to achieve convergence for the RMSE and the ANPD, respectively, while the APD always requires much less than 1 millisecond. The APD is mathematically invariant to upsampling, which heavily decreases time costs. Additionally, the APD values closely resembled the ANPD limits, while the RMSE limits tended to be more different.

For the two-dimensional cases, the average APD value, ANPD asymptotic value, and RMSE asymptotic value were 0.543, 0.439, and 0.668, respectively, with standard deviations of 0.258, 0.199, and 0.194, respectively. Case-by-case, relative to the ANPD asymptotic value, the average APD value and RMSE asymptotic value were 130% and 171%, with standard deviations of 62% and 88%. A two-sided paired t-test indicates that these values are significantly different ($p<10^{-6}$). These results suggest that the APD is a closer approximation to the ANPD limit than the RMSE limit, although the APD is still somewhat different from the ANPD asymptotic value. The average calculation times for the ANPD and the RMSE at their asymptotic values were 8.9 ms and 0.4 ms, respectively, with standard deviations of 8.2 ms and 2 ms and maxima of 62.0 ms and 15.9 ms, respectively. The APD calculation times were always below 0.1 ms.

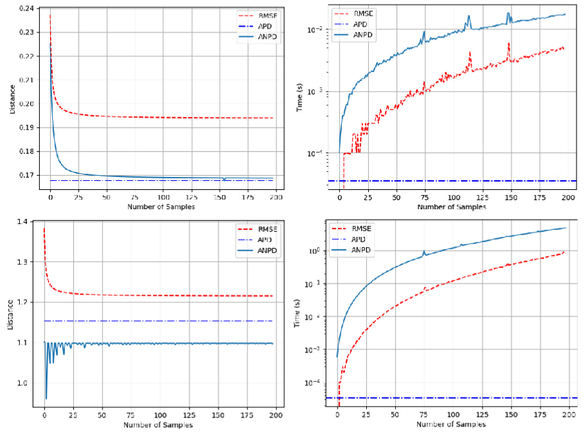

Figure 2: Sample distances as a function of the number of intra-simplex samples per barycentric coordinate (left) and the times required to compute these distances (right) for the 2D (top) and 3D (bottom) cases.

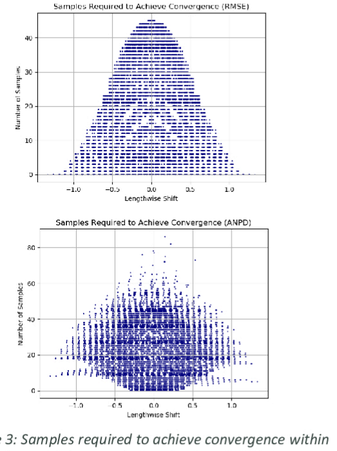

Figure 3: Samples required to achieve convergence within 1% of the metric's asymptotic value for all 2D test cases for the RMSE (top) and ANPD (bottom). Pairs are separated by their lengthwise shift relative to each other.

CONCLUSIONS

In this study, we have presented, compared, and analyzed several Pareto surface similarity metrics for use in MCO prediction assessment. Based on our analysis, we believe that the average nearest-point distance (ANPD) is the most suitable metric for Pareto surface comparison, with metric convergence being reached at approximately 80 interior samples per barycentric dimension per simplex. When possible, it is recommended to use the ANPD to compute the most accurate and appropriate distances because the ANPD formulation exactly represents the similarity between Pareto surfaces. However, the average projected distance (APD) may be more useful in time-sensitive applications due to its much higher computation speed.

REFERENCES

1. Berezkin V, Lotov A. Comparison of two Pareto frontier approximations. Computational Mathematics and Mathematical Physics. 2014;54:1402-1410.
2. Teichert K, Süss P, Serna JI, Monz M, Küfer KH, Thieke C. Comparative analysis of Pareto surfaces in multi-criteria IMRT planning. Physics in medicine and biology. 2011;56(12):3669-3684.
3. Zitzler E, Thiele L, Laumanns M, Fonseca CM, Fonseca VGd. Performance assessment of multiobjective optimizers: an analysis and review. IEEE Transactions on Evolutionary Computation. 2003;7(2):117-132.
4. Bosman PAN, Thierens D. The balance between proximity and diversity in multiobjective evolutionary algorithms. IEEE Transactions on Evolutionary Computation. 2003;7(2):174-188.
5. Veldhuizen DAV. Multiobjective evolutionary algorithms: classifications, analyses, and new innovations, Air Force Institute of Technology; 1999.
6. Grosan C, Oltean M, Dumitrescu D. Performance metrics for multiobjective optimization evolutionary algorithms. 2003.
7. Lizárraga G, Hernandez-Aguirre A, Botello S. G-Metric: an M-ary quality indicator for the evaluation of non-dominated sets. 2008.
8. Zitzler E, Thiele L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Transactions on Evolutionary Computation. 1999;3(4):257-271.
9. Li M, Yang S, Liu X. A Performance Comparison Indicator for Pareto Front Approximations in Many-Objective Optimization. 2015.

ACKNOWLEDGEMENTS

This work was supported by NIH grant R01CA201212.

CONTACT INFORMATION

Corresponding Author: Q. Jackie Wu, PhD, Department of Radiation Oncology, DUMC Box 3295, Duke University Medical Center, Durham, NC 27710 USA. Phone: +1-919-618-1841; E-mail: jackie.wu@duke.edu