Adaptive Sampling with Topological Scores

Dan Maljovec[1], Bei Wang[1], Ana Kupresanin[2], Gardar Johannesson[2], Valerio Pascucci[1], and Peer-Timo Bremer[2]

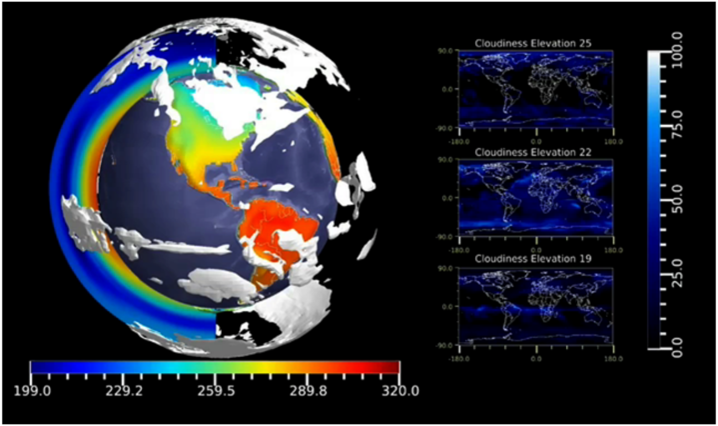
[1]SCI Institute, University of Utah
[2]Lawrence Livermore National Laboratory

Oct 24, 2011

Motivation

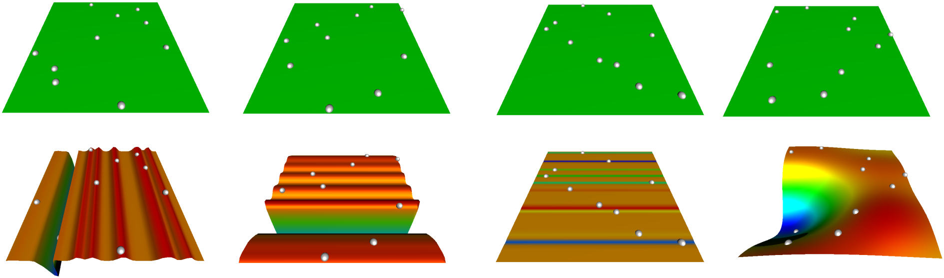
Model or simulate a phenomenom where a certain level of uncertainty exists.

- Weather and climate data
- Building, plant, and automotive design
- Socio-economic conditions/trends

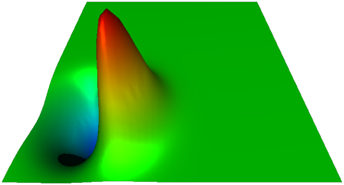


Choosing the "Right" Points

Understanding of a simulation depends heavily on where we query.

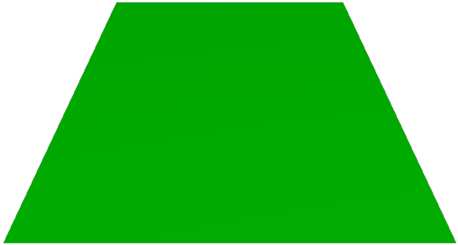


True Response:



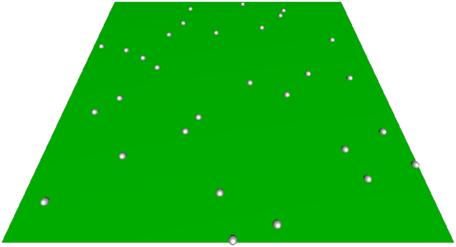
Space-Filling Sampling

No prior knowledge of the dataset:

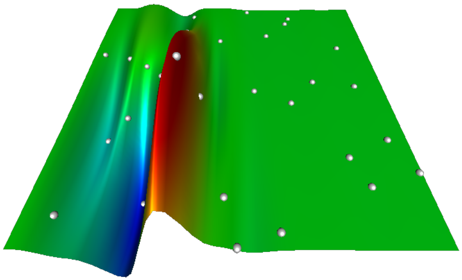


Where should we sample the model?

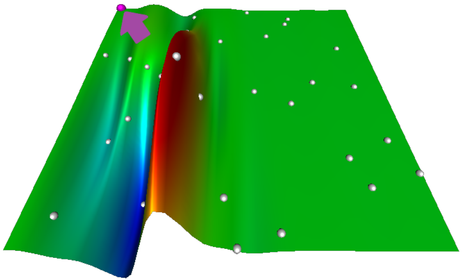
Space-filling



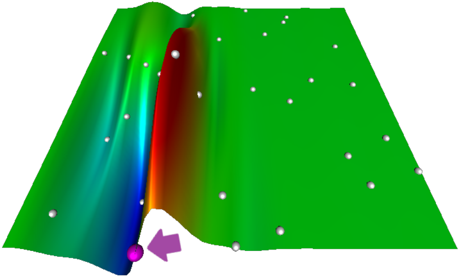
Space-Filling Sampling



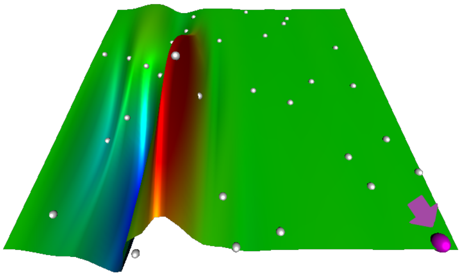
Space-Filling Sampling



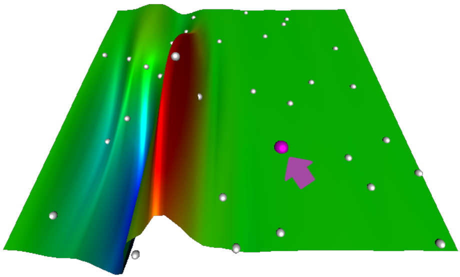
Space-Filling Sampling



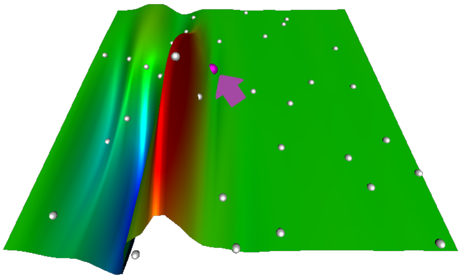
Space-Filling Sampling



Space-Filling Sampling

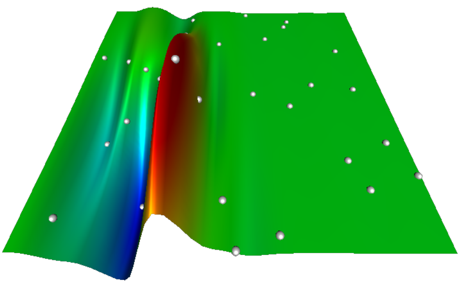


Space-Filling Sampling

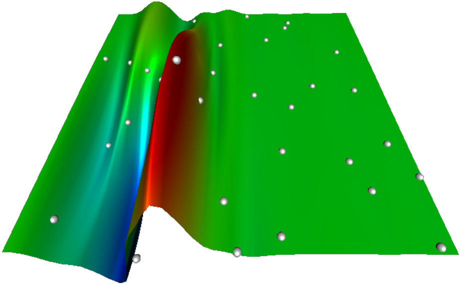


Space-Filling Sampling

What have we learned from the addition of 5 points? Not much



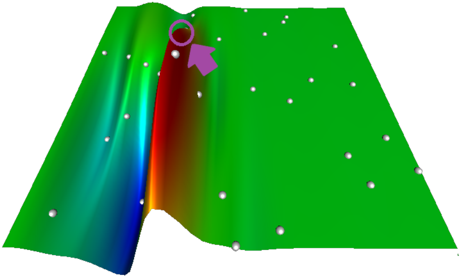
Initial fit



Model refit after adding 5 points

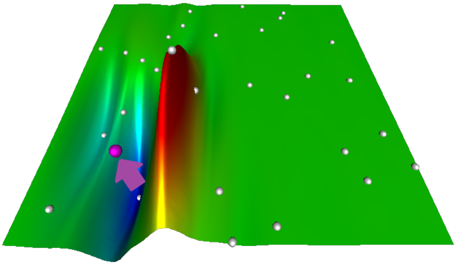
Topologically-inspired Adaptive Sampling

Sample adaptively by "learning" the model



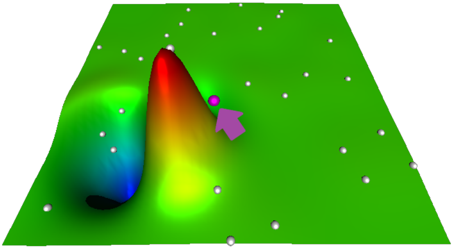
Topologically-inspired Adaptive Sampling

Sample adaptively by "learning" the model



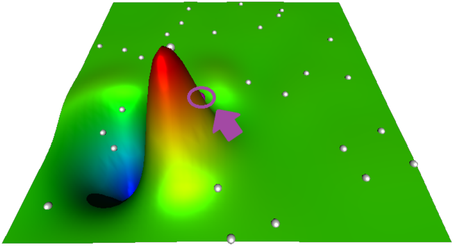
Topologically-inspired Adaptive Sampling

Sample adaptively by "learning" the model



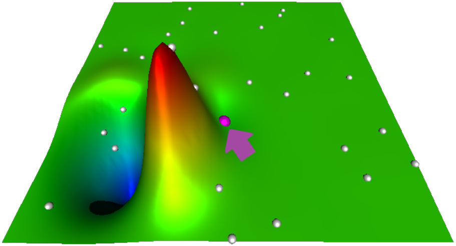
Topologically-inspired Adaptive Sampling

Sample adaptively by "learning" the model

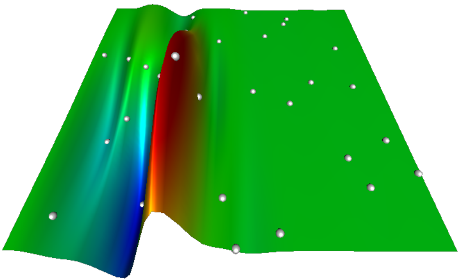


Topologically-inspired Adaptive Sampling

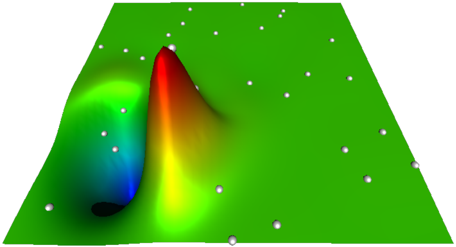
Sample adaptively by "learning" the model



Topologically-inspired Adaptive Sampling

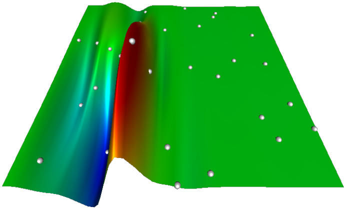


Initial fit

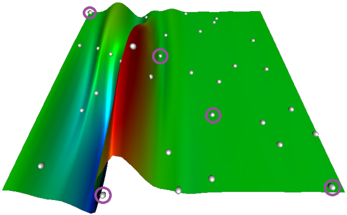


Model refit after adding 5 points

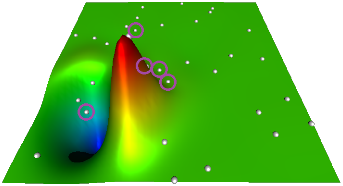
Comparison: Space-Filling Sampling vs. Adaptive Sampling



Initial Predicting Model

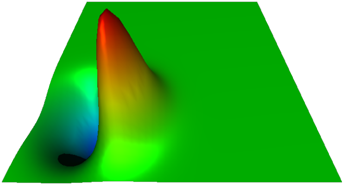


Space-filling Sampling

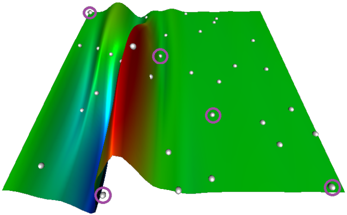


Adaptive Sampling

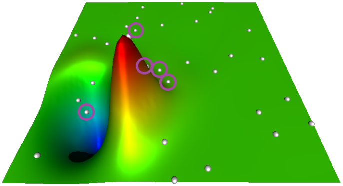
Comparison: Space-Filling Sampling vs. Adaptive Sampling



True Function Response



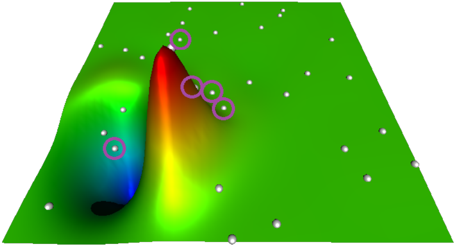
Space-filling Sampling



Adaptive Sampling

Why Topological Sampling

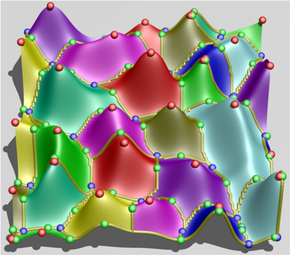
The points selected were in topologically significant regions



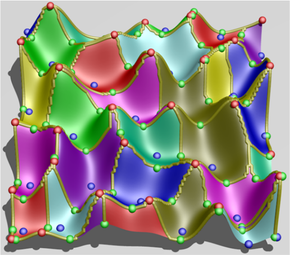
How can we measure topological impact?

Morse-Smale Complex

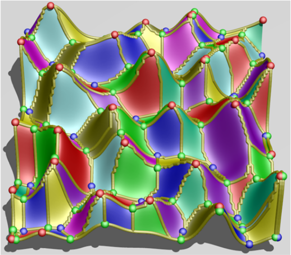
A partition of the data into monotonic regions



stable manifolds



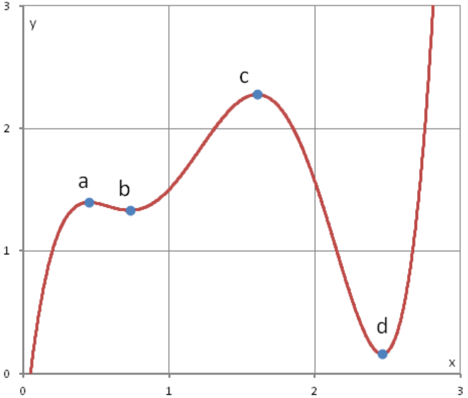
unstable manifolds



Morse-Smale Complex

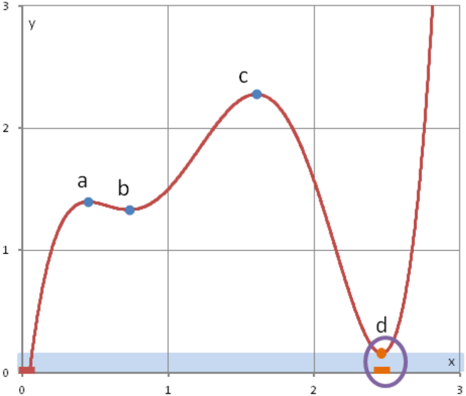
Persistence Simplification of the Morse-Smale Complex

Track birth and death of topological features



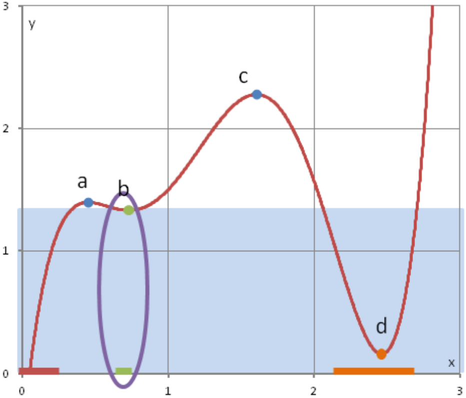
Persistence Simplification of the Morse-Smale Complex

Track birth and death of topological features



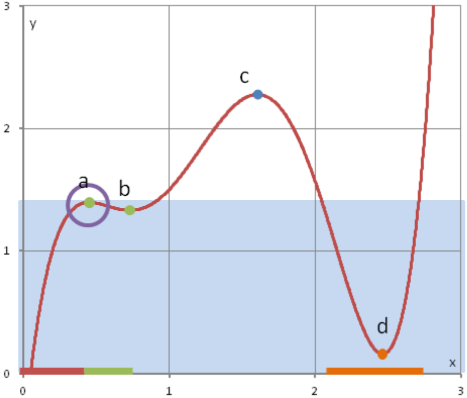
Persistence Simplification of the Morse-Smale Complex

Track birth and death of topological features



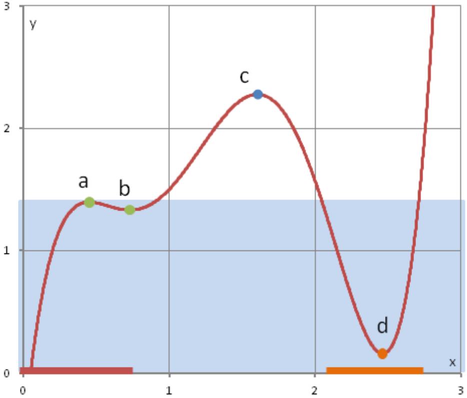
Persistence Simplification of the Morse-Smale Complex

Track birth and death of topological features



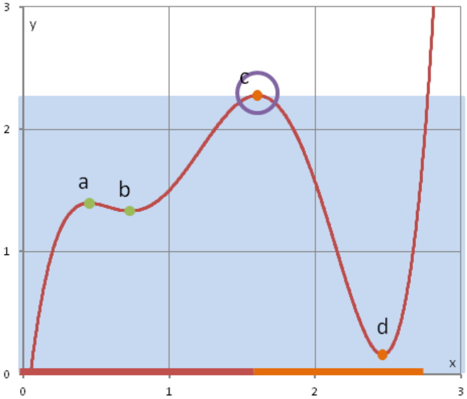
Persistence Simplification of the Morse-Smale Complex

Track birth and death of topological features



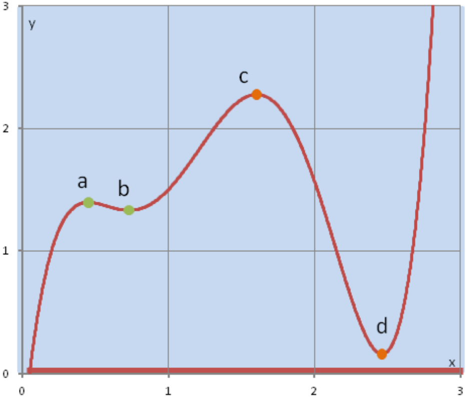
Persistence Simplification of the Morse-Smale Complex

Track birth and death of topological features



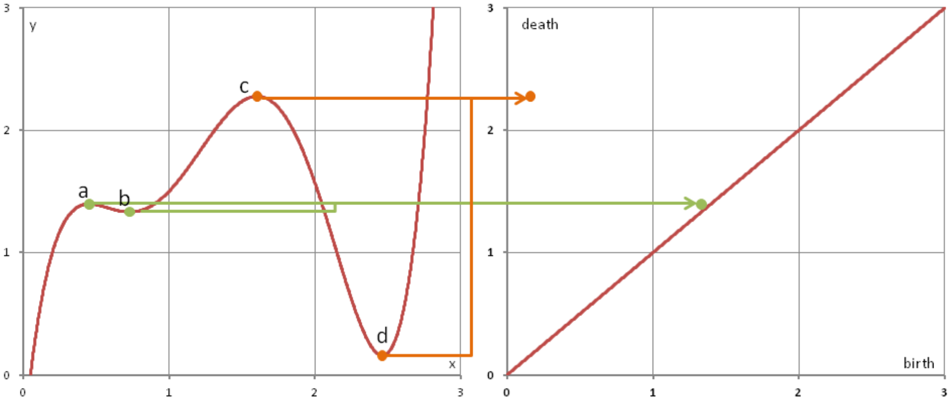
Persistence Simplification of the Morse-Smale Complex

Track birth and death of topological features

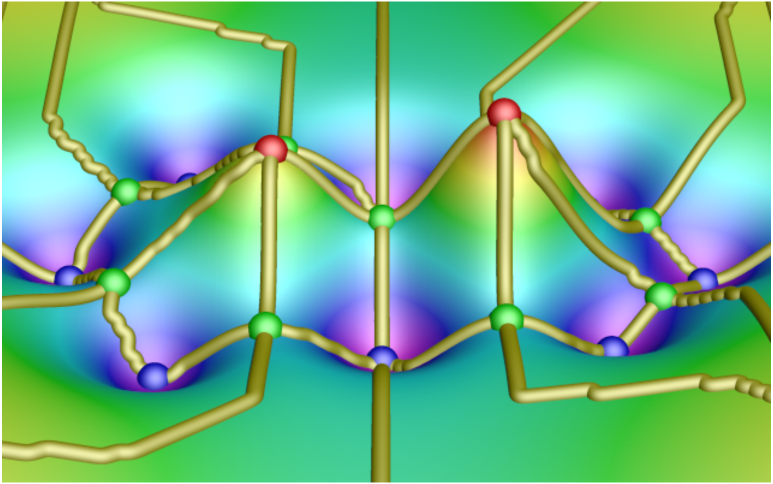


Persistence Simplification of the Morse-Smale Complex

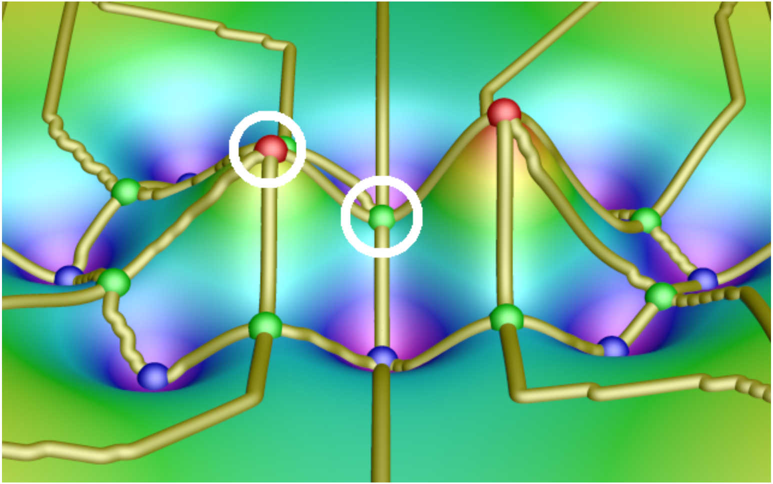
Associate pairs of critical points to a birth and death pair



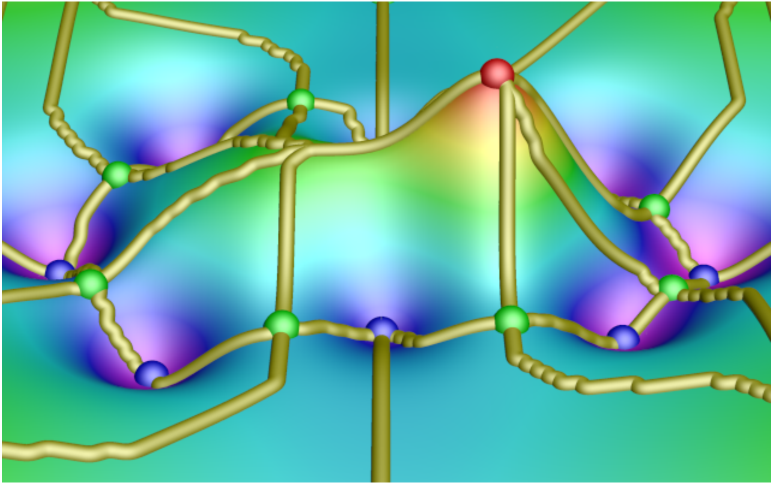
Persistence Simplification 2D Example



Persistence Simplification 2D Example

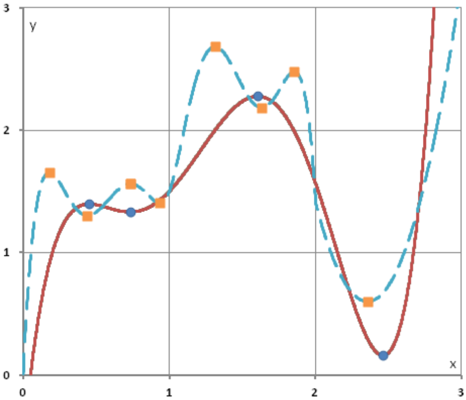


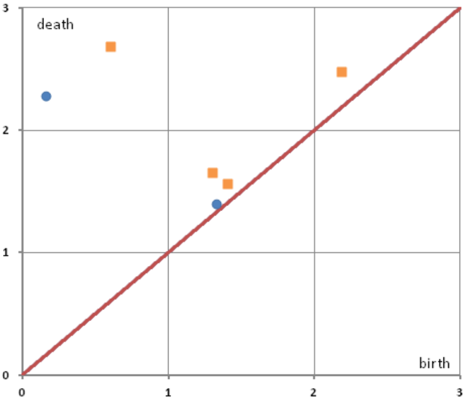
Persistence Simplification 2D Example



Bottleneck Distance

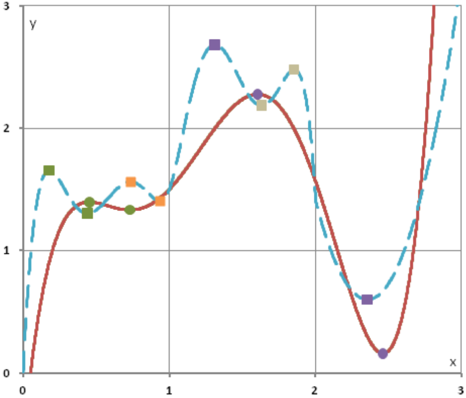
Comparing Mores-Smale complex of two similar function responses

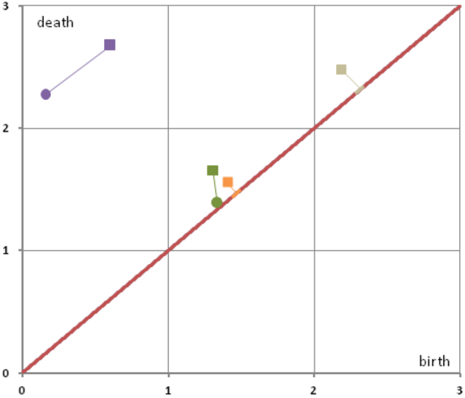




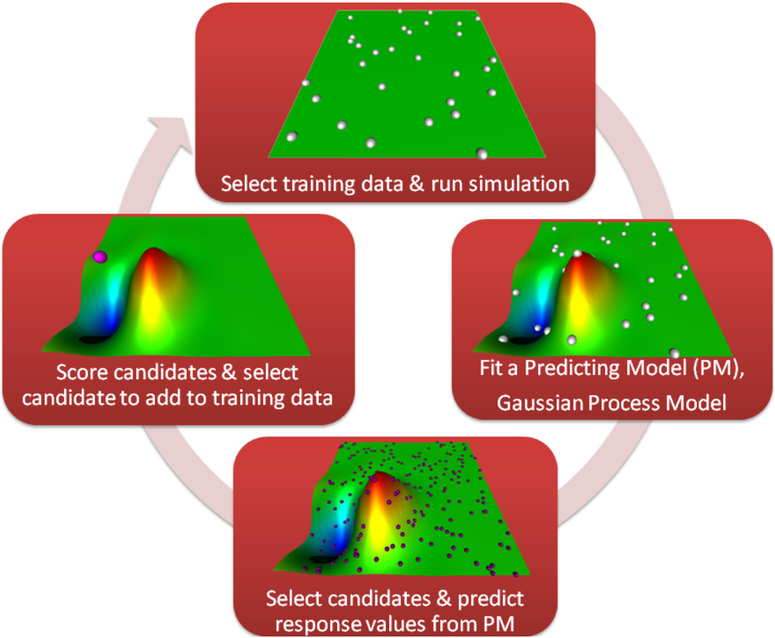
Bottleneck Distance

Comparing Mores-Smale complex of two similar function responses



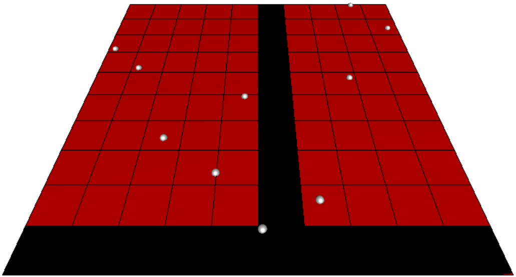


General Pipeline



Selecting Initial Training Set

Use space-filling algorithm



Our implementation uses Latin Hypercube Sampling (LHS)

- Fill axis-aligned hyperplanes evenly

Fitting a Statistical Model

Gaussian Process Model

- Stochastic model based on treating inputs as having normal distributions
- Inputs have multivariate normal distribution

Classical Scoring

Use LHS to choose candidates

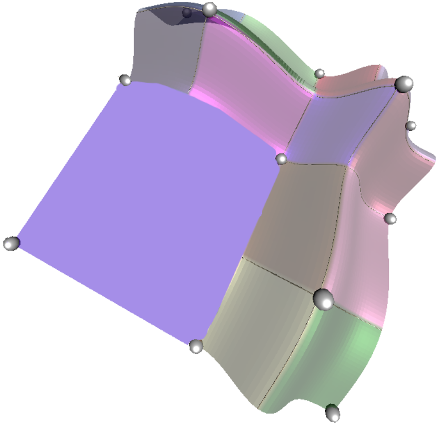
Most classic scoring functions rely on geometric or statistical concepts of the data:

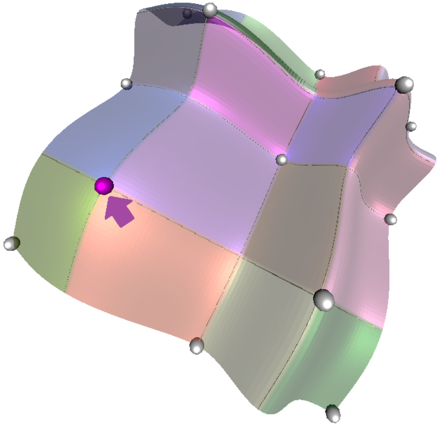
- Active-Learning McKay(ALM)
  - sample high-frequency or low-confidence regions
- Delta
  - distribute samples in the range space or areas of steep gradient
- Expected Improvement (EI)
  - Select points with high uncertainty or large discrepancy with existing data
- Distance (*DP)
  - Scaling factor applied to above, creating 3 new scoring functions (ALMDP, DeltaDP, EIDP)

Topological Scoring

Average Change in Persistence (TOPOP)

- Average change in persistence between before and after inserting a candidate into Morse-Smale

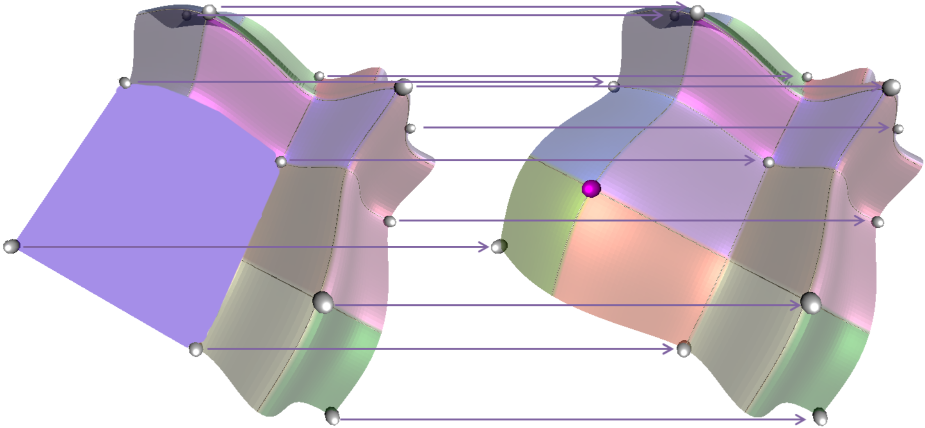




Topological Scoring

Average Change in Persistence (TOPOP)

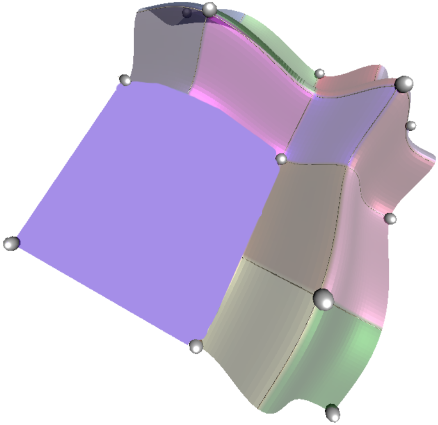
- Average change in persistence between before and after inserting a candidate into Morse-Smale

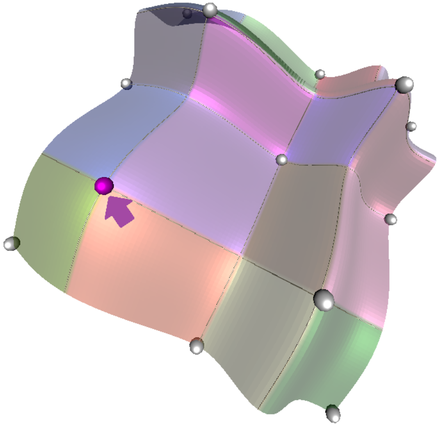


Topological Scoring

Bottleneck Distance in Persistence (TOPOB)

- Bottleneck distance between before and after inserting a candidate into Morse-Smale

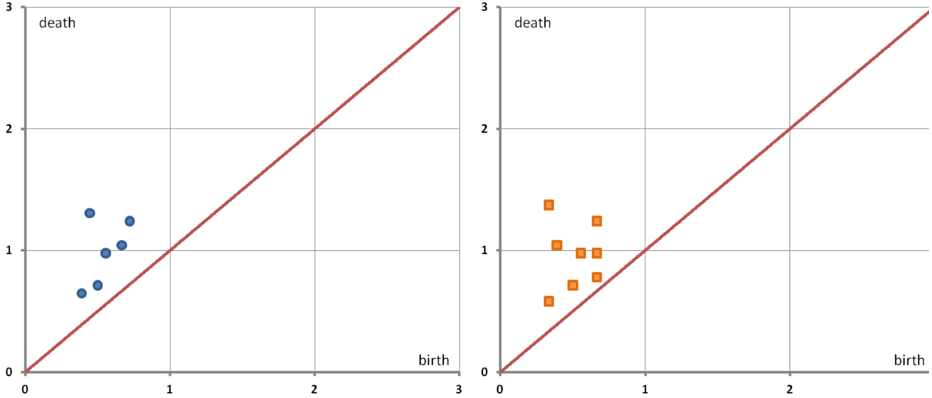




Topological Scoring

Bottleneck Distance in Persistence (TOPOB)

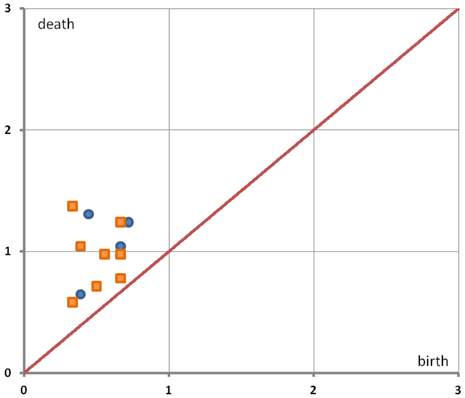
- Bottleneck distance between before and after inserting a candidate into Morse-Smale



Topological Scoring

Bottleneck Distance in Persistence (TOPOB)

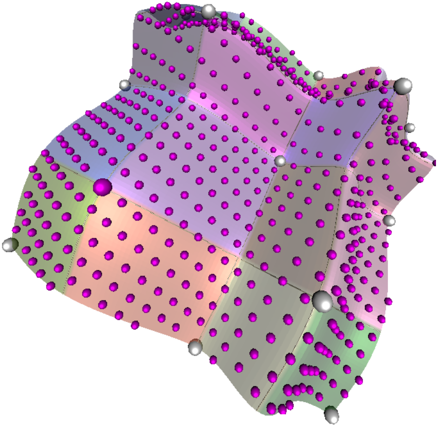
- Bottleneck distance between before and after inserting a candidate into Morse-Smale



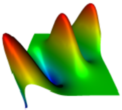
Topological Scoring

Highest Persistence (TOPOHP)

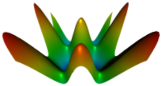
- Find highest persistence critical point in Morse-Smale complex constructed from training data and predicted responses of candidates



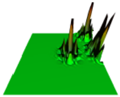
Testing Functions



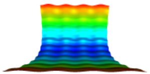
Generalized Diagonal (DmM)



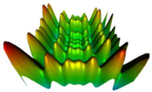
Ackley (Ack)



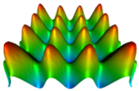
Modified Langerman (ML)



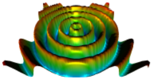
Mis-Scaled Generalized Rastrigin (MGR)



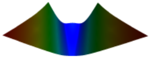
Normalized Schwefel (NS)



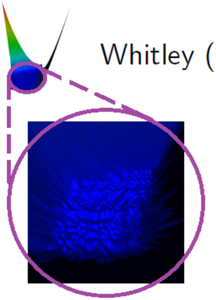
Generalized Rastrigin (Rast)



Salomon (Sal)

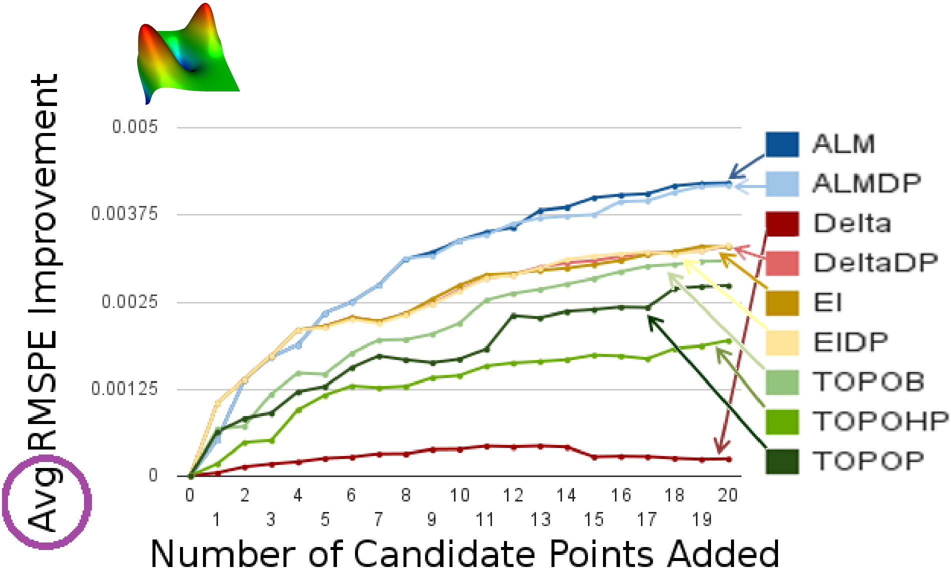


Generalized Rosenbrock (Rose)

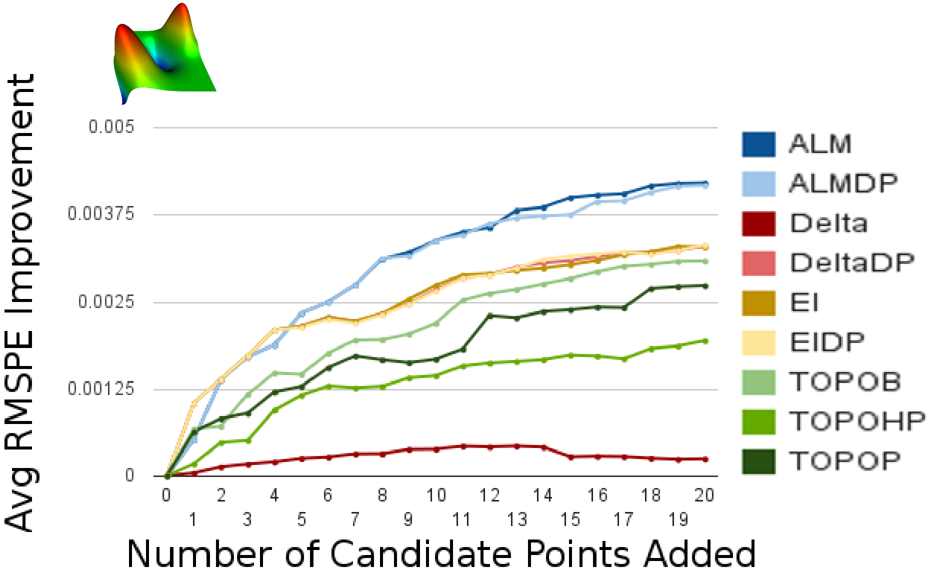


Whitley (Whit)

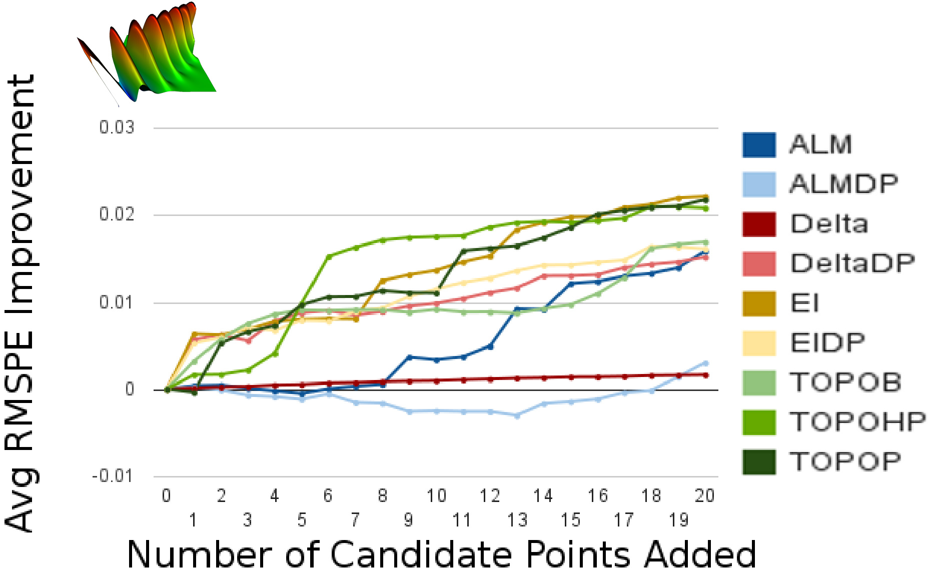
2 Maxima along Main Diagonal in 2D



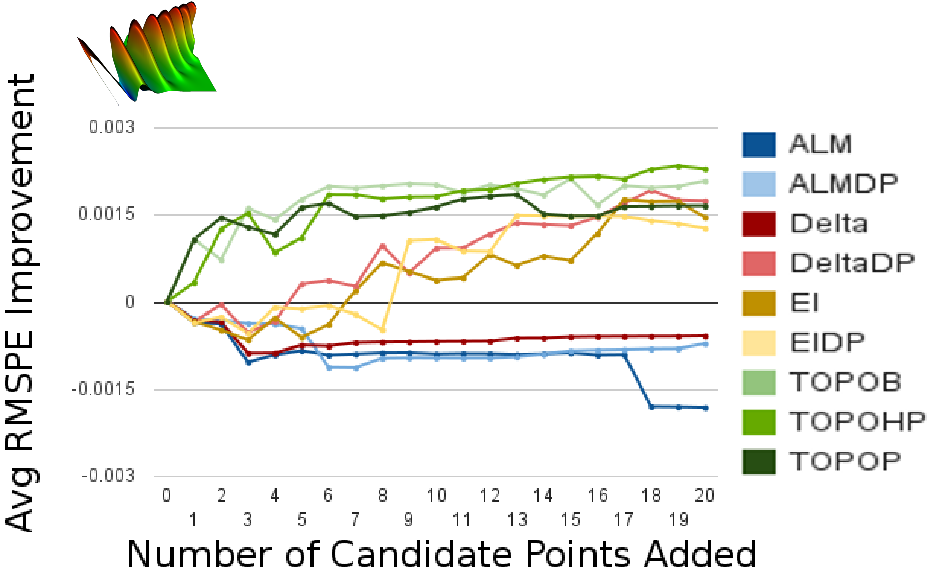
2 Maxima along Main Diagonal in 2D



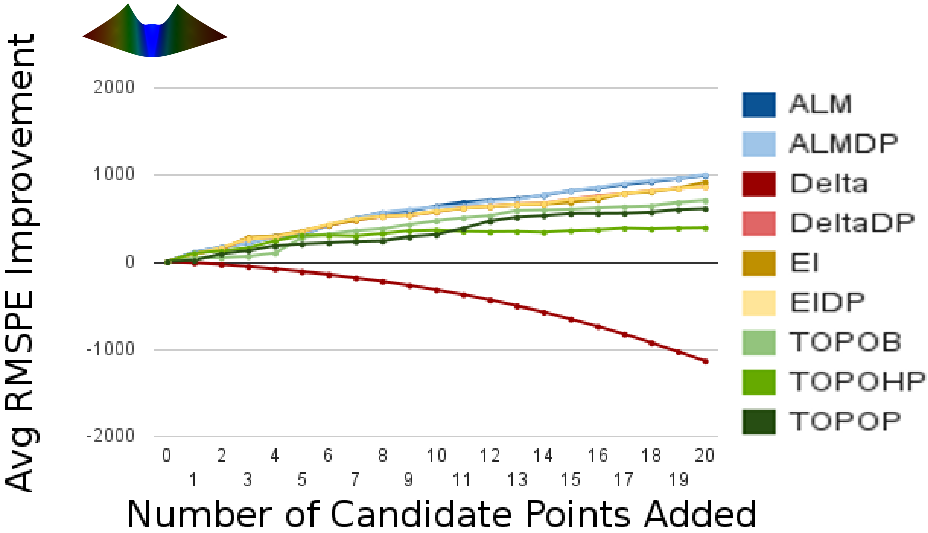
8 Maxima along Main Diagonal in 2D



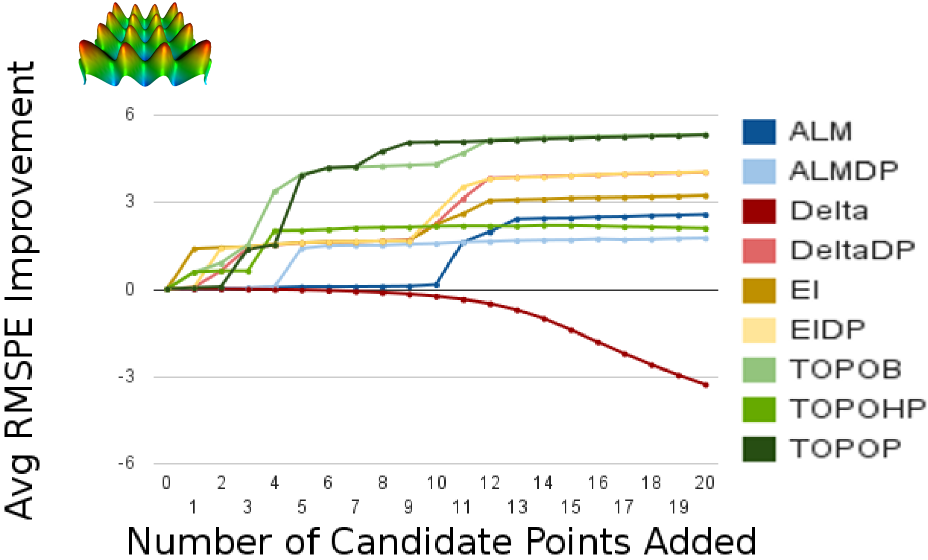
30 Maxima along Main Diagonal in 5D



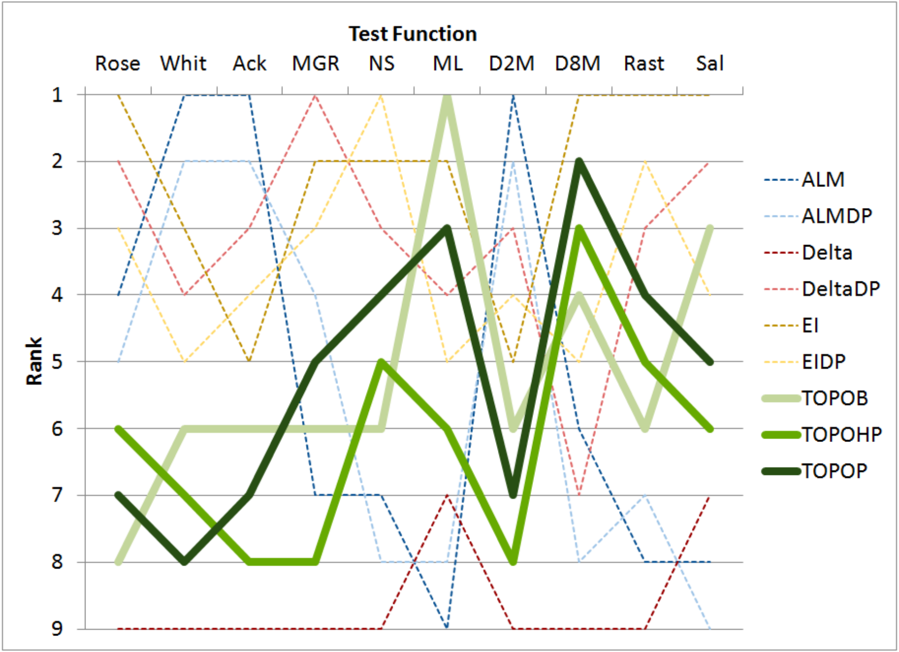
Rosenbrock in 4D



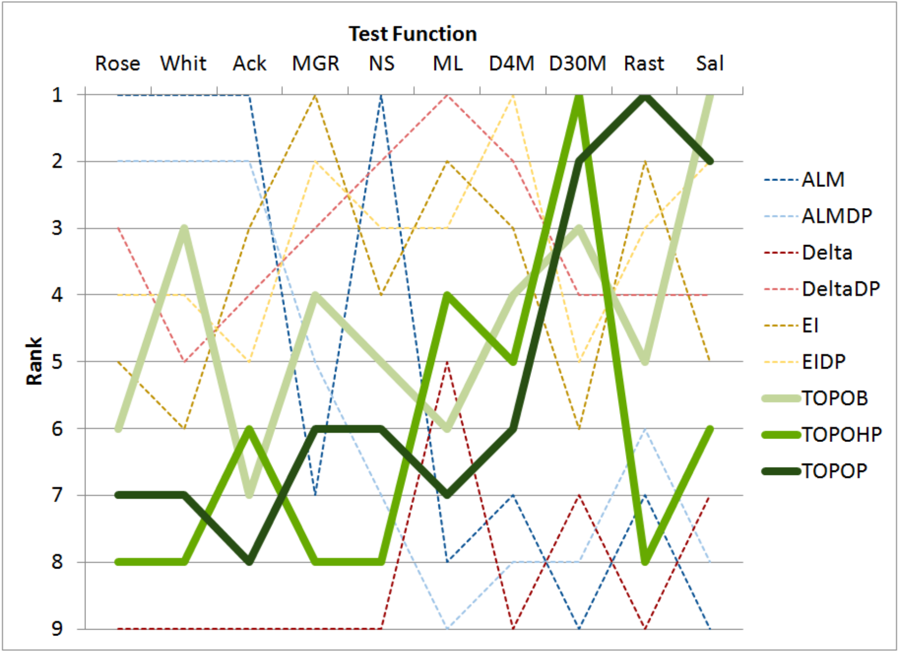
Rastrigin in 5D



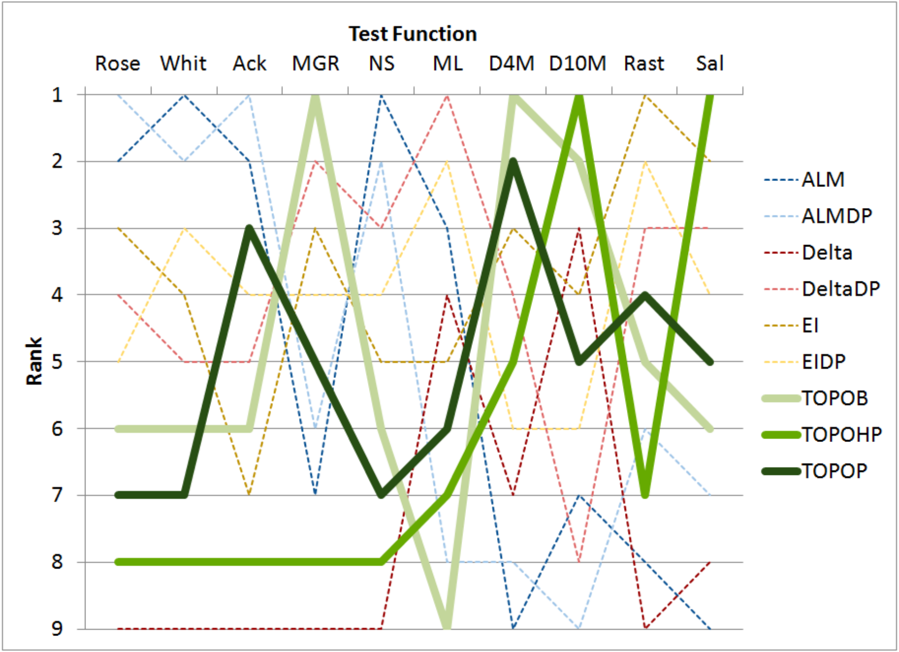
Ranking Trend in 2D



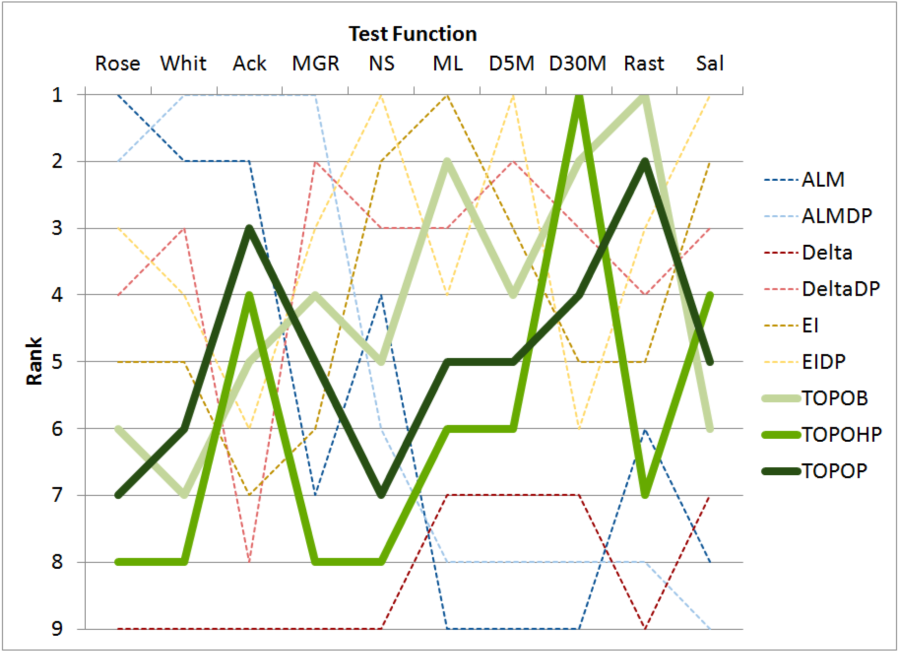
Ranking Trend in 3D



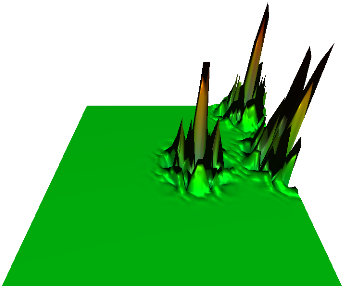
Ranking Trend in 4D



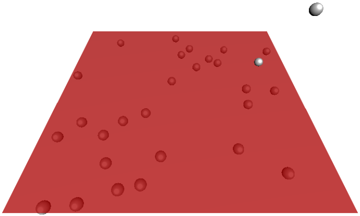
Ranking Trend in 5D



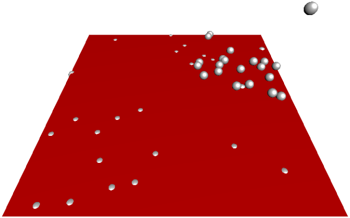
Discussion: GPM



Ground Truth Function



Initial Fit



Fit after addition of 20 points

Current/Future Work

- Use different predicting regression models
- Investigate different metrics for measuring effectiveness of adaptive sampling technique besides RMSPE
- Further investigate how to measure "topological impact" of a candidate point and hybrid methods
- Gain better understanding of "function classification" problem

Thank you

Questions?