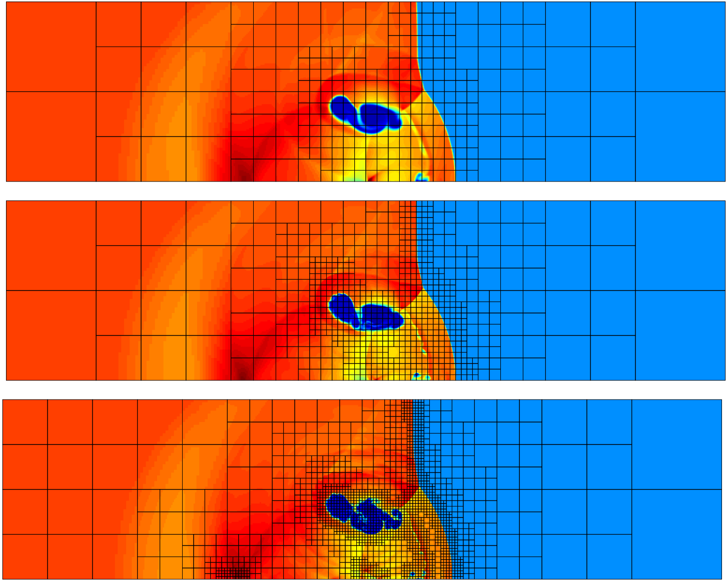
Active learning for cost-aware model reduction

Dmitry Duplyakin (U Utah and CU), **Jed Brown** (CU Boulder),
Donna Calhoun (Boise State)
jed.brown@colorado.edu

Copper Mountain Conference on Iterative Methods
https://jedbrown.org/files/20180327-ActiveLearning.pdf

AMR shock-bubble with 4, 5, and 6 levels



What goes wrong?

```
=====================================================================================
= BAD TERMINATION OF ONE OF YOUR APPLICATION PROCESSES
= PID 17873 RUNNING AT joule
= EXIT CODE: 139
= CLEANING UP REMAINING PROCESSES
= YOU CAN IGNORE THE BELOW CLEANUP MESSAGES
=====================================================================================
YOUR APPLICATION TERMINATED WITH THE EXIT STRING: Segmentation
fault (signal 11)
```

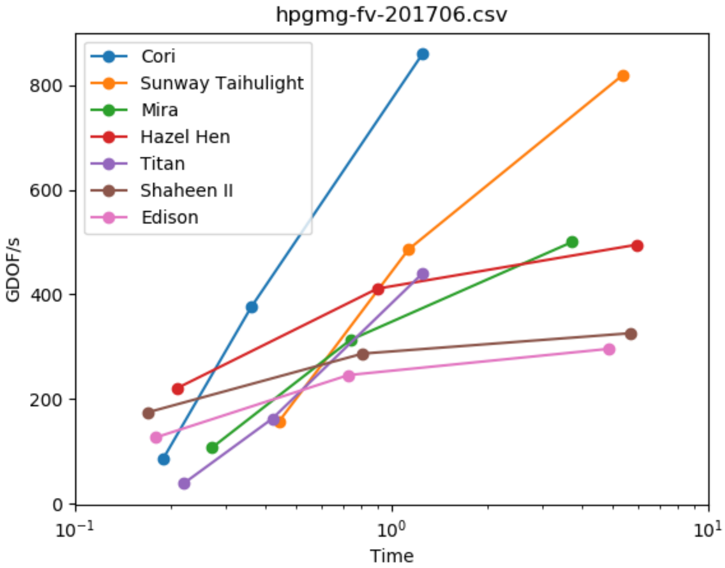
- Resubmit batch job, this time using more nodes.
- Tweak a refinement parameter.

On the last time step?

```
=====================================================================
= BAD TERMINATION OF ONE OF YOUR APPLICATION PROCESSES
= PID 16824 RUNNING AT joule
= EXIT CODE: 152
= CLEANING UP REMAINING PROCESSES
= YOU CAN IGNORE THE BELOW CLEANUP MESSAGES
=====================================================================
YOUR APPLICATION TERMINATED WITH THE EXIT STRING: CPU time limit
exceeded (signal 24)
```

- Checkpoint more often?
- Need to wait through the queue again.

New computer



Modeling

$$(\text{response}) = f(x) + \mathcal{N}(0, \sigma_n^2)$$

x user-relevant parameters

- Physics: bubble size, density, shock intensity
- Numerics: box size, max levels, refinement criteria
- Machine: # nodes, MPI/OpenMP, compilers

f Response

- CPU time
- Wall-clock time
- Peak memory usage
- Physics: Δ entropy, decay time

σ_n unbiased Gaussian noise

LOL Gaussian!

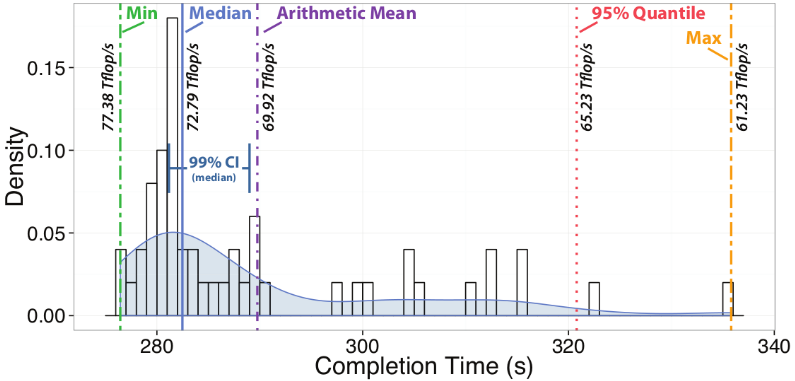


Figure 1: Distribution of 50 HPL measurement results.

[Hoefler & Belli, SC15]

Gaussian process regression

$$p\big(f(x_\star) \mid X, y\big) \sim \mathcal{N}\left(\mu_\star, \sigma_\star^2\right)$$

$$\mu_\star = k_\star^\mathsf{T} K_y^{-1} y$$
$$\sigma_\star^2 = k_{\star\star} - k_\star^\mathsf{T} K_y^{-1} k_\star$$
$$K_y = K + \sigma_n^2 I$$

where

$$[K]_{ij} = k(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{|x_i - x_j|^2}{2\ell^2}\right)$$

in terms of *hyperparameters* ℓ, σ_f, σ_n.

Optimizing hyperparameters

The evidence provided by (X, y) in support of $(\ell, \sigma_f, \sigma_n)$ is quantified by

$$\mathscr{L}(\ell, \sigma_f, \sigma_n) = \log p(y \mid X, l, \sigma_f^2, \sigma_n^2) = -\frac{1}{2}\left(y^\mathsf{T} K_y^{-1} y + \log|K_y|\right) + C$$

- Non-convex optimization problem
- Determinant $|K_y|$ due to normalization

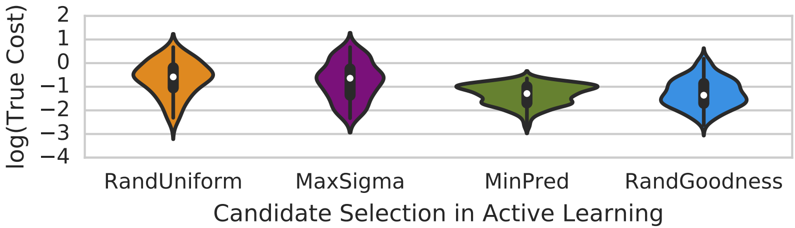
Offline Active Learning

- Precompute database of features and responses
- Partition data (X, y) into Initial, Active, Test
- Compare many “trajectories” using different partitions

Algorithm

- Train GPR for each feature (e.g., cost and memory) in Initial set
- Repeat
  1. Consult GPR models to select next observation from Active
  2. Make that observation (incurring cost, etc.)
  3. Retrain GPRs with new observation (including failure)

Selection procedure



RandUniform Uniform random sampling

MaxSigma Choose candidate with largest uncertainty

MinPred Maximize $\sigma_{\text{cost}}/\mu_{\text{cost}}$

RandGoodness Probability density $\sim \sigma_{\text{cost}}/\mu_{\text{cost}}$

RandGoodness with Memory Awarness As above, but exclude cases that violate memory bound

Metrics

Accuracy

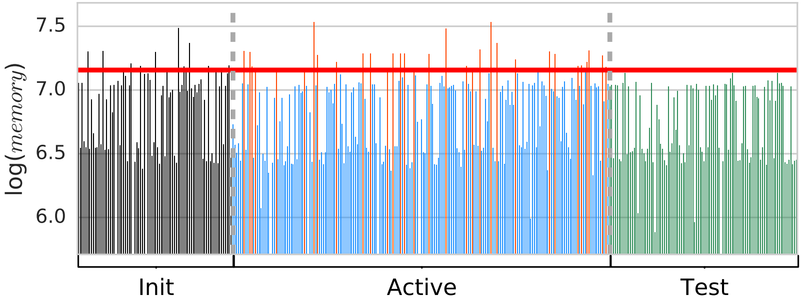
$$RMSE = \sqrt{\frac{1}{n_{\text{Test}}} e^T e}$$

where $e = \mu^{\text{Test}} - y^{\text{Test}}$

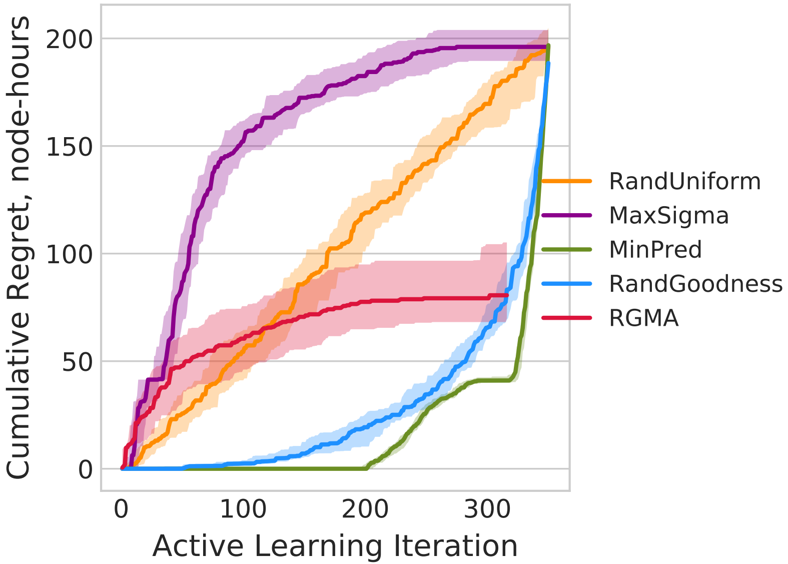
Cumulative Cost Total cost of selected experiments, including failures

Cumulative Regret Costs incurred attempting failures

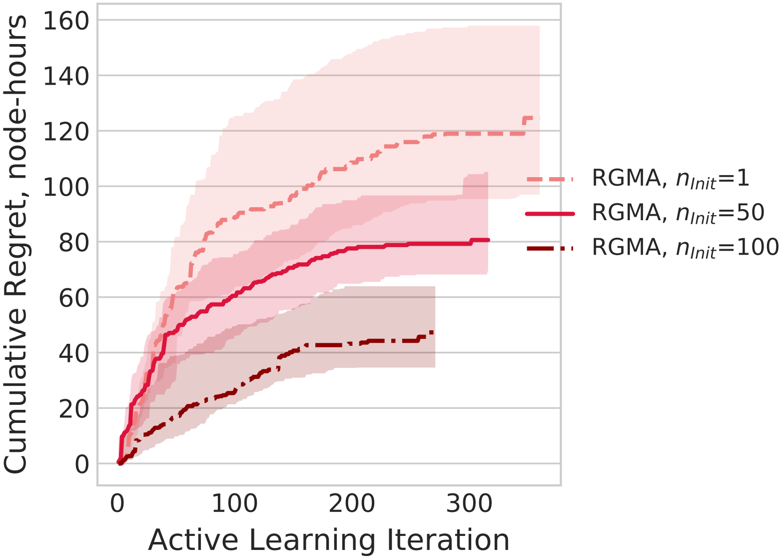
Memory limits



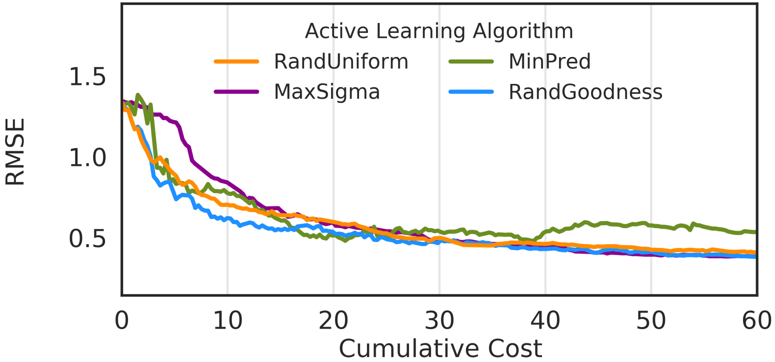
Cumulative Regret by Algorithm



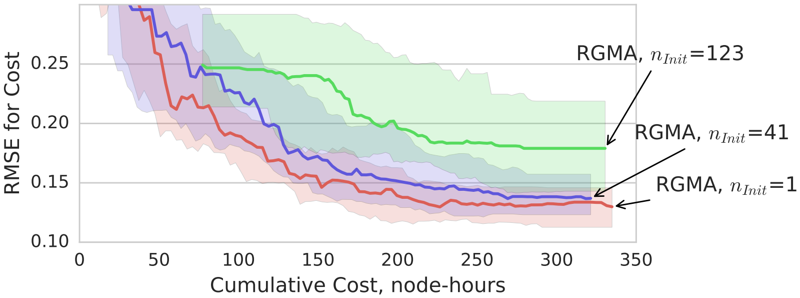
Cumulative Regret by nInit



RMSE vs Cumulative Cost



RMSE vs Cumulative Cost



Outlook

- Better measure of “risk” of exceeding memory bound
- Domain-specific kernel functions
- Non-Gaussian distributions
- Online mode
- Thanks to NERSC Edison and Cori, and to DOE ASCR