Multi-Phase Task-Based HPC Applications: Quickly Learning how to Run Fast

3rd Workshop of the LIG SRCPR Axis (2022)

Lucas Leandro Nesi, Lucas Mello Schnorr, Arnaud Legrand
May 12, 2022

POLARIS - Performance evaluation and Optimization of LARge Infrastructures and Systems
Institute of Informatics, Federal University of Rio Grande do Sul (UFRGS), Brazil
University Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, France
Heterogeneity in HPC – At System-Level

System-Level Heterogeneity: Hybrid Nodes with Different Computational Power

Santos Dumont

Five Partitions/Systems:
- Base CPU: 504 nodes
- Base Hybrid: 54 nodes
- Base GPU: 198 nodes
- BS CPU: 282 nodes
- BS GPU: 94 nodes

Jean Zay

Four Partitions/Systems:
- CPU Only: 1508 nodes
- CPU + GPU 1: 612 nodes
- CPU + GPU 2: 31 nodes
- CPU + GPU 3: 3 nodes
System-Level Heterogeneity: Hybrid Nodes with Different Computational Power

**Santos Dumont**

- Five Partitions/Systems:
  - Base CPU: 504 nodes
  - Base Hybrid: 54 nodes
  - Base GPU: 198 nodes
  - BS CPU: 282 nodes
  - BS GPU: 94 nodes

**Jean Zay**

- Four Partitions/Systems:
  - CPU Only: 1508 nodes
  - CPU + GPU 1: 612 nodes
  - CPU + GPU 2: 31 nodes
  - CPU + GPU 3: 3 nodes

**Modern HPC Applications**

- dpotrf
- dtrsm
- dsyrk
- dgemm

\[
\begin{array}{cccc}
0,0 & 1,0 & 2,0 & 3,0 \\
1,1 & 2,2 & 3,3 & 3,1 \\
2,1 & 3,2 & 3,3 & 3,2 \\
3,0 & 3,1 & 3,2 & 3,3 \\
\end{array}
\]
ExaGeoStat - Application Structure

Asynchronous Phases

- New θ
- Generation
- Cholesky
- Determinant
- Solve
- Dot Product

Optimization Iteration

(1) Generation O(n²)
(2) Cholesky O(n³)
(3) Determinant O(n)
(4) Solve O(n²)
(5) Dot product O(n)
ExaGeoStat - Application Structure

Asynchronous Phases

- New θ
- Generation
- Cholesky
- Determinant
- Solve
- Dot Product

Optimization
Iteration

(1) Generation O(n²)
(2) Cholesky O(n³)
(3) Determinant O(n)
(4) Solve O(n²)
(5) Dot product O(n)

ExaGeoStat - Sync

ExaGeoStat - Async

Node Occupation

Node Occupation

Time [ms]

Time [ms]
Behavior varying the number of nodes

On a different number of machines:

![Graph showing behavior varying the number of nodes on different machines.](image)
Behavior varying the number of nodes

On a different number of machines:

![Graph showing behavior varying the number of nodes.]
Traditional Multi-armed bandit overview

• Independent actions (arms)
Traditional Multi-armed bandit overview

• Independent actions (arms)

Legend

True Mean

- UCB
  - Select the action with the higher reward plus confidence
  - Good regret
Traditional Multi-armed bandit overview

- Independent actions (arms)

Legend: Sample Mean \( \rightarrow \) True Mean

- UCB
- Select the action with the higher reward plus confidence

<table>
<thead>
<tr>
<th>Action (Times Selected)</th>
<th>Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (20)</td>
<td>0.00</td>
</tr>
<tr>
<td>2 (4)</td>
<td>0.25</td>
</tr>
<tr>
<td>3 (10)</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Sample Mean

True Mean
Traditional Multi-armed bandit overview

• Independent actions (arms)

- **UCB**
  - Select the action with the higher reward plus confidence
  - Good regret
Gaussian Process overview

• Assumes a form of smoothness over the data
Gaussian Process overview

- Assumes a form of smoothness over the data
Gaussian Process overview

- Assumes a form of smoothness over the data
Gaussian Process overview

- Assumes a form of smoothness over the data
Gaussian Process overview

• Assumes a form of smoothness over the data
Gaussian Process overview

- Assumes a form of smoothness over the data
Gaussian Process overview

- Assumes a form of smoothness over the data
Gaussian Process overview

- Assumes a form of smoothness over the data
Gaussian Process overview

- Assumes a form of smoothness over the data
Gaussian Process overview

- Assumes a form of smoothness over the data
Experiments - GP-Discontinuous Step by Step (6L-30S)

GP-Discontinuous (Model overhead, Search-space limit, Discontinuities) evolution:

Iteration 1

Legend
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error

Time [s] vs. Factorization Nodes

6/7
GP-Discontinuous (Model overhead, Search-space limit, Discontinuities) evolution:

Iteration 8

Legend
- \(\text{Real Behavior}\)
- \(\text{Next Action}\)
- \(\text{Measurements}\)
- \(\text{Predictive Mean + UCB}\)
- \(\text{LP Prediction}\)
- \(\text{LP Error}\)
Experiments - GP-Discontinuous Step by Step (6L-30S)

GP-Discontinuous (Model overhead, Search-space limit, Discontinuities) evolution:

Iteration 10

Legend
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error
GP-Discontinuous (Model overhead, Search-space limit, Discontinuities) evolution:

Iteration 15

Legend:
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error
Experiments - GP-Discontinuous Step by Step (6L-30S)

GP-Discontinuous (Model overhead, Search-space limit, Discontinuities) evolution:

Iteration 20

Legend
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error
Experiments - GP-Discontinuous Step by Step (6L-30S)

GP-Discontinuous (Model overhead, Search-space limit, Discontinuities) evolution:

Iteration 25

Legend:
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error
Experiments - GP-Discontinuous Step by Step (6L-30S)

GP-Discontinuous (Model overhead, Search-space limit, Discontinuities) evolution:

Iteration 30

Legend
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error
Experiments - GP-Discontinuous Step by Step (6L-30S)

GP-Discontinuous (Model overhead, Search-space limit, Discontinuities) evolution:

Iteration 50

Legend
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error
Experiments - GP-Discontinuous Step by Step (6L-30S)

GP-Discontinuous (Model overhead, Search-space limit, Discontinuities) evolution:

Iteration 100
Conclusion

• Predicting the ideal number of resources to use is complex
• The application can learn it online and adapt
• We have a collection of setups and comparisons with other algorithms
  • Available online: https://adaphetnodes.shinyapps.io/shiny/
• We hope this strategy can be helpful for your problems as well
Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, the National Council for Scientific and Technological Development (CNPq), grant no 141971/2020-7 to the first author, the projects: FAPERGS (Data Science – 19/711-6), CAPES/Cofecub (04/2017) and the INRIA Associated Team ReDaS.

Some experiments were carried out using Grid’5000, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations (https://www.grid5000.fr). The authors acknowledge the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil) for providing HPC resources of the SDumont supercomputer, which have contributed to the research results reported in this paper (http://sdumont.lncc.br).
Thanks for your attention!

Contact: lucas.nesi@inf.ufrgs.br

Lucas Leandro Nesi, Lucas Mello Schnorr, Arnaud Legrand

May 12, 2022

POLARIS - Performance evaluation and Optimization of LARge Infrastructures and Systems
Institute of Informatics, Federal University of Rio Grande do Sul (UFRGS), Brazil
University Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, France
ExaGeoStat - Different number of machines

Using only 8 fast nodes

Node Occupation

Time [ms]

- dcmg
- dgeadd
- dgemm
- dlacpy
- dmdet
- dpotrf
- dsconv
- dsyrk
- dtrsm
- sdmat_reg
ExaGeoStat - Different number of machines

Using only 8 fast nodes

Using 23 het nodes (15 + 8)
ExaGeoStat - Different number of machines

- How many nodes a application should use?
ExaGeoStat - Different number of machines

- How many nodes a application should use?
Experiments - GP-UCB Step by Step (6L-30S)

GP-UCB evolution:

Iteration 1

Legend
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error
GP-UCB evolution:

Iteration 8

- Initial measurements
Experiments - GP-UCB Step by Step (6L-30S)

GP-UCB evolution:

Legend
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error

Iteration 10

Factorization Nodes

Time [s]

0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36
Experiments - GP-UCB Step by Step (6L-30S)

GP-UCB evolution:

Iteration 15

Legend
- Red: Real Behavior
- Blue: Next Action
- Cross: Measurements
- Black: Predictive Mean + UCB
- Blue Circle: LP Prediction
- Purple: LP Error

Factorization Nodes vs. Time [s]
GP-UCB evolution:

Iteration 20

Legend:
- Red circle: Real Behavior
- Red plus sign: Next Action
- Purple cross: Measurements
- Black circle: Predictive Mean + UCB
- Blue line: LP Prediction
- Purple line: LP Error

Factorization Nodes vs. Time [s]
GP-UCB evolution:

Iteration 25

Legend:
- Red Circle: Real Behavior
- Red Cross: Next Action
- Black Circle: Predictive Mean + UCB
- Blue Line: LP Prediction
- Purple Line: LP Error
Experiments - GP-UCB Step by Step (6L-30S)

GP-UCB evolution:

Iteration 30

Legend:
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error

- Few repetitions ⇒ Poor confidence estimation
Experiments - GP-UCB Step by Step (6L-30S)

GP-UCB evolution:

Iteration 50

Legend
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error
Experiments - GP-UCB Step by Step (6L-30S)

GP-UCB evolution:

Iteration 100

Legend
- Red Circle: Real Behavior
- Red Cross: Next Action
- Black Circle: Predictive Mean + UCB
- Blue Line: LP Prediction
- Purple Line: LP Error
Experiments - GP-UCB Step by Step (6L-30S)

GP-UCB evolution:

Iteration 100

Legend
- Real Behavior
- Next Action
- Measurements
- Predictive Mean + UCB
- LP Prediction
- LP Error

- Measuring every possibility is bad!