

Scheduling Bags of Non-identical Tasks

Henri Casanova and Matthieu Gallet and Frédéric Vivien

November 13, 2009

The Problem

- ▶ A master-worker platform
- ▶ Several bag-of-tasks applications
(Each application is a collection of similar tasks)
- ▶ Objective: maximizing the throughput
- ▶ **Bad news:** a bag is made of similar but not identical tasks

Presentation outline

Offline Case: Identical Tasks

Offline Case: Tasks With Different Characteristics

Online Case: Tasks With Different Characteristics

Simulations

Presentation outline

Offline Case: Identical Tasks

Offline Case: Tasks With Different Characteristics

Online Case: Tasks With Different Characteristics

Simulations

Notation

- ▶ A master P_0 which has an output bandwidth of bw_0
- ▶ n workers: P_1, \dots, P_n
- ▶ Processor P_i has
 - ▶ a speed of s_i
 - ▶ an input bandwidth of bw_i
- ▶ m bag-of-tasks applications
- ▶ Tasks of bag k have
 - ▶ a volume of computation of $V_{comp}(k)$
 - ▶ a volume of computation of $V_{comm}(k)$
- ▶ Communication model:
bounded multi-port with linear communications times

Constraints

1. Cumulative throughput of T_k :

$$\rho^{(k)} = \sum_{1 \leq i < n} \rho_i^{(k)}$$

2. Throughput of T_k proportional to its priority:

$$\frac{\rho^{(k)}}{\pi_k} = \frac{\rho^{(1)}}{\pi_1}$$

Objective

$$\text{MAXIMIZE } \rho^{(1)}$$

Constraints (continued)

3. Constraint on computation capabilities of worker P_i

$$\sum_{1 \leq k \leq m} \rho_i^{(k)} \frac{V_{comp}(k)}{s_i} \leq 1$$

4. Constraint on communication capabilities of worker P_i

$$\sum_{1 \leq k \leq m} \rho_i^{(k)} \frac{V_{comm}(k)}{bw_i} \leq 1$$

5. Constraint on communication capabilities of the master

$$\sum_{1 \leq i < n} \sum_{1 \leq k \leq m} \rho_i^{(k)} \frac{V_{comm}(k)}{bw_0} \leq 1$$

Complete Linear Program

$$\left\{ \begin{array}{l} \text{MAXIMIZE } \rho^{(1)} \text{ UNDER THE CONSTRAINTS} \\ \forall k \in [1, m], \quad \sum_{1 \leq i < n} \rho_i^{(k)} = \rho^{(k)} \\ \forall k \in [1, m], \quad \frac{\rho^{(k)}}{\pi_k} = \frac{\rho^{(1)}}{\pi_1} \\ \forall i \in [1, n], \quad \sum_{1 \leq k \leq m} \rho_i^{(k)} \frac{V_{comp}(k)}{s_i} \leq 1 \\ \forall i \in [1, n], \quad \sum_{1 \leq k \leq m} \rho_i^{(k)} \frac{V_{comm}(k)}{bw_i} \leq 1 \\ \sum_{1 \leq i < n} \sum_{1 \leq k \leq m} \rho_i^{(k)} \frac{V_{comm}(k)}{bw_0} \leq 1 \end{array} \right.$$

Presentation outline

Offline Case: Identical Tasks

Offline Case: Tasks With Different Characteristics

Online Case: Tasks With Different Characteristics

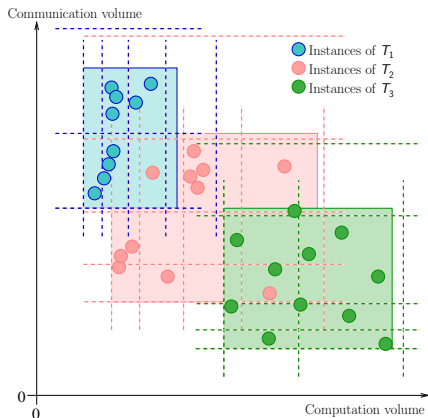
Simulations

Notation

- ▶ A master P_0 which has an output bandwidth of bw_0
- ▶ n workers: P_1, \dots, P_n
- ▶ Processor P_i has
 - ▶ a speed of s_i
 - ▶ an input bandwidth of bw_i
- ▶ m bag-of-tasks applications
- ▶ Tasks of bag k have
 - ▶ $X_{comm}^{(k)}$ is a random variable
the u -th instance has a communication volume of $X_{comm}^{(k)}(u)$
 $\min_{comm}^{(k)} \leq X_{comm}^{(k)}(u) \leq \max_{comm}^{(k)}$
 - ▶ $X_{comp}^{(k)}$ is a random variable
the u -th instance has a computation volume of $X_{comp}^{(k)}(u)$
 $\min_{comp}^{(k)} \leq X_{comp}^{(k)}(u) \leq \max_{comp}^{(k)}$
- ▶ Communication model:
bounded multi-port with linear communications times

An ε -approximation scheme

Underlying principle: split each application into several virtual applications in which two instances only have small differences in term of communication and computation volumes.



Formal splitting

$$\gamma_q^{(k)} = (1 + \varepsilon)^q \min_{comp}^{(k)}, \text{ with } 0 \leq q \leq Q^{(k)} = 1 + \left\lfloor \frac{\ln\left(\frac{\max_{comp}^{(k)}}{\min_{comp}^{(k)}}\right)}{\ln(1+\varepsilon)} \right\rfloor$$

$$\delta_r^{(k)} = (1 + \varepsilon)^r \min_{comm}^{(k)}, \text{ with } 0 \leq r \leq R^{(k)} = 1 + \left\lfloor \frac{\ln\left(\frac{\max_{comm}^{(k)}}{\min_{comm}^{(k)}}\right)}{\ln(1+\varepsilon)} \right\rfloor$$

Instance u of T_k belongs to $I_{q,r}^{(k)} = [\gamma_q^{(k)}; \gamma_{q+1}^{(k)}] \times [\delta_r^{(k)}; \delta_{r+1}^{(k)}]$ if

- ▶ $\gamma_q^{(k)} \leq X_{comp}^{(k)}(u) \leq \gamma_{q+1}^{(k)}$ and
- ▶ $\delta_r^{(k)} \leq X_{comm}^{(k)}(u) \leq \delta_{r+1}^{(k)}$

Virtual applications

- ▶ Instances of T_k in $I_{q,r}^{(k)}$ define virtual application $T_{k,q,r}$
- ▶ $p_{q,r}^{(k)}$ probability of an instance of T_k to belong to virtual application $T_{k,q,r}$:

$$p_{q,r}^{(k)} = \mathcal{P} \left(\gamma_q^{(k)} \leq X_{comp}^{(k)} < \gamma_{q+1}^{(k)}; \delta_r^{(k)} \leq X_{comm}^{(k)} < \delta_{r+1}^{(k)} \right)$$

$$\forall k, \sum_{q,r} p_{q,r}^{(k)} = 1$$

- ▶ $\rho_{i,q,r}^{(k)}$: contribution of processor P_i to the throughput of virtual application $T_{k,q,r}$
- ▶ Throughput of virtual application $T_{k,q,r}$ is related to the throughput of T_k :

$$\forall k, \forall q < Q^{(k)}, \forall r < R^{(k)}, \sum_{1 \leq i < n} \rho_{i,q,r}^{(k)} = p_{q,r}^{(k)} \rho^{(k)}$$

Transposing the constraints

- ▶ Throughput of T_k is still proportional to its priority:

$$\forall k \in [1, m], \frac{\rho^{(k)}}{\pi_k} = \frac{\rho^{(1)}}{\pi_1}$$

- ▶ Constraint on computation capabilities of worker P_i
Problem: We do not know the execution time of instances
Solution: We (conservatively) over-approximate them

$$\forall i \in [1, n], \sum_{k=1}^m \sum_{\substack{q < Q^{(k)} \\ r < R^{(k)}}} \left(\rho_{i,q,r}^{(k)} \frac{\gamma_{r+1}^{(k)}}{s_j} \right) \leq 1$$

Transposing the constraints (cont.)

- ▶ Constraint on communication capabilities of worker P_i

$$\forall 1 \leq i < n, \sum_{k=1}^m \sum_{\substack{q < Q^{(k)} \\ r < R^{(k)}}} \left(\rho_{i,q,r}^{(k)} \frac{\delta_{r+1}^{(k)}}{bw_i} \right) \leq 1$$

- ▶ Constraint on communication capabilities of the master

$$\sum_{k=1}^m \sum_{\substack{q < Q^{(k)} \\ r < R^{(k)}}} \left(\rho_{i,q,r}^{(k)} \frac{\delta_{r+1}^{(k)}}{bw_0} \right) \leq 1$$

New linear program

MAXIMIZE $\rho = \rho^{(1)}$ UNDER THE CONSTRAINTS

$$\forall k \in [1, m], \forall q < Q^{(k)}, \forall r < R^{(k)}, \quad \sum_{i=1}^n \rho_{i,q,r}^{(k)} = p_{q,r}^{(k)} \rho^{(k)}$$

$$\forall k \in [1, m], \quad \frac{\rho^{(k)}}{\pi_k} = \frac{\rho^{(1)}}{\pi_1}$$

$$\forall i \in [1, n], \quad \sum_{k=1}^m \sum_{\substack{q < Q^{(k)} \\ r < R^{(k)}}} \left(\rho_{i,q,r}^{(k)} \frac{\gamma_{q+1}^{(k)}}{s_i} \right) \leq 1$$

$$\forall i \in [1, n], \quad \sum_{k=1}^m \sum_{\substack{q < Q^{(k)} \\ r < R^{(k)}}} \left(\rho_{i,q,r}^{(k)} \frac{\delta_{r+1}^{(k)}}{bw_i} \right) \leq 1$$

$$\sum_{i=1}^n \sum_{k=1}^m \sum_{\substack{q < Q^{(k)} \\ r < R^{(k)}}} \left(\rho_{i,q,r}^{(k)} \frac{\delta_{r+1}^{(k)}}{bw_0} \right) \leq 1$$

Performance

Theorem.

An optimal solution of the Linear Program describes a solution with a throughput ρ larger than $\rho^*/(1 + \varepsilon)$, where ρ^* is the optimal throughput.

Presentation outline

Offline Case: Identical Tasks

Offline Case: Tasks With Different Characteristics

Online Case: Tasks With Different Characteristics

Simulations

Aim

- ▶ Non-clairvoyant about computation volumes
- ▶ Communication volumes can be supposed to be known
- ▶ Underlying distributions are unknown

Is there any hope?

Case with dominant computations

Theorem.

ON-DEMAND policy is asymptotically optimal when

- ▶ Computations are always dominant:

$$\forall i \in [1, n], \quad \min_{k,u} \frac{X_{comp}^{(k)}(u)}{s_i} \geq \max_{k',u'} \frac{X_{comm}^{(k')}(u')}{bw_i}$$

- ▶ The master's bandwidth is not constraining:

$$bw_0 \geq \sum_{i=1}^n bw_i$$

- ▶ Each worker as a limited number of buffers ($\in [2, n_{buffers}]$)

Principle of the proof (1/3)

Notation

- ▶ Γ : worst computation time
- ▶ Δ : worst communication time
- ▶ R_i : computation volume allocated to worker P_i
- ▶ T_i : completion time of worker P_i

We consider the scheduling of N tasks

Principle of the proof (2/3)

- ▶ t : time the first worker completes its work
- ▶ $Makespan = \max_i T_i \leq t + (b + 1)\Gamma$

$$Makespan - (b + 1)\Gamma \leq t \quad (1)$$

- ▶ Dominating computations

$$t \leq T_i \leq \Delta + \frac{R_i}{s_i} \quad (2)$$

- ▶ Combining Equations 1 and 2

$$Makespan - (b + 1)\Gamma \leq \Delta + \frac{R_i}{s_i}$$

Principle of the proof (3/3)

- ▶ Combining Equations 1 and 2

$$\text{Makespan} - (b + 1)\Gamma \leq \Delta + \frac{R_i}{s_i}$$

- ▶ By summation

$$\sum_i s_i (\text{Makespan} - (b + 1)\Gamma) \leq \left(\sum_i s_i \right) \Delta + \sum_i R_i$$

- ▶ Trivial bound: $\frac{\sum_i R_i}{\sum_i s_i} \leq \text{Makespan}_{opt}$
- ▶ Asymptotic optimality

$$\text{Makespan}_{opt} - (b + 1)\Gamma \leq \text{Makespan} - (b + 1)\Gamma \leq \Delta + \text{Makespan}_{opt}$$

Case with dominant computations (extension)

Theorem.

ON-DEMAND policy is asymptotically optimal when

- ▶ Processor P_i is always granted at least a fraction α_i of its input bandwidth when it requests data
- ▶ Computations are always dominant:

$$\forall i \in [1, n], \frac{\min_{k,u} X_{comp}^{(k)}(u)}{s_i} \geq \frac{\max_{k',u'} X_{comm}^{(k')}(u')}{\alpha_i bw_i}$$

- ▶ Each worker as a limited number of buffers ($\in [2, n_{buffers}]$)

Case with infinite buffers

Theorem.

ON-DEMAND has no constant competitive ratio

- ▶ 1 application with N tasks and unitary communication and computation volume, master's bandwidth not constraining
- ▶ $bw_1 = \frac{1}{2N}$; $bw_2 = \dots = bw_n = 1$
- ▶ $s_1 = 2(n-1)N$; $s_2 = \dots = s_n = 1$
- ▶ Possible schedule: ignore worker P_1 :
$$makespan_{opt} \leq \left\lceil \frac{N}{n-1} \right\rceil + 1$$
- ▶ solution of ON-DEMAND 1 task each for P_2, \dots, P_n ,
 $N - (n-1)$ tasks for P_1 .
$$Makespan_{ON-DEMAND} \geq (N - (n-1))s_1 \geq N \times Makespan_{opt}$$

(for $N \geq 4n$).

Case with dominant communications

Theorem.

ON-DEMAND policy is asymptotically optimal when

- ▶ Communications are always dominant:

$$\forall i \in [1, n], \quad \max_{k,u} \frac{X_{comp}^{(k)}(u)}{s_i} \leq \min_{k',u'} \frac{X_{comm}^{(k')}(u')}{bw_i}$$

- ▶ Each worker has a limited number of buffers ($\in [2, n_{buffers}]$)

Practical heuristics

- ▶ Use the first 10% of instances to gather data on applications
- ▶ From this sample, split applications into virtual applications
 - ▶ arithmetical buckets
 - ▶ geometrical buckets
 - ▶ recursive buckets

(We only report on Geometrical buckets as they lead to (slightly) better results)

- ▶ Apply the multi-application linear program on the virtual applications (with the rounding used for tasks with different characteristics)
- ▶ Schedule realized using a 1D load-balancing among processors (per virtual application)

Presentation outline

Offline Case: Identical Tasks

Offline Case: Tasks With Different Characteristics

Online Case: Tasks With Different Characteristics

Simulations

Simulation settings

- ▶ 3 or 4 applications
- ▶ 100, 1000, or 5000 instances per application
- ▶ Communication volume uniformly picked in $[\min_{comm}; \max_{comm}]$ with $\max_{comm} / \min_{comm}$ in $\{1, 1.35, 1.65, 2.35, 2.65\}$.
- ▶ Correlation factor $\phi \in [0, 1]$ (0: no correlation).
For instance u : $\exists \lambda, X_{comm}^{(k)}(u) = \lambda \min_{comm}^{(k)} + (1 - \lambda) \max_{comm}^{(k)}$
 $V_{comp}(i)$ is randomly picked in

$$\left[(\phi\lambda + 1 - \phi) \min_{comp}^{(k)} + \phi(1 - \lambda) \max_{comp}^{(k)}, \right. \\ \left. \phi\lambda \min_{comp}^{(k)} + (1 - \lambda\phi) \max_{comp}^{(k)} \right]$$

- ▶ Platforms: 3, 5, 10, or 15 workers.
Master's bandwidth = 1, 5, or 100 times the average bandwidth of workers

Overall results

Heuristic	Normalized to best	Normalized to UB
ON-DEMAND	0.87 ($\sigma = 0.108$)	0.821 ($\sigma = 0.109$)
ROUND-ROBIN	0.779 ($\sigma = 0.123$)	0.736 ($\sigma = 0.126$)
LP_SAMP(ARITH, 1, 1)	0.971 ($\sigma = 0.0362$)	0.917 ($\sigma = 0.0651$)
LP_SAMP(GEOM, 2, 1)	0.875 ($\sigma = 0.106$)	0.829 ($\sigma = 0.122$)
LP_SAMP(GEOM, 4, 1)	0.819 ($\sigma = 0.13$)	0.777 ($\sigma = 0.144$)
LP_SAMP(GEOM, 8, 1)	0.795 ($\sigma = 0.136$)	0.754 ($\sigma = 0.149$)
LP_SAMP(GEOM, 2, 2)	0.842 ($\sigma = 0.129$)	0.799 ($\sigma = 0.144$)
LP_SAMP(GEOM, 4, 4)	0.812 ($\sigma = 0.139$)	0.771 ($\sigma = 0.153$)
0.05-approx	0.993 ($\sigma = 0.022$)	0.937 ($\sigma = 0.0555$)
0.2-approx	0.985 ($\sigma = 0.0201$)	0.93 ($\sigma = 0.0513$)

Communication / Computation Ratio = 0.05

Heuristic	Normalized to best	Normalized to UB
ON-DEMAND	0.993 ($\sigma = 0.00687$)	0.937 ($\sigma = 0.0397$)
ROUND-ROBIN	0.716 ($\sigma = 0.101$)	0.676 ($\sigma = 0.104$)
LP_SAMP(ARITH, 1, 1)	0.97 ($\sigma = 0.0443$)	0.917 ($\sigma = 0.0749$)
LP_SAMP(GEOM, 2, 1)	0.878 ($\sigma = 0.118$)	0.832 ($\sigma = 0.132$)
LP_SAMP(GEOM, 4, 1)	0.797 ($\sigma = 0.144$)	0.756 ($\sigma = 0.155$)
LP_SAMP(GEOM, 8, 1)	0.76 ($\sigma = 0.151$)	0.722 ($\sigma = 0.162$)
LP_SAMP(GEOM, 2, 2)	0.86 ($\sigma = 0.148$)	0.816 ($\sigma = 0.16$)
LP_SAMP(GEOM, 4, 4)	0.801 ($\sigma = 0.157$)	0.761 ($\sigma = 0.169$)
0.05-approx	0.979 ($\sigma = 0.0392$)	0.926 ($\sigma = 0.0714$)
0.2-approx	0.974 ($\sigma = 0.0376$)	0.92 ($\sigma = 0.0691$)

Communication / Computation Ratio = 1

Heuristic	Normalized to best	Normalized to UB
ON-DEMAND	0.81 ($\sigma = 0.115$)	0.763 ($\sigma = 0.115$)
ROUND-ROBIN	0.81 ($\sigma = 0.12$)	0.764 ($\sigma = 0.125$)
LP_SAMP(ARITH, 1, 1)	0.958 ($\sigma = 0.0419$)	0.903 ($\sigma = 0.0686$)
LP_SAMP(GEOM, 2, 1)	0.866 ($\sigma = 0.103$)	0.819 ($\sigma = 0.121$)
LP_SAMP(GEOM, 4, 1)	0.815 ($\sigma = 0.124$)	0.772 ($\sigma = 0.14$)
LP_SAMP(GEOM, 8, 1)	0.794 ($\sigma = 0.129$)	0.752 ($\sigma = 0.144$)
LP_SAMP(GEOM, 2, 2)	0.841 ($\sigma = 0.122$)	0.796 ($\sigma = 0.139$)
LP_SAMP(GEOM, 4, 4)	0.819 ($\sigma = 0.132$)	0.776 ($\sigma = 0.148$)
0.05-approx	0.995 ($\sigma = 0.0128$)	0.938 ($\sigma = 0.0528$)
0.2-approx	0.985 ($\sigma = 0.0166$)	0.929 ($\sigma = 0.0499$)

Applications with 100 instances

Heuristic	Normalized to best	Normalized to UB
ON-DEMAND	0.879 ($\sigma = 0.101$)	0.788 ($\sigma = 0.104$)
ROUND-ROBIN	0.781 ($\sigma = 0.119$)	0.702 ($\sigma = 0.124$)
LP_SAMP(ARITH, 1, 1)	0.951 ($\sigma = 0.0448$)	0.853 ($\sigma = 0.0691$)
LP_SAMP(GEOM, 2, 1)	0.793 ($\sigma = 0.127$)	0.713 ($\sigma = 0.129$)
LP_SAMP(GEOM, 4, 1)	0.718 ($\sigma = 0.142$)	0.647 ($\sigma = 0.146$)
LP_SAMP(GEOM, 8, 1)	0.696 ($\sigma = 0.144$)	0.627 ($\sigma = 0.149$)
LP_SAMP(GEOM, 2, 2)	0.734 ($\sigma = 0.145$)	0.661 ($\sigma = 0.148$)
LP_SAMP(GEOM, 4, 4)	0.698 ($\sigma = 0.147$)	0.629 ($\sigma = 0.152$)
0.05-approx	0.98 ($\sigma = 0.034$)	0.879 ($\sigma = 0.0618$)
0.2-approx	0.98 ($\sigma = 0.0308$)	0.878 ($\sigma = 0.0596$)

Applications with 5000 instances

Heuristic	Normalized to best	Normalized to UB
ON-DEMAND	0.863 ($\sigma = 0.112$)	0.84 ($\sigma = 0.109$)
ROUND-ROBIN	0.779 ($\sigma = 0.127$)	0.758 ($\sigma = 0.125$)
LP_SAMP(ARITH, 1, 1)	0.984 ($\sigma = 0.0241$)	0.958 ($\sigma = 0.0245$)
LP_SAMP(GEOM, 2, 1)	0.935 ($\sigma = 0.0492$)	0.91 ($\sigma = 0.0489$)
LP_SAMP(GEOM, 4, 1)	0.899 ($\sigma = 0.0713$)	0.875 ($\sigma = 0.0706$)
LP_SAMP(GEOM, 8, 1)	0.88 ($\sigma = 0.0818$)	0.857 ($\sigma = 0.081$)
LP_SAMP(GEOM, 2, 2)	0.928 ($\sigma = 0.0456$)	0.903 ($\sigma = 0.0454$)
LP_SAMP(GEOM, 4, 4)	0.91 ($\sigma = 0.0566$)	0.886 ($\sigma = 0.0565$)
0.05-approx	0.999 ($\sigma = 0.00442$)	0.972 ($\sigma = 0.00563$)
0.2-approx	0.986 ($\sigma = 0.0102$)	0.959 ($\sigma = 0.0109$)

Simulations with large relative deviations ($\nu^{(k)} = 10.7$)

Heuristic	Normalized to best	Normalized to UB
ON-DEMAND	0.986 ($\sigma = 0.00707$)	0.953 ($\sigma = 0.00768$)
ROUND-ROBIN	0.728 ($\sigma = 0.0993$)	0.704 ($\sigma = 0.0976$)
LP_SAMP(ARITH, 1, 1)	0.983 ($\sigma = 0.0126$)	0.95 ($\sigma = 0.0143$)
LP_SAMP(GEOM, 2, 1)	0.975 ($\sigma = 0.0165$)	0.942 ($\sigma = 0.0172$)
LP_SAMP(GEOM, 4, 1)	0.955 ($\sigma = 0.0293$)	0.923 ($\sigma = 0.0297$)
LP_SAMP(GEOM, 8, 1)	0.926 ($\sigma = 0.0419$)	0.895 ($\sigma = 0.0415$)
LP_SAMP(GEOM, 2, 2)	0.948 ($\sigma = 0.0327$)	0.916 ($\sigma = 0.0331$)
LP_SAMP(GEOM, 4, 4)	0.9 ($\sigma = 0.0617$)	0.87 ($\sigma = 0.0612$)
0.05-approx	0.995 ($\sigma = 0.0221$)	0.963 ($\sigma = 0.025$)
0.2-approx	0.992 ($\sigma = 0.00587$)	0.958 ($\sigma = 0.00781$)

Overall results

Heuristic	Normalized to best	Normalized to UB
ON-DEMAND	0.87 ($\sigma = 0.108$)	0.821 ($\sigma = 0.109$)
ROUND-ROBIN	0.779 ($\sigma = 0.123$)	0.736 ($\sigma = 0.126$)
LP_SAMP(ARITH, 1, 1)	0.971 ($\sigma = 0.0362$)	0.917 ($\sigma = 0.0651$)
LP_SAMP(GEOM, 2, 1)	0.875 ($\sigma = 0.106$)	0.829 ($\sigma = 0.122$)
LP_SAMP(GEOM, 4, 1)	0.819 ($\sigma = 0.13$)	0.777 ($\sigma = 0.144$)
LP_SAMP(GEOM, 8, 1)	0.795 ($\sigma = 0.136$)	0.754 ($\sigma = 0.149$)
LP_SAMP(GEOM, 2, 2)	0.842 ($\sigma = 0.129$)	0.799 ($\sigma = 0.144$)
LP_SAMP(GEOM, 4, 4)	0.812 ($\sigma = 0.139$)	0.771 ($\sigma = 0.153$)
0.05-approx	0.993 ($\sigma = 0.022$)	0.937 ($\sigma = 0.0555$)
0.2-approx	0.985 ($\sigma = 0.0201$)	0.93 ($\sigma = 0.0513$)

Conclusion

- ▶ Always worth to distinguish applications

- ▶ Further splitting worthwhile if
 - ▶ Lots of instances
 - ▶ Comparable communication and computation costs
 - ▶ Communication-to-computation ratio depends of communication volume