

Table des matières

1	A la recherche de performances	1
1.1	L'organisation de projet	1
1.1.1	Ne jamais optimiser prématurément	1
1.1.2	Le processus de développement	2
1.1.3	Bonnes pratiques de développement	3
1.2	Les différentes approches en R pour gagner en performances	9
1.2.1	Apprendre à mesurer les performances temps et mémoire	9
1.2.2	Améliorer/optimiser le code R	15
1.2.3	Implémenter les points chauds de calcul avec des langages compilés	21
1.2.4	Utiliser plusieurs unités de calcul	24
2	Fondamentaux du calcul parallèle	25
2.1	Évolution des ordinateurs et nécessité du calcul parallèle	25
2.2	Les architectures parallèles	27
2.2.1	Éléments de vocabulaire autour du processeur	28
2.2.2	Éléments de vocabulaire autour de la mémoire	29
2.2.3	Typologie des infrastructures de calcul	35
2.3	Le calcul parallèle	37
2.3.1	Éléments de vocabulaire	37
2.3.2	Penser parallèle	39
2.4	Limites aux performances	41
2.4.1	Loi d'Amdahl	41
2.4.2	Loi de Gustafson	42
2.4.3	Et dans la vraie vie	43
3	Calcul parallèle avec R sur machine multi-cœurs	45
3.1	Principe général	45
3.2	Le package <code>parallel</code> et son utilisation	46
3.2.1	L'approche <code>snow</code>	48
3.2.2	L'approche <code>multicore</code>	55
3.2.3	<code>foreach</code> + <code>doParallel</code>	60

xii	Calcul parallèle avec R	
	3.2.4 Génération de nombres pseudo-aléatoires	62
3.3	L'équilibrage de charge	63
	3.3.1 Durée des tâches connue et ordonnancement statique	65
	3.3.2 Durée des tâches inconnue et ordonnancement dynamique	66
	3.3.3 Granularité et équilibrage de charge	70
4	C++ parallèle interfacé avec R	73
4.1	Principe général	73
4.2	openMP	78
	4.2.1 Les bases d'openMP	78
	4.2.2 R et openMP	82
4.3	Les threads de C++11	84
	4.3.1 Les bases des threads de C++11	84
	4.3.2 R et les threads de C++11	85
4.4	RcppParallel	87
5	Calculs et données distribués avec R sur un cluster	89
5.1	Cluster orienté HPC	89
	5.1.1 Les bases de MPI	90
	5.1.2 R et MPI	92
5.2	Cluster orienté <i>big data</i>	100
A	Notions complémentaires	103
A.1	Problématique du calcul flottant	103
A.2	La complexité	104
A.3	Typologie des langages	106
A.4	Les accélérateurs	107
	A.4.1 GP-GPU, General-Purpose computation on Graphic Processing Unit	107
	A.4.2 Carte many-cœurs	109
A.5	Exemples d'utilisation de Rcpp avec la fonction <code>cppfunction</code> du package inline	109
	Bibliographie	114
	Index	115