

Rapport INRIA 1994 — Programme 1

Architecture

ACTION ARCHI

3 mai 1995

ACTION ARCHI

Architecture

Localisation : *Rocquencourt*

Mots-clés : architecture d'ordinateur (1), optimisation (2, 4), réécriture (3), réseaux de transports hiérarchisés (6), routage (1).

1 Composition de l'équipe

Responsable scientifique

Dimitri Tusera, directeur de recherche

Personnel INRIA

Dominique Fortin, chargé de recherche

Chercheurs doctorants

Polina Strogova, boursière M.A.E., université de Nancy I
Alexandre Tusera, boursier Inria, université de Versailles

2 Présentation de l'action

Les architectures de machines parallèles peuvent être regroupées en deux classes :

- les architectures qui partagent des ressources (bus et mémoire) ou machines fortement couplées,
- les architectures pour lesquelles les processeurs topologiquement voisins ne partagent que les connexions, ou machines faiblement couplées.

L'action Architecture s'intéresse aux problèmes de routage liés aux machines faiblement couplées. Au contraire des algorithmes généraux qui ne tiennent pas compte de la structure du réseau d'interconnexion, le but recherché est, ici, de concevoir des algorithmes respectant au mieux cette structure ; ainsi, l'algorithme lui-même trouvera-t-il une implémentation naturellement parallèle sur le réseau en question.

3 Actions de recherche

3.1 Routage dans une grille multidimensionnelle

Participants : Alexander Tusera, Dominique Fortin

Une stratégie *divide-and-conquer*, respectant fortement la structure d'une grille bi-dimensionnelle, conduit à diviser le problème en 4 sous-problèmes de taille moitié moindre et respectant certaines conditions de compatibilité aux limites de subdivision. La mise en équation de ce problème d'optimisation conduit à un problème d'assignation linéaire avec contraintes. La levée des contraintes s'effectue par augmentation de la dimension du problème, autorisant ainsi une présentation sous forme d'assignation linéaire standard mais pour des matrices creuses.

Parmi les méthodes heuristiques déterministes pour résoudre ce problème NP-hard, les trois suivantes sont plus particulièrement connues : relaxation linéaire, relaxation Lagrangienne couplée avec un subgradient et enfin *branch-and-bound*. Puisque la taille de notre problème est grande (> 100), d'un point de vue pratique, seule la relaxation Lagrangienne est utilisable.

La méthode standard de subgradient conduisant très souvent à l'échec (phénomène de zig-zag entre plusieurs optima locaux), nous étudions les moyens d'enrichir celle-ci en nous référant à la structure polyédrale du problème (facettes associées au polytope)[4]. Plus particulièrement, étant données une solution relaxée et une solution faisable, nous approchons le polytope original dans un voisinage défini par les facettes *locales* au couple, c'est-à-dire les facettes violées par la solution relaxée mais satisfaites par la solution faisable. Outre la diminution considérable du nombre de facettes introduites, cette méthode présente l'avantage de modifier dynamiquement la perception du polytope vis-à-vis du voisinage, au cours de la trajectoire suivie par le subgradient.

Cette méthode constitue une alternative à la méthode *bundle-trust* qui approxime le subgradient tout en conservant le polytope original [3]. Bien que nos expérimentations [5] concluent favorablement à l'approximation du polytope en termes de difficulté de calcul (*bundle-trust* requiert la résolution d'un problème d'optimisation intermédiaire quadratique alors que tout reste linéaire dans le premier cas) et de vitesse de convergence, les deux méthodes doivent être considérées comme extrêmes dans la palette des méthodes de subgradient et leurs différentes combinaisons, mériteraient d'être étudiés.

Enfin, une approche statique de l'approximation du polytope est également examinée pour éviter les difficultés de la gestion dynamique des voisinages au cours des itérations successives. Des expérimentations ont montré que la version statique conduisait également à de bons résultats.

L'objectif que nous poursuivons vise à établir des compromis efficaces entre les approximations globales (statiques) et locales (dynamiques) et à comparer les résultats expérimentaux avec les autres méthodes évoquées.

Ce travail s'effectue en collaboration avec C. Roucairol, de l'Université de Versailles, pour ce qui concerne les aspects d'optimisation en entiers, liés au problème d'assignation linéaire.

3.2 Routage dans des réseaux de Cayley

Participants : Polina Strogova, Dominique Fortin

Un autre pôle d'intérêt de l'action concerne le problème de routage dans des réseaux basés sur la structure d'un groupe. Ce travail est conjoint avec l'équipe Prothéo de l'INRIA Lorraine.

Pour un traitement parallèle efficace, il est important de posséder un réseau d'interconnexion performant, minimisant les délais de communication (le diamètre) et le matériel (le nombre de connexions pour chaque nœud). Les réseaux dits de Cayley, basés sur la structure d'un groupe, tiennent un record en ce qui concerne leur taille par rapport au degré et au diamètre. Par exemple, l'hypercube 9-dimensionnel contient $2^9 = 512$ nœuds, tandis qu'il existe un graphe de Cayley avec les mêmes degré et diamètre, ayant 4773696 nœuds, soit dix mille fois plus ! Cette propriété des graphes de Cayley est un avantage important quant à la conception des architectures. Une autre propriété prometteuse est leur symétrie, puisqu'elle permet d'éviter les encombrements sur le réseau. De plus, la

présence d'une structure de groupe offre une possibilité de traitement automatique.

Notre ambition est de considérer de très grands groupes, pour lesquels les techniques traditionnelles de routage ne sont plus efficaces. Nous proposons une méthode originale, utilisant des techniques de réécriture [2].

Les données sur le réseau sont codées en un système d'équations sur les éléments du groupe dont on ne connaît pas de présentation. La complétion du système permet d'obtenir, en particulier, une présentation minimale pour l'ensemble générateur donné (on notera ici la différence avec le problème de recherche d'un système canonique associé à un groupe dont on connaît une présentation: problème déjà abondamment traité dans la littérature). Un routage minimal dans le réseau est obtenu par normalisation d'un terme schématisant le chemin recherché.

Bien que le problème semble être résolu théoriquement, la réalisation pratique de la méthode pose de nombreux problèmes. Ils sont liés à la génération, lors du calcul d'un système canonique, d'un nombre exponentiel de paires critiques [1].

Notre idée est donc de décomposer le problème, suivant une chaîne stabilisatrice du groupe initial, et faire de la complétion incrémentale. Pour arriver au système canonique et construire une présentation minimale, nous comptons exploiter au mieux l'arsenal moins coûteux de simplification et de complétion close, ainsi que les lois de simplification forte, valides dans les groupes.

3.3 Matrices de Monge faibles

Participants : Dominique Fortin

Pour un semi-groupe commutatif et totalement ordonné (H, \oplus, \preceq) pour lequel \oplus est compatible avec \preceq , c'est-à-dire :

$$a \preceq b \implies a \oplus c \preceq b \oplus c, \forall a, b, c \in H$$

une matrice $m \times n$ C de coûts dans H , est une matrice de Monge ssi elle satisfait les conditions suivantes :

$$c[i, j] \oplus c[r, s] \preceq c[i, s] \oplus c[r, j] \quad \forall 1 \leq i < r \leq m, 1 \leq j < s \leq n$$

Les différentes classes de matrices obtenues par spécialisation de l'opération correspondent à des situations pour lesquelles la solution optimale

des problèmes d'optimisation associés est connue. Par exemple, pour $(H, \oplus, \preceq) = (\mathbb{R}, +, \leq)$, si la matrice C possède la propriété de Monge, alors la solution optimale du problème de transport associé:

$$\begin{aligned} \min_x z &= \sum_{i=1}^m \sum_{j=1}^n c[i, j] x_{ij} \\ \sum_{j=1}^n x_{ij} &= a_i^1, \forall i \in [1, m] \\ \sum_{i=1}^m x_{ij} &= a_j^2, \forall j \in [1, n] \\ x_{ij} &\geq 0 \forall i, j \end{aligned}$$

est donnée par la diagonale.

Une autre propriété algébrique, essentielle à la classification, réside dans la simplification permise par l'opération \oplus :

- simplification forte

$$a \oplus c = b \oplus c \implies a = b, \forall a, b, c \in H$$

- simplification faible

$$a \oplus c = b \oplus c \implies a = b \text{ ou } a \oplus c = c, \forall a, b, c \in H$$

Dans l'exemple précédent, si l'on remplace $+$ par \max , la simplification passe de forte à faible mais le même résultat demeure quant à l'optimum pour une fonction objective *bottleneck*, pourtant réputée plus difficile qu'une linéaire.

Une restriction des conditions de Monge aux seuls éléments diagonaux de la matrice ($j = i$ dans les conditions) conduit à la définition de matrices de Monge faibles pour lesquelles la même subdivision, suivant la propriété de simplification, peut être effectuée. Cette dernière classe correspond justement aux problèmes d'affectation linéaire mentionnés dans les paragraphes précédents (problèmes de transport pour lesquels les seconds membres a_i^1 et a_j^2 sont tous égaux à 1).

La propriété de compatibilité de l'opération \oplus induit une structure de cône :

$$\begin{aligned} A, B \in \mathcal{C} &\implies A \oplus B \in \mathcal{C} \\ A \in \mathcal{C} &\implies \lambda A \in \mathcal{C} \end{aligned}$$

pour les classes décrites ci-dessus. L'université de Graz (Autriche) a déjà abordé, depuis quelques années, l'étude de la structure du cône

des matrices de Monge et établi, notamment, la liste de ses arêtes. En collaboration avec R. Rudolf, co-auteur des résultats sur le cas fort, nous prolongeons ce travail au cas des matrices de Monge faibles, où la situation est plus compliquée, à cause du nombre de contraintes restreint par rapport au cas fort. La généralisation au cas multidimensionnel, et non plus seulement 2D, est également étudiée.

4 Actions industrielles

Programme Praxitèle

Participants : Alexander Tusera

Ce programme propose de porter l'accent sur trois grands domaines :

- la modélisation des réseaux de transport hiérarchisés,
- le contrôle dynamique des réseaux de transport,
- le développement de nouveaux modes de transport,

en faisant appel aux compétences de différents projets suivant les domaines cités.

Une vision macroscopique des transports qui s'attache aux flux des passagers et comment ces flux se répartissent entre différents modes suivant certains critères, est une approche nouvelle dans le domaine des transports. Dans ce cadre, les techniques de routage hiérarchiques que nous avons mentionnées précédemment trouvent tout naturellement leur intérêt pour le premier domaine du programme. Plus généralement, ce domaine nécessite l'usage des problèmes de transport ou d'affectation, bien connus de la R.O.; notre collaboration s'inscrit donc dans le sens de développement de méthodes et d'outils pour faire face aux problèmes soulevés.

5 Publications

Communications à des congrès, colloques, etc.

- [1] D. FORTIN, C. KIRCHNER, P. STROGOVA, «Routing in Regular Networks Using Rewriting», *in: Proceedings of the CADE international workshop on automated reasoning in algebra (ARIA)*, J. Slaney (éd.), p. 5-8, Nancy, June 1994.

- [2] D. FORTIN, P. STROGOVA, «Routing in Regular Networks Generated by Permutation Groups», in : *Proceedings of the international workshop on Cayley graphs in computer science*, Lyon, Décembre 1993.
- [3] A. TUSERA, D. FORTIN, «Revisiting Subgradient Techniques Applied to 3D Linear Assignment Problem», in : *4th Conference on Operations Research*, Berlin, September 1994.
- [4] A. TUSERA, «Facettes associées au dD Linear Assignment Problem Polytope», in : *Conference/Workshop on Algorithmic Approaches to Large and Complex Combinatorial Optimization Problems*, Giens, Mai 1994.

Rapports de recherche et publications internes

- [5] A. TUSERA, D. FORTIN, «Revisiting Subgradient Techniques Applied to 3D Linear Assignment Problem», *Rapport de Recherche n° 2266*, INRIA, Mai 1994.

6 Abstract

The ARCHI project aims at designing efficient routing algorithms on various interconnexion networks. Routing schemes heavily rely on the underlying structure of the given network, contrary to general purpose algorithms whose decision making occurs at a local level. Networks under studies are multi-dimensional meshes and Cayley graphs associated with permutation groups.

Table des matières

1	Composition de l'équipe	1
2	Présentation de l'action	1
3	Actions de recherche	2
3.1	Routage dans une grille multidimensionnelle	2
3.2	Routage dans des réseaux de Cayley	3
3.3	Matrices de Monge faibles	4
4	Actions industrielles	6
5	Publications	6
6	Abstract	7