SMP マシン上での BLAS ライブラリ用自動チューニング機構の設計と実装

A Design and Implementation of Auto-tuning Facilities for BLAS libraries on SMP Machines

木下靖夫[*1] 片桐孝洋[*1*2] 本多弘樹[*1] 弓場敏嗣[*1]
Yasuo Kinoshita Takahiro Katagiri Hiroki Honda Toshitsugu Yuba

[*1] 電気通信大学大学院情報システム学研究科 Graduate School of Information Systems
[*2] 科学技術振興機構さきがけ PRESTO, JST
The University of Electro-Communications

1. はじめに

数値計算処理を高速に行うためには、適切に数値計算ライブラリのパラメータ設定を行うことが必要である。現在これらを自動的にチューニングする機構が開発されている。自動チューニングを行うソフトウェアには、インストール時最適化型と実行時最適化型の２種類が知られており、前者では ATLAS[2]、後者では Autopilot が知られている。SMP マシン上ではパラメータを事前に設定するインストール時最適化型が適しており、ATLAS では数値計算ライブラリ BLAS を自動並列化コンパイラを用いて並列化することで高速化を行っている。本稿では ATLAS にスレッドライブラリを導入し、BLAS を SMP マシンに適するように再帰化する AutoTuned-RB(Recursive BLAS)方式を提案する。

2．AutoTuned-RB の概要

2.1 再帰 BLAS[1]

プログラムを再帰化表現し、再帰終了時、すなわち再帰ノードで BLAS ライブラリをコールする。再帰ノードの割り当てを SMP マシンのプロセッサ台数に合わせてスレッド化することで、効率良く並列計算することが出来る。

2.2 スレッドライブラリの追加

ATLAS にスレッドライブラリのサポート機能を加える。ATLAS はコンパイラによる自動並列化の機能を備えているが、並列実行環境によってはパフォーマンスが悪くなることもある。ここでは Posix Pthread ライブラリによる並列化を行う。

2.3 パラメータチューニング

ATLAS では BLAS をチューニングする際に、キャッシュサイズに合わせたブロック幅を微小に変化させ、最も速くなるブロック幅パラメータを使用する。本方式では再帰の深さのパラメータを追加し、プロセッサ数または指定されたプロセッサ数よりも少ない数で最も速くなるように調整する機能を追加する。

3. 提案する自動チューニング機構

ATLAS は次の手順でインストールを行う。

i．環境設定ファイルの作成
ii. BLAS のインストール

手順 i では CPU のアーキテクチャ、C と Fortran コンパイラ、およびコンパイラオプションを選択する。その環境設定ファイルを元に手順 ii で BLAS ライブラリを実行し Flops 値を測定する。その結果を基に最適なパラメータを選択する。

AutoTuned-RB では ATLAS のインストール処理に加え、下記の 4 つの処理を追加する。

① スレッドライブラリをサポートするか否かの選択
② プロセッサ数入力
③ 実測によるパラメータの選択
④ 再帰 BLAS ライブラリチューニング機能

このフローチャートを図１に示す。

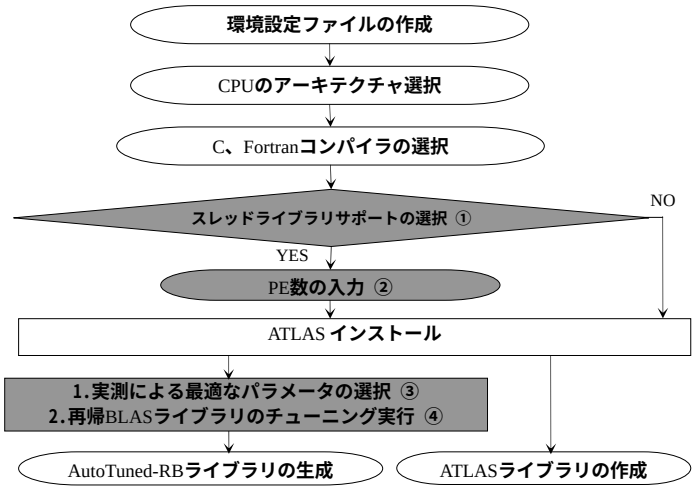

図 1：AutoTuned-RB の流れ図

4. 予備評価

予備評価として電気通信大学総合情報処理センタの Origin3400 を 16PE 用いた。ATLAS により最適化した BLAS を利用した再帰 BLAS について、ATLAS BLAS、通常の BLAS との比較を行った。測定プログラムは行列積ルーチン GEMM でサイズ 1000×1000 であり、再帰 BLAS のカーネルでは ATLAS BLAS をコールする。コンパイラは IRIX C コンパイラでオプションは”-64 -O3”を用いた。

表２：再帰 BLAS と ATLAS BLAS との比較

ライブラリ	Times(s.)	FLOPS	ピーク性能に対する効率(%)
BLAS	11.71	170.8	1.1%
ATLAS BLAS	0.31	6425.8	40.2%
再帰 BLAS	2.46	813.0	5.1%

表２から再帰 BLAS は、コンパイラによる自動並列化では高速化できないといえる。再帰 BLAS は 75%程度のピーク性能に対する効率が出ることが確認されている[1]ので、スレッド化することにより ATLAS BLAS を上回る性能の達成が期待される。

参考文献

[1]F. Gustavson, A. Henriksson, I. Jonsson, B. Kågström and P. Ling, “ Recursive Blocked Data Formats and BLAS’s for Dense Linear Algebra Algorithms,” Proceedings of PARA98, LNCS, Vol.1541, pp.195-206, Springer,1998.

[2]R.Clint Whaley, A. Petitet, Jack J. Dongarra, “Automated Empirical Optimizations of Software and the ATLAS project, ” Parallel Computing, Vol.27, pp.3-35, 2001