2J-02

共有メモリ環境における Matrix Powers Kernel の評価

黒田勝汰†　　藤井昭宏†　　田中輝雄†

工学院大学†

1　はじめに

科学技術シミュレーションで用いられるクリロフ部分空間法は{Ar,…,A^nr}の計算に疎行列ベクトル積（SpMV）を用いる．一般的に SpMV の性能はベクトルに対するメモリアクセス性能に制約される．近年，Demmel らがメモリアクセスを改善する手法である Matrix Powers Kernel（MPK）[1]を提案している．MPK は計算を指数方向も含めてブロック化することでキャッシュを効率的に利用する手法で，ブロックの重複により計算量が増えるが，メモリアクセスが改善できる．また，並列度の低い処理を並列化する手法として Multicolor ordering が知られている．これは同時実行可能な処理に同じ色を割り当て，処理の並列性を抽出する手法である．

本研究では 2 次元 5 点差分，3 次元 7 点差分に対しMPK と Multicolor ordering を組合せた手法を提案し共有メモリ環境で実装，評価した．

2　Matrix Powers Kernel

MPK は{Ar,…,A^nr}の計算で A^nr を分割し，その計算に必要な要素を一度に計算しキャッシュヒット率をあげる手法である．

要素数 14，n=5 の 1 次元 3 点差分での MPK の計算を図 1 に示す.行は A^ir の要素である.白点，黒点はA^5rの分割された要素とその計算に必要な要素をあらわす.灰色の点は黒点と白点の領域の重複部分であり，重複して計算する．A^ir の j 番目の要素を更新する際は A^{i-1}r の j-1，j，j+1 番目の要素が計算されている必要がある．

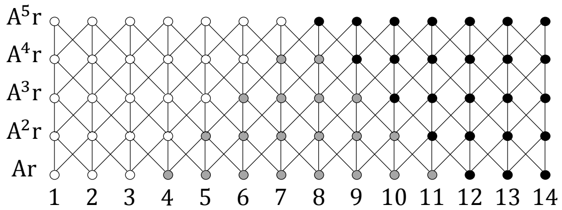

図 1　1 次元 3 点差分での MPK の計算

An Evaluation of Matrix Powers Kernel in Shared Memory Enviroment
Shota Kuroda†, Akihiro Fujii† and Teruo Tanaka†
†Kogakuin University

3　提案する手法

本研究では Multicolor ordering を用いて計算順序を変え，MPK の重複計算を減らす手法を提案し，有限差分法で用いられる 2 次元 5 点差分，3 次元 7 点差分に適用した．疎行列は CRS 形式（Compressed Row Storage）[2]で保持する．

3.1　領域の分割方法

2 次元 5 点差分では A^nr に相当する領域を Xbs×Ybs のブロックに分割した．3 次元 7 点差分では領域を Xbs×Ybs×Zbs のブロックで分割した．2 次元の分割を図 2 に示す．図 2 は左から分割された計算領域, 分割されたブロックを計算するための A の指数も含めた計算領域である．右図の横軸，奥行き軸は領域の X 軸，Y 軸を，高さ軸は A の指数 n である．

3.2　色分け方法

重複計算が残る色分け A と重複計算が発生しない色分け B の 2 種類の方法を提案する．2 次元 5 点差分における色分けを図 3 に示す．図 3 は左から色分けなし，色分け A，色分け B のときのそれぞれの色のブロックと重複部分である．領域に付けられた番号は色番号である．

色分け B では Xbs または Ybs が 2n より小さい場合重複計算が発生するので，Xbs，Ybs は 2n 以上となるように設定する．色番号は実行順序をあらわす．3 次元 7 点差分では重複計算が発生するため色分け B の色数を 4 色に増やした．

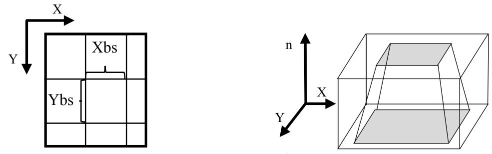

図 2　2 次元 5 点差分での
領域分割と計算範囲

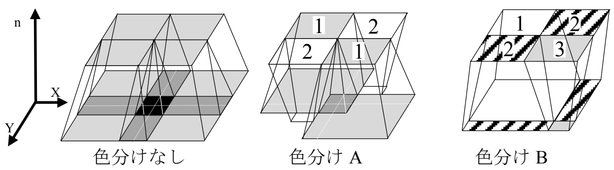

図 3　2 次元差分でのそれぞれの
色分け方法における計算領域と重複部分

4　実験

実験には Intel Core i7-4770 （L3 cache 8MB）, gcc 4.4.7 を用いた．コンパイルオプションは“-O3 -fopenmp”を用いた．スレッド並列化する際，ブロックの計算処理を OpenMP によるスレッド並列で処理し，スレッド数は 8 とした．

4.1　2 次元 5 点差分

図 4 に各色分けにおいて Xbs，Ybs を最小値，最大値とその間の 2 のべきの値をとるように変化させたときの最大性能と通常の SpMV により計算したときの性能を示す．このとき，n は 2 から 64，領域サイズは 2048×2048 である．各色分けの最大性能時のパラメタを表 1 に示す.

ほとんどの n について色分け B，色分け A，色分けなしの順に性能がよい. 重複計算の減少量に反比例していると考えられる．

色分け A から色分け B への性能向上が色分けなしから色分け A へのものと比べて小さいのは重複計算の減る量の差によるものである．実際，色分けなしと色分け A の最大性能のときの重複計算増加率は約 12%，約 2.6%であり，色分け A の重複計算量は十分小さい.

すべての MPK で n が大きくなると性能が低下する傾向が見られる．これは n が大きいと，指定できる最小のブロックサイズが大きく，キャッシュに収まらないためである．

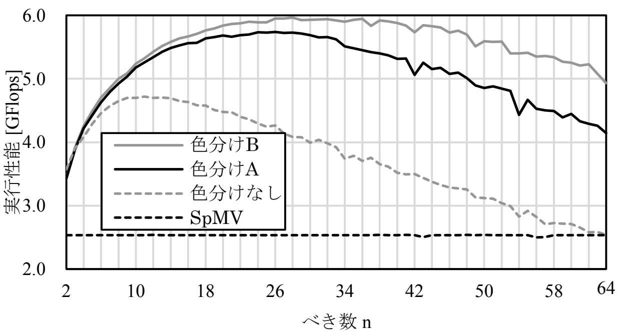

図 4　2 次元 5 点差分での
最適なパラメタのときの性能

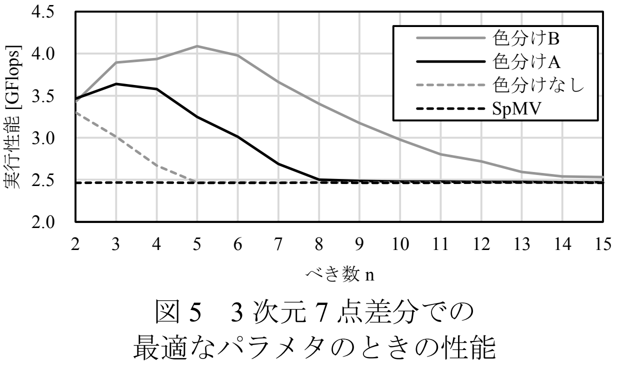

図 5　3 次元 7 点差分での
最適なパラメタのときの性能

表 1　2 次元 5 点差分における各色分けの
最大性能時のパラメタ

色分け方法	色分けなし	色分け A	色分け B
n	11	26	28
Xbs	256	512	512
Ybs	128	64	56
重複計算増加率[%]	約 12	約 2.6	0.0

表 2　3 次元 7 点差分における各色分けの
最大性能時のパラメタ

色分け方法	色分けなし	色分け A	色分け B
n	2	3	5
Xbs	256	256	256
Ybs	8	8	10
Zbs	16	16	16
重複計算増加率[%]	約 19	約 0.073	0.0

4.2　3 次元 7 点差分

図 5 に各色分け方法において Xbs，Ybs，Zbs を最小値，最大値とその間の 2 のべきの値をとるように変化させたときの最大性能と通常の SpMV により計算したときの性能を示す．n は 2 から 15，領域サイズは 256×256×256 である．各色分けの最大性能時のパラメタを表 2 に示す.

全体的な傾向は 2 次元 5 点差分と同じである．しかし，2 次元に比べ 3 次元では分割された $A^n r$ を計算するために必要な領域が大きくなり，キャッシュに収まらなくなる．また重複計算量の増加率も 2 次元に対して大きくなるので性能の下がり始める n が小さくなっている．

5　おわりに

本研究では MPK と Multicolor ordering を組み合わせ，重複計算は残るが色数が少なくなる色分け（色分け A），重複計算は発生しないが色数が多くなる色分け（色分け B）を提案し実装した．2 次元 5 点差分においては色分け B によって従来手法と比べて約 1.26 倍の性能が得られた．3 次元 7 点差分においては色分け B によって従来手法と比べて約 1.24 倍の性能が得られた．色分けによる重複計算削減量が多い色分け B の性能向上の効果が高いことがわかった．また今後の課題として，本手法を一般の疎行列に対して拡張したときの効 1 果を検証する必要がある．

参考文献

[1] Demmel, J., et al.: Avoiding Communication in Sparse Matrix Communications, in IEEE International Parallel and Distributed Processing Symposium(2008)

[2] Barret, R., et al.: Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, SIAM pp. 57-65 (1944)