

## 研究課題別評価

### 1 研究課題名： 特異値分解法の革新による実用化基盤の構築

### 2 研究者氏名： 中村佳正

研究員： 峯崎征隆（研究期間 H.15.4 ～ H.16.5）

研究員： 高田雅美（研究期間 H.16.4 ～ H.18.3）

研究員： 岩崎雅史（研究期間 H.16.8 ～ H.18.3）

### 3 研究のねらい：

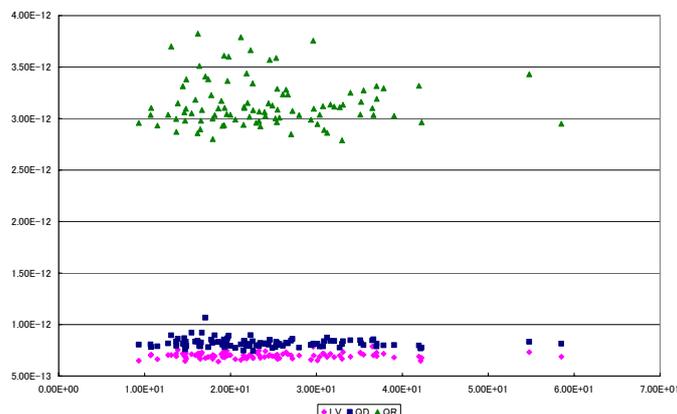
MATLAB、Mathematica など汎用ソフトウェアで使われる行列の特異値分解ルーチンは QRs 法に基づく LAPACK の DBDSQR である。QRs 法の計算量は  $O(N^3)$  であり大規模行列では収束が遅く、一部の特異値・特異ベクトルのみを求めることも構造上困難である。本研究では、大規模情報処理の基盤技術となる新しい特異値分解法を開発し、その高速性、高精度性を理論的かつ実験的に検証する。具体的には、原点シフトの導入による特異値計算の高速化と高精度化、収束性の証明、収束次数の解明、誤差評価、前進後退の安定性の証明に加え、2重コレスキー分解による直交性に優れた特異ベクトルの高速計算法の研究を行う。さらに、一層の高速化を目指して特異値と特異ベクトルのそれぞれの並列計算法を開発する。また、大規模特異値分解のボトルネックを解消するため、新しい特異値分解法を実装した逐次型高速特異値分解ライブラリ群を構築し、特許化・ライセンス契約を経て公開するとともに、画像処理やデータ検索などその実用化研究を開始する。

### 4 研究成果：

#### [1] mdLVs アルゴリズムによる特異値計算ライブラリ DLVS の開発

3年間のさきがけ研究における成果の第一は、収束性と数値安定性が保証された新しい特異値計算法である「mdLVs アルゴリズム」を開発したことである。さらに、mdLVs アルゴリズムを実装したライブラリ DLVS を構築し様々なチューニングを行った。原点シフトの大きさをジョンソン境界より小さくとれば、mdLVs アルゴリズムは、丸め誤差があっても、非常に高い相対精度をもって必ず特異値に収束することを数学的に証明している。収束次数は最高で3次である。以下は数値実験の一部である。

[1-1] 図は 1000 次上2重対角行列をランダムに 100 個生成し、横軸にその条件数の対数を、縦軸に計算された 1000 個の特異値の相対誤差の総和を、DLVS (ピンク)、DLASQ (青)、DBDSQR (緑) で色分けしたものである。DBDSQR (緑) は信頼性の高い Demmel-Kahan シフト付き QR (QRs) 法を実装した LAPACK ルーチン、DLASQ (青) は dqds 法を実装した LAPACK ルーチンである。100 個の全てにおいて相対誤差の総和は「DLVS (ピンク) < DLASQ (青) << DBDSQR (緑)」となり、DLVS の高精度性が裏付けられる。特異値一個あたりの相対誤差はマシンイプシロン程度である。なお、テスト行列は、数式処理を用いた整数計算によりランダムな特異値をもつ上2重対角行列をテスト行列として誤差なしで生成しており実験の信頼性は高い。



図[1-1] 特異値の相対誤差の総和

[1-2] 計算時間(sec.)の比較は次の通りである。対象はランダムに生成した 10,000 次上2重対角行列 100 個である。DBDSQRと比べると十分に速く、DLASQとほぼ同等の速度で特異値を計算できる。

表[1-2] 特異値計算の実行時間(sec.)

Pentium 4, GNU compiler 2.95.4			Itanium2, Intel Fortran compiler 8.0		
DBDSQR	DLASQ	DLVS	DBDSQR	DLASQ	DLVS
18.22	4.73	9.62	-----	6.82	6.32

[2] I-SVD アルゴリズムによる特異値分解ライブラリ DBDSL<sub>V</sub> の開発

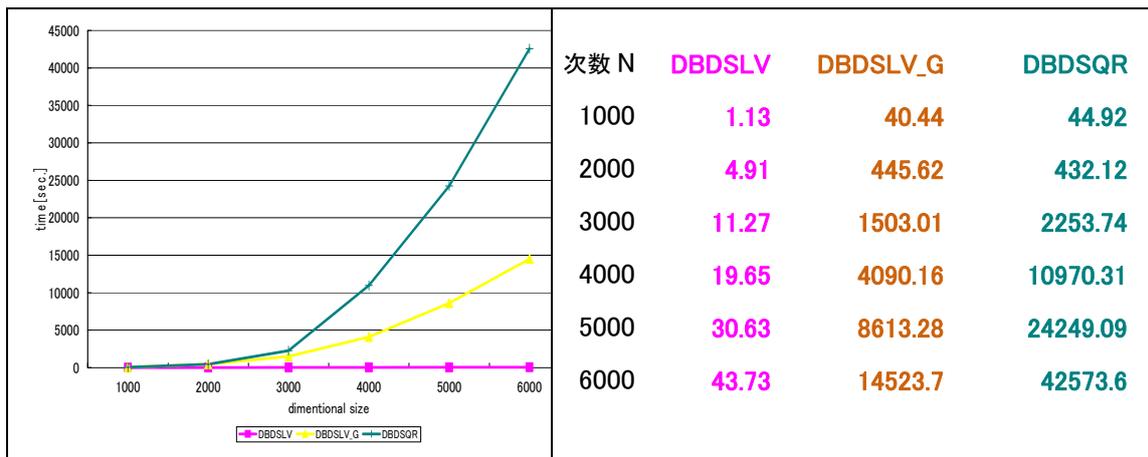
本研究の第二の成果は、高精度ツイスト分解法を開発し、与えられた行列とその特異値を用いて、特異ベクトルを  $O(N^2)$  の計算量で高精度かつ十分な直交性をもって計算できるようになったことである。悪条件の行列の2重コレスキー分解の計算に 2 種類の dLV 型変換を組み合わせ使用し、その結果、高精度なツイスト分解が可能となった。高精度ツイスト分解の成功により、mdLVs アルゴリズムによる特異値計算を前段階とする新しい特異値分解法「I-SVD アルゴリズム」が定式化された。特異値計算部と特異ベクトル計算部の分離により QRs 型特異値分解法と比べてはるかに高速に特異値分解が完了する。高精度ツイスト分解によって、単純な逆反復法や最近アメリカで開発されている MRRR 法と比べて、より高精度で数値安定なベクトル計算が可能となる。本研究では、DLVS ルーチンの特異値計算部とし、新しいツイスト分解の特異ベクトル計算部で行う I-SVD アルゴリズムを実装したライブラリ DBDSL<sub>V</sub> を開発した。以下は DBDSL<sub>V</sub> の評価実験の一部である。

[2-1] 1,000 次元行列の特異値分解の結果が次表である。U は左特異ベクトル行列、V が右特異ベクトル行列、 $\Sigma$  が特異行列、 $\hat{T}$  は転置、数値は誤差の 2 ノルムである。DBDSQR の特異ベクトルの直交性は良好であるが真値に対する誤差が大きい。一方、DBDSL<sub>V</sub> では特異ベクトルの直交性はやや劣るものの、真値に対する誤差は小さい。1 ノルム、 $\infty$  ノルムでは両者にはほとんど精度の差は現れない。また、特異ベクトルの直交性を重視して修正グラムシュミット法による再直交化を行うライブラリ DBDSL<sub>V\_G</sub> を開発した。

表[2-1] 特異ベクトルと特異値分解の誤差の比較

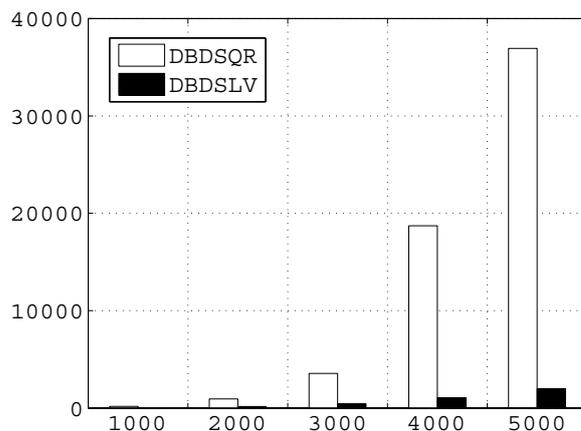
	DBDSL <sub>V</sub>	DBDSL <sub>V_G</sub>	DBDSQR
$V-V'$	$4.23 \times 10^{-9}$	$4.15 \times 10^{-9}$	$1.02 \times 10^{-7}$
$U-U'$	$4.14 \times 10^{-9}$	$3.92 \times 10^{-9}$	$1.02 \times 10^{-7}$
$V^{\hat{T}}TV-I$	$3.24 \times 10^{-10}$	$7.53 \times 10^{-12}$	$8.31 \times 10^{-11}$
$U^{\hat{T}}TU-I$	$3.15 \times 10^{-10}$	$8.65 \times 10^{-12}$	$8.28 \times 10^{-11}$
$B-U \Sigma V^{\hat{T}}$	$3.98 \times 10^{-9}$	$6.21 \times 10^{-11}$	$6.90 \times 10^{-11}$

[2-2] ランダムに生成した 1000 次から 6000 次上2重対角行列 100 個について、DBDSQR(青)、DBDSL<sub>V</sub>(ピンク)、DBDSL<sub>V\_G</sub>(黄)ルーチンによる特異値分解の平均計算時間(sec.)を比較したのが次の図表である。例えば、4000 次では、DBDSQR が 3 時間もかかるのに対して、DBDSL<sub>V</sub> では 20 秒である。DBDSL<sub>V</sub> の計算量が  $O(N^2)$  であることと良好なスケーラビリティをもつことが確認される。



図[2-2] 2重対角行列の特異値分解の実行時間(sec.)

[2-3] 与えられた密行列 A をハウスホルダー変換により上2重対角化(前処理)し、I-SVD アルゴリズムを実装した DBDSLV ルーチンにより上2重対角行列 B の特異値と特異ベクトルをすべて計算し、さらに行列を1次変換して元の行列 A の左右の特異ベクトルを計算する(後処理)ことで特異値分解が完了する。図は前処理・後処理まで含めた特異値分解の全工程について、現在の標準解法の DBDSQR(□)と DBDSLV(■)の実行時間の比較である。横軸は密行列の次数、縦軸は特異値分解完了までの総時間(sec.)である。DBDSLV ルーチンの優位性が顕著に現れている。全部の特異ベクトルが必要ではないとき、両者の差はもっと大きくなる。



図[2-3] 密行列の特異値分解の実行時間(sec.)

### [3] I-SVD アルゴリズムの並列化について

本研究の成果の第三はI-SVD アルゴリズムの並列化である。I-SVD アルゴリズムの特異ベクトル計算部分(ツイスト分解と逆反復)は個々の特異値に対する特異ベクトルを独立に計算するため、並列実行は容易である。しかしながら、特異ベクトル計算部のみ並列化した I-SVD では、特異値計算部は高速の mdLVs アルゴリズムであるが、逐次的構造のため、この部分の並列実行が難しい。特異値計算部は全実行時間全体の約 3 割を占めるため、全体の並列化効率は低いものとなっていた。そこで、本研究では、分割統治型特異値分解法の特異値計算部のみ取り出して二分木を用いて並列実行し、特異値計算部の並列度を高めることに成功した。特異値は 25 次程度の小規模な対称 3 重対角行列に対するニュートン法を利用した固有値解法に帰着する。特異ベクトル部は I-SVD アルゴリズムにより並列計算する。

表は「並列 I-SVD アルゴリズム」による上重対角行列の特異値分解の並列化効率と計算時間である。現在の標準並列特異値分解ルーチンの PSTEQR (ScaLAPACK、アメリカ)と比べても十分な高速性をもつことがわかる。写真は本研究で購入した PC クラスタである。

表[3] 並列特異値分解の実行時間(sec.)と並列化効率

アルゴリズム・ルーチン	CPUの数	1	2	4	8	16
“並列I-SVD”	3000次SVD(sec)	5.78	2.96	1.50	0.83	0.49
	Speed up	1.00	1.95	3.85	6.96	11.84
	並列化効率	1.00	0.98	0.96	0.87	0.75
	7000次SVD(sec)	37.86	19.27	9.81	5.07	2.81
	Speed up	1.00	1.96	3.86	7.47	13.47
	並列化効率	1.00	0.98	0.96	0.93	0.84
ベクトル計算部のみ	3000次SVD(sec)	5.53	3.44	3.36	1.80	1.52
並列化したI-SVD	Speed up	1.00	1.61	2.34	3.07	3.64
並列化したI-SVD	並列化効率	1.00	0.80	0.59	0.38	0.23
PSTEQR (ScaLAPACK) (ScaLAPACK)	3000次SVD(sec)	867.98	433.61	173.44	66.78	34.71

#### 5 自己評価:

本研究は、離散可積分系ロトカ・ボルテラ(dLV)方程式という従来にない方法による行列の特異値分解の実用化を目指して開始されたものである。新しい特異値分解法を開発し、その高速性、高精度性を理論的かつ実験的に検証するという目標はほぼ達成された。本研究で開発したのは次のアルゴリズムとライブラリ群である。(\*\*)は現在チューニング中のライブラリである。

	アルゴリズム	ライブラリ名	特徴
特異値計算	mdLVs	DLVS	収束証明、数値安定性、高精度
特異値分解	I-SVD	DBDSLVS	高速 $O(N^2)$ 、高精度ベクトル
分割統治特異値分解	dDC	(DBDDDC)	集積した特異値分布で高速
並列特異値分解	Parallel dDC	(PDBDDDC)	高い並列化効率

本研究で開発したライブラリ DBDSLVS は LAPACK の DBDSQR に対して、速度で圧倒し、特異ベクトルの直交性を除いて、特異値、特異ベクトルの精度で勝り、収束証明など信頼性は同水準である。当初の LAPACK レベルでの逐次型ライブラリ開発という目標はほぼ実現されたといえる。現在は、計画を早めて、分割統治ライブラリと並列ライブラリの構築とチューニングを行っている。

また、これらのライブラリはいずれも容易に対称行列の固有値分解に転用可能である。収束証明はないものの現在、最速最高精度とされ 2006 年リリース予定の LAPACK 4.0 の対称3重対角行列の固有値分解ルーチンとして採用される dqds 法、MRRR 法に対しては、mdLVs 法、I-SVD 法は、逐次計算の速度はやや劣るものの、信頼性、固有値、固有ベクトルの精度と直交性、並列化効率で勝る。

以上の結果、従来、数日が必要とされた CT、MRI、PET 等、医療画像処理に現れる 10000 次の密行列の特異値分解が I-SVD 法により数時間に短縮されるなど、幅広い情報処理基盤が整備されたと考えられる。なお、当初、研究計画にあげた実用化研究の開始のうち、画像処理ソフトウェアの開発についてはウェブ画圧縮への高速特異値分解の応用という形で研究が進められ、成果が出始めており、医療画像処理と合わせて、いずれ学会発表・論文発表の予定である。しかし、データ検索ソフトウェアの開発には多くの人手が要ることがわかり、研究の規模を考慮して、他のテーマの研究を優先することとした。

本研究は、それぞれ、並列計算機システムの構築、ライブラリのプログラミング、収束証明など理論研究と持ち味の異なる3名のポスドク研究員を雇用して研究を進めた。目標が達成できた影には彼らの大きな貢献があった。うち1名は既に大学助手、1名は大学助手に内定しており、別の1名は発展研究(SORST)の主力メンバーとして今後とも貢献を期待している。この3年間に科研費への応募が可能になり、雇用期間が平成18年3月まで延長されるなど、3名とも誇りと意欲をもって研究に専念することができたようである。参加したポスドク研究員のキャリア形成に大いに役に立ったことも本研究の成果である。

領域会議では研究総括やアドバイザーから毎回、実に多くのご指導をいただいた。領域会議の発表後、直ちに質疑をまとめ、京都で待つメンバーに電子メールで送って即日検討させてきた。研究の節目でいただいた有益なご指導が次の目標の設定に役立ち、それを達成することで、本研究の成果につながったのだと思う。研究総括とアドバイザーの方々に深く感謝する。

#### 6 研究総括の見解:

行列の特異値計算、特異値分解は様々な情報処理、情報検索の基礎となる重要な線形数値演算である。また、最近では、大規模な疎行列に限定した特異値を求めるために、新たな近似計算法が研究されてはいるものの、数値不安定性や精度悪化の問題が残り、その対処法は今のところ確立されていない。中村研究者は、これらの問題を解決するために、新しい特異値計算・特異値分解法の改良、実装、並列化等の研究開発を行った。本アルゴリズムは並列化が容易なので、計算の一層の高速化に顕著な効果が期待でき、時系列画像から対象物体の立体形状を求めるなど実用化の見通しが始まっており研究成果を高く評価できる。

#### 7 主な論文等:

##### [1] 成果発表論文(5件以内)

[1-1] M. Iwasaki and Y. Nakamura, An application of the discrete Lotka–Volterra system with variable step-size to singular value computation, *Inverse Problems*, Vol. 20(2004), 553–563.

[1-2] 高田雅美, 岩崎雅史, 木村欣司, 中村佳正, 高精度特異値計算ルーチンの開発とその性能評価, *情報処理学会論文誌*, Vol. 46, No. SIG12(2005), 299–311.

[1-3] M. Takata, M. Iwasaki, K. Kimura and Y. Nakamura, An evaluation of singular value computation by the discrete Lotka–Volterra system, *Proceedings of The 2005 International Conference on Parallel and Distributed Processing Techniques and Applications*, Vol. II, pp. 410–416.

[1-4] 岩崎雅史, 中村佳正, 特異値計算アルゴリズム dLV の基本性質について, *日本応用数理学会論文誌*, Vol. 15, No. 3(2005), 287–306.

[1-5] 岩崎雅史, 阪野真也, 中村佳正, 実対称3重対角行列の高精度ツイスト分解とその特異値分解への応用, *日本応用数理学会論文誌*, Vol. 15, No. 3(2005), 461–481.

##### [2] 特許

###### [2-1] 発明者: 中村佳正, 岩崎雅史, 阪野真也

発明の名称: 行列の高速高精度特異値分解法, プログラムおよび装置

出願人: 科学技術振興機構

出願日: 特願 2004-166437 号(平成 16 年 6 月 3 日)

PCT 出願日: PCT/JP2005/10084(平成 17 年 6 月 1 日)(米国, カナダ, EP(英, 独, 仏))

###### [2-2] 発明者: 中村佳正, 萱田太郎, 岩崎雅史, 阪野真也, 高田雅美

発明の名称: 特異値分解装置、及び特異値分解方法

出願人: 国立大学法人京都大学

出願日: 特願 2005-351089 号(平成 17 年 12 月 5 日)

[3] 受賞

[3-1] 2006 年情報処理学会ハイパフォーマンスコンピューティングと計算科学シンポジウム (HPCS2006)最優秀論文賞 受賞

論文題目:分割統治法とツイスト分解法による新しい特異値分解アルゴリズム

著者: 菅田太郎, 高田雅美, 岩崎雅史, 中村佳正

[4] 招待講演等

[4-1] Y. Nakamura, I-SVD: A new singular value decomposition algorithm with a high relative accuracy, GAMM Workshop Applied and Numerical Linear Algebra, FernUniversitat in Hagen, Germany, 平成 16 年 7 月 3 日.

[4-2] 中村佳正, 「可積分系から可積分アルゴリズムへー 一行列の特異値分解を中心としてー」, 平成 17 年 11 月 4 日, 応用数理学会秋の学校, 大阪大学サイバーメディアセンター.