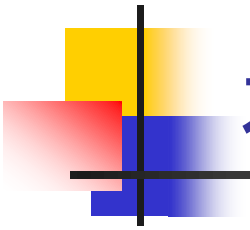
第2回CMSI人財育成シンポジウム

線形計算における誤差解析の事例

2013年12月2日

電気通信大学 情報理工学研究科
情報・通信工学専攻
山本有作

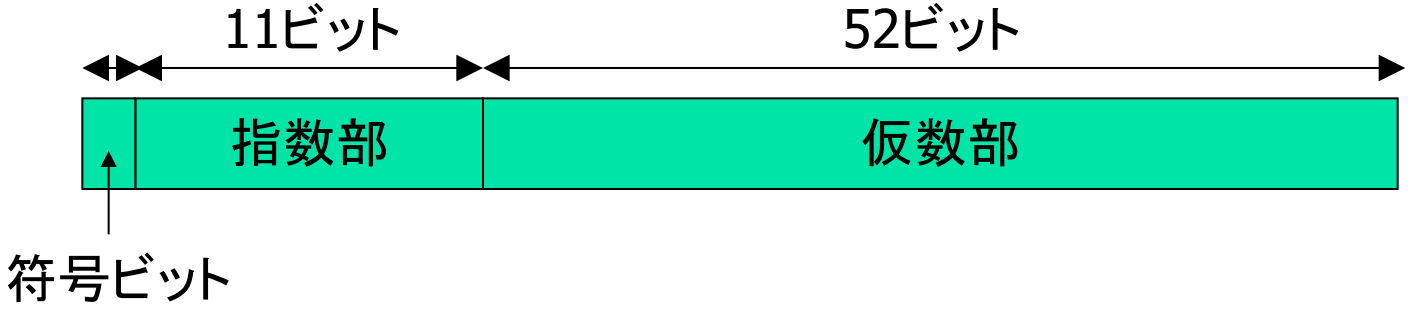


本講演の目的

- 丸め誤差の性質と誤差解析の基本的な手法を紹介する
  - 浮動小数点演算と丸め誤差
  - 前進誤差解析と後退誤差解析
  - 摂動論と条件数
- 基本的な線形計算について，誤差解析の結果を紹介する
  - 内積
  - ハウスホルダー変換
  - 固有値計算
- 並列化と誤差の振る舞いについて，実例を紹介する
  - TSQRアルゴリズムの誤差解析
  - MPI_Reduce における演算順序

浮動小数点による演算

- IEEE754
  - 1985年に制定された2進浮動小数点演算規格
  - 次のような内容について規定
  - 浮動小数点フォーマット，四則演算と丸め，浮動小数点例外などについて規定
  - Intel x86，SUN SPARC，IBM PowerPCなど，多くのCPUが採用



倍精度浮動小数点フォーマット

四則演算と丸め

- 丸め
  - ある実数 x を浮動小数点数で近似することを丸めと呼び，丸めた結果を $fl(x)$ と書く
  - IEEE754 では，最も近い浮動小数点数への丸めが既定値
- 四則演算に対する丸めと誤差
  - 2つの浮動小数点数 x, y に対する四則演算の結果 $x \odot y$（ただし $\odot = +, -, *, /$）は一般に浮動小数点にならない
  - IEEE754では，$x \odot y$ を正確に計算し，その結果を浮動小数点数に丸める
  - 最近点への丸めでは，次の式が成り立つ

$$fl(x \odot y) = (x \odot y)(1 + \delta), \quad |\delta| \leqq \mathbf{u}$$

  - ただし，$\mathbf{u}$ は丸め誤差の単位（IEEE754 の倍精度では $\mathbf{u} = 2^{-53}$）

簡単な計算の誤差解析

- $a/(b*c)$ の計算

$$
\begin{aligned}
fl(a/(b*c)) &= fl(a / ((b*c)(1+\delta_1))) \\
&= a / ((b*c)\,(1+\delta_1))\,(1+\delta_2) \\
&= a / (b*c)\,(1+\varepsilon)
\end{aligned}
$$

ただし，$|\varepsilon| = |(1+\delta_1) / (1+\delta_2) - 1| \fallingdotseq 2\mathbf{u}$

- すなわち，浮動小数点による $a/(b*c)$ の計算結果は，$2\mathbf{u}$ 程度の相対誤差を含む
- このように，計算結果に含まれる誤差を直接見積もる手法を，前進誤差解析と呼ぶ

便利な補題

- 誤差の表式の簡単化
  - 誤差解析では，$1+\delta$, $1/(1+\delta)$ などの因子の積がよく現れる
  - これを簡単化するために，次の補題[1]が使える

$|\delta_i| \leqq \mathbf{u}$, $\rho_i = \pm 1$ $(i = 1, 2, \ldots, n)$, $n\mathbf{u} < 1$ とする
このとき，次の不等式が成り立つ

$$\prod_{i=1}^{n}(1+\delta_i)^{\rho_i} = 1 + \theta_n$$

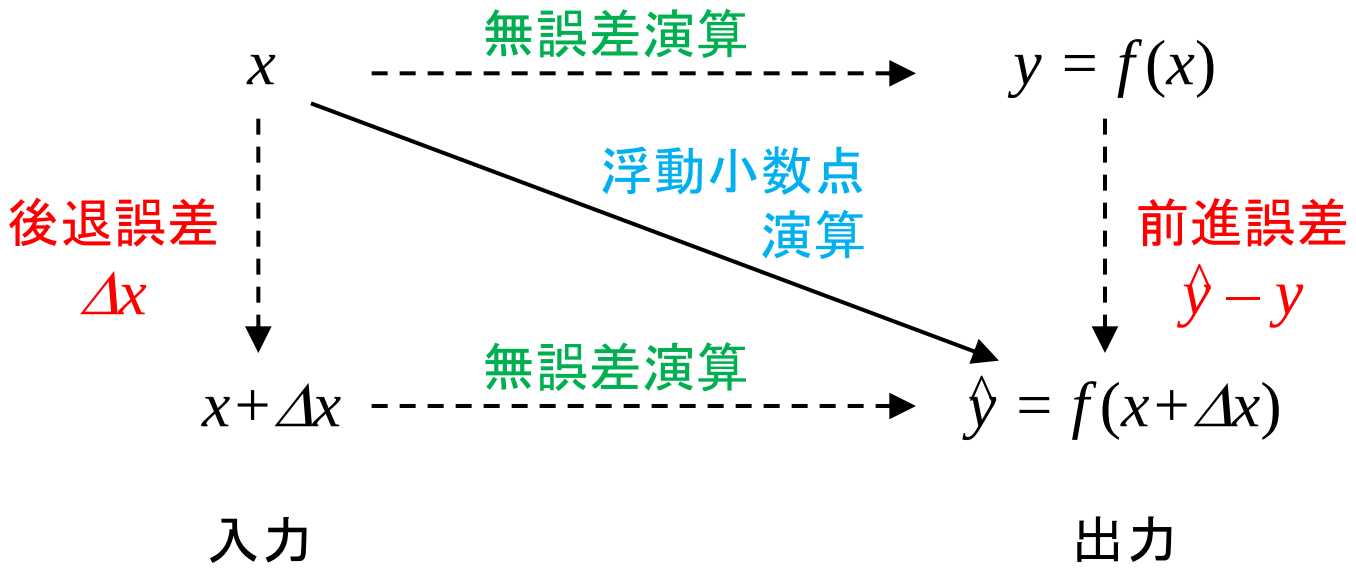
ただし，

$$|\theta_n| \leq \frac{n\mathbf{u}}{1-n\mathbf{u}} =: \gamma_n$$

[1] N. J. Higham: “Accuracy and Stability of Numerical Algorithms”, SIAM, 1996.

後退誤差解析

- 定義
  - 計算過程における誤差を，入力の誤差のせいにするタイプの誤差解析を，後退誤差解析と呼ぶ
  - すなわち，「計算は誤差なく行われたが，入力に誤差があったために，結果が真の解と異なった」と考え，その誤差の上界を求める



総和演算の誤差解析

- $y = x_1 + x_2 + \cdots + x_n$ の計算

$$
\begin{aligned}
\hat{y} &= fl(x_1 + x_2 + x_3 + \cdots + x_n) \\
&= fl(fl(\cdots fl(fl(x_1 + x_2) + x_3) + \cdots) + x_n) \\
&= fl(fl(\cdots fl((x_1 + x_2)(1 + \delta_1) + x_3) + \cdots) + x_n) \\
&= fl(fl(\cdots ((x_1 + x_2)(1 + \delta_1) + x_3)(1 + \delta_2) + \cdots) + x_n) \\
&= \cdots \\
&= (\cdots(\cdots((x_1 + x_2)(1 + \delta_1) + x_3)(1 + \delta_2) + \cdots) + x_n)(1 + \delta_{n-1}) \\
&= \sum_{i=1}^{n} x_i \prod_{j=i-1}^{n-1} (1 + \delta_j) \qquad (\delta_0 \equiv 0)
\end{aligned}
$$

ここで, $\displaystyle\prod_{j=i-1}^{n-1} (1 + \delta_j) = 1 + \theta_{n-i}, \quad |\theta_{n-i}| \le \gamma_{n-i}.$

総和演算の誤差解析(続き)

- 前進誤差
  - 前ページの結果より，前進誤差は次のようになる

$$|\hat{y}-y| = \left|\sum_{i=1}^{n} x_i\theta_{n-i}\right| \le \sum_{i=1}^{n} |x_i||\theta_{n-i}| \le \gamma_n \sum_{i=1}^{n} |x_i|.$$

- 後退誤差
  - $\hat{y} = \sum_{i=1}^{n} x_i(1+\theta_{n-i})$ だから，入力 x_i を

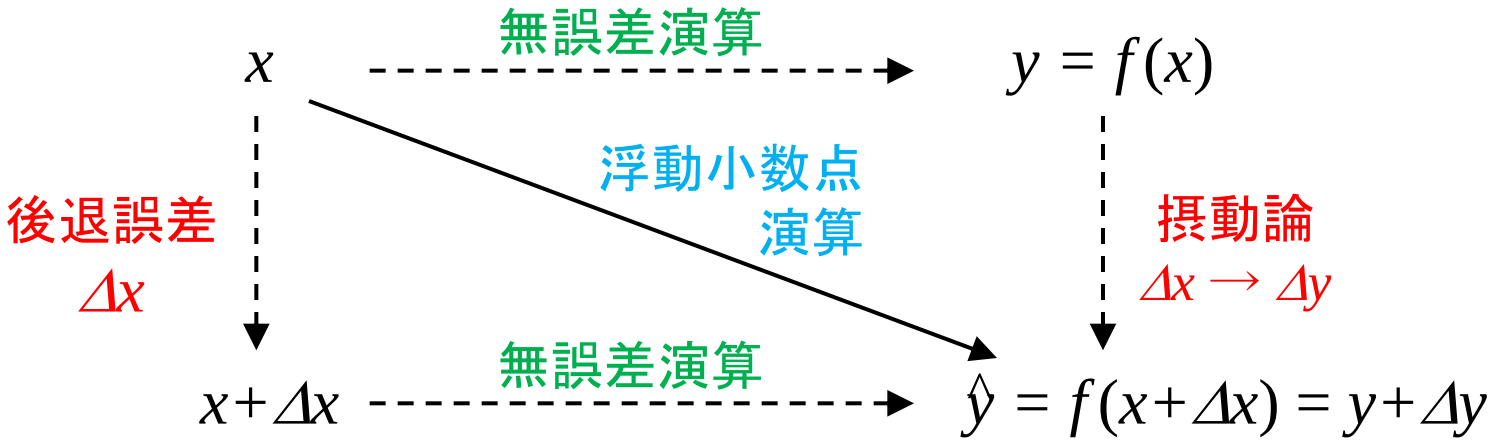
$$x_i \to (1+\theta_{n-i})x_i, \quad |\theta_{n-i}| \le \gamma_{n-i}$$

と変化させたときの正確な出力が y となる

  - よって後退誤差は $\Delta x_i = \theta_{n-i}\, x_i$.
  - 後退誤差が **u** * (n の多項式) 程度の場合，後退安定であるという

摂動論と条件数

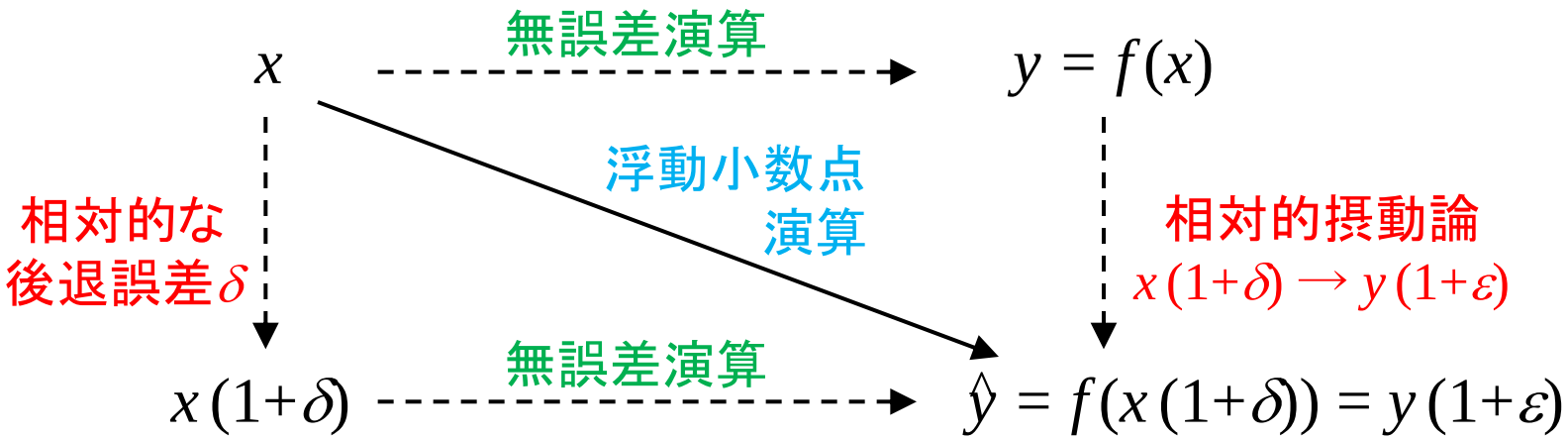
- 摂動論
  - 入力が $x \rightarrow x+\Delta x$ と変化したときの出力 y の変化 Δy
  - 後退誤差解析と摂動論を組み合わせることで，計算結果 y に含まれる誤差が評価できる



- 条件数
  - $\lim_{\varepsilon \rightarrow 0} \sup_{|\Delta x| \leqq \varepsilon} |\Delta y| / \varepsilon$ を条件数と呼ぶ
  - 条件数の大きい問題は誤差が入りやすい

相対的摂動論と相対的条件数

- 相対的摂動論
  - 入力が $x \rightarrow x(1+\delta)$ と変化したときの出力の変化 $y \rightarrow y(1+\varepsilon)$
  - 相対誤差の意味での後退誤差解析と相対的摂動論を組み合わせることで，計算結果 y に含まれる相対誤差が評価できる



- 相対的条件数
  - $\lim_{\delta \rightarrow 0} \sup_{|\delta_i| \leqq \delta} |y - \Delta y| / \delta |y|$ を相対的条件数と呼ぶ
  - 相対的条件数の大きい問題は誤差が入りやすい

総和演算の相対的条件数

- 相対的条件数
  - $x_i \to x_i(1+\delta_i), |\delta_i| \leqq \delta$ とすると

$$\frac{|y-\Delta y|}{\delta|y|} = \frac{|\sum_{i=1}^n \delta_i x_i|}{\delta|\sum_{i=1}^n x_i|} \leq \frac{\delta\sum_{i=1}^n |x_i|}{\delta|\sum_{i=1}^n x_i|} = \frac{\sum_{i=1}^n |x_i|}{|\sum_{i=1}^n x_i|}.$$

  - したがって，$\sum_{i=1}^n x_i \neq 0$ ならば，相対的条件数は $\frac{\sum_{i=1}^n |x_i|}{|\sum_{i=1}^n x_i|}$.
  - すなわち，和が 0 に近いとき，悪条件になる

- 後退誤差解析との組み合わせ
  - 後退誤差解析の結果より，$|\delta_i| \leqq \gamma_{n-i} \leqq \gamma_n$.
  - したがって，出力 y の相対誤差限界は，

$$\gamma_n \frac{\sum_{i=1}^n |x_i|}{|\sum_{i=1}^n x_i|}.$$

（前進誤差解析と同じ結果）

基本的な線形計算の後退誤差解析(1)

- 内積
  - 総和演算と同じ手法で後退誤差解析を行うことにより，$x, y \in \mathbf{R}^n$ の内積について，次の結果が成り立つ

$$fl(\mathbf{x}^\top \mathbf{y}) = (\mathbf{x} + \Delta\mathbf{x})^\top \mathbf{y} = \mathbf{x}^\top (\mathbf{y} + \Delta\mathbf{y}),$$
$$|\Delta\mathbf{x}| \le \gamma_n |\mathbf{x}|, \quad |\Delta\mathbf{y}| \le \gamma_n |\mathbf{y}|. \qquad \text{（後退安定）}$$

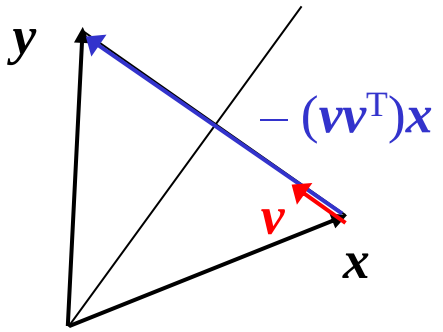
  - ただし，
    - $|x|$ は x の各要素の絶対値を要素とするベクトル
    - $\leqq$ は全ての要素について不等号が成り立つことを意味する
  - これより，前進誤差も次のように得られる

$$|\mathbf{x}^\top \mathbf{y} - fl(\mathbf{x}^\top \mathbf{y})| \le \gamma_n \sum_{i=1}^{n} |x_i y_i| = \gamma_n |\mathbf{x}|^\top |\mathbf{y}|.$$

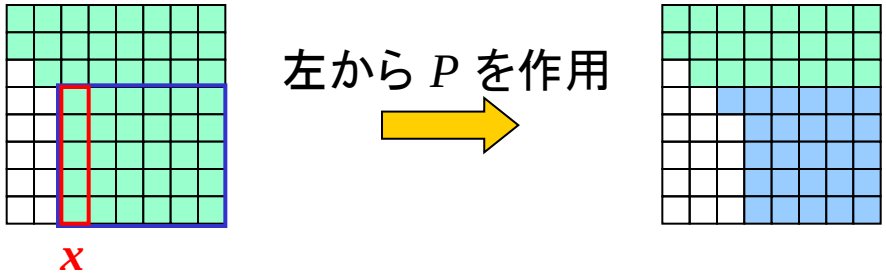
  - これらの結果は，行列ベクトル積，行列乗算にも適用可能

基本的な線形計算の後退誤差解析(2)

- ハウスホルダー変換
  - 長さ $\sqrt{2}$ のベクトル $\boldsymbol{v}$ を用いて $P = I - \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}$ と定義される直交変換
  - $\boldsymbol{v}$ に垂直な平面に関する折り返しを表す
  - $\| \boldsymbol{x} \| = \| \boldsymbol{y} \|$ ならば，$\boldsymbol{x}$ を $\boldsymbol{y}$ に移すハウスホルダー変換が存在する



- ハウスホルダー変換の利用法
  - $\boldsymbol{y} = [\, \| \boldsymbol{x} \| \,, 0, \ldots, 0]^{\mathrm{T}}$ とすると，ハウスホルダー変換を用いて $\boldsymbol{x}$ の2番目以降の要素を消去できる
  - QR分解や，固有値問題における3重対角化で使われる



基本的な線形計算の後退誤差解析(3)

- ハウスホルダー変換の後退誤差解析
  - ハウスホルダー変換のベクトルを $\boldsymbol{v}$ とし，浮動小数点演算によって（ある決まったやり方で）計算されたベクトルを $\hat{\boldsymbol{v}}$ とする
  - $\hat{\boldsymbol{v}}$ を用いて，ベクトル $\boldsymbol{b}$ に対して次のようにハウスホルダー変換を行うとする

$$\mathbf{y} = \hat{P}\mathbf{b} = (I - \hat{\mathbf{v}}\hat{\mathbf{v}}^{\top})\mathbf{b} = \mathbf{b} - \hat{\mathbf{v}}(\hat{\mathbf{v}}^{\top}\mathbf{b}).$$

  - このとき，浮動小数点演算によって計算されたベクトル $\boldsymbol{y}$ は，次の式を満たす

$$\hat{\mathbf{y}} = (P + \Delta P)\mathbf{b}, \quad \|\Delta P\|_F \leq \gamma_{cn}.$$ （後退安定）

  - これは，次のようにも書き直せる

$$\hat{\mathbf{y}} = P(\mathbf{b} + \Delta\mathbf{b}), \quad \|\Delta\mathbf{b}\|_2 \leq \gamma_{cn}\|\mathbf{b}\|_2.$$

基本的な線形計算の後退誤差解析(4)

- 複数のハウスホルダー変換に対する後退誤差解析
  - r 個のハウスホルダー変換 $P_k = I - \boldsymbol{v}_k \boldsymbol{v}_k^{\mathrm{T}}$ $(k=1, \ldots, r)$ があるとする
  - いま，行列 $A_0 = A \in \mathrm{R}^{m \times n}$ に対して，次のように r 個のハウスホルダー変換を作用させたとする

$$A_k = P_k A_{k-1}$$

  - ただし，$\boldsymbol{v}_k$ は浮動小数点で計算し，A_{k-1} への作用も浮動小数点で計算するものとする
  - このとき，計算結果 $\hat{A}_r$ は次の式を満たす

$$\hat{A}_r = Q^{\top}(A + \Delta A).$$

  - ただし，

$$Q^{\top} = P_r P_{r-1} \cdots P_1,$$

$$\|\Delta A\|_F \leq r\gamma_{cm}\|A\|_F. \quad \text{（後退安定）}$$

ハウスホルダー変換の応用

- QR分解（直交変換による上三角化）
  - 行列 $A \in \mathrm{R}^{m \times n}$ $(m \geqq n)$を，浮動小数点演算を用いたハウスホルダー変換により上三角化し，行列 $\hat{R} \in \mathrm{R}^{m \times n}$ が得られたとする
  - このとき，ある（厳密な）直交行列 Q が存在し，$\hat{R}$ は次の式を満たす

$$A + \Delta A = Q\hat{R}, \quad \|\Delta A\|_F \leq n\gamma_{cm}\|A\|_F. \quad \text{（後退安定）}$$

- 3重対角化・ヘッセンベルグ化
  - 行列 $A \in \mathrm{R}^{n \times n}$ を，浮動小数点演算を用いた両側ハウスホルダー変換によりヘッセンベルグ化（A が対称の場合は3重対角化）し，行列 $\hat{H} \in \mathrm{R}^{n \times n}$ が得られたとする
  - このとき，ある（厳密な）直交行列 Q が存在し，$\hat{H}$ は次の式を満たす

$$Q^{\top}(A + \Delta A)Q = \hat{H}, \quad \|\Delta A\|_F \leq 2n\gamma_{cn}\|A\|_F. \quad \text{（後退安定）}$$

固有値計算の精度

- 非対称行列の場合
  - $A \in \mathrm{R}^{n \times n}$ が $X^{-1}AX = \Lambda$ と対角化可能とする。いま，$A \rightarrow A + \Delta A$ という摂動が加わったとすると，A の任意の固有値 λ の摂動は，次の式で抑えられる（Bauer-Fike Theorem）

$$|\Delta\lambda| \leq \kappa(X)\|\Delta A\|_2.$$

  - ヘッセンベルグ化の後退誤差は小さいが，固有ベクトル行列 X の条件数 $\kappa(X)$ が大きいと，固有値の誤差は大きくなる可能性がある

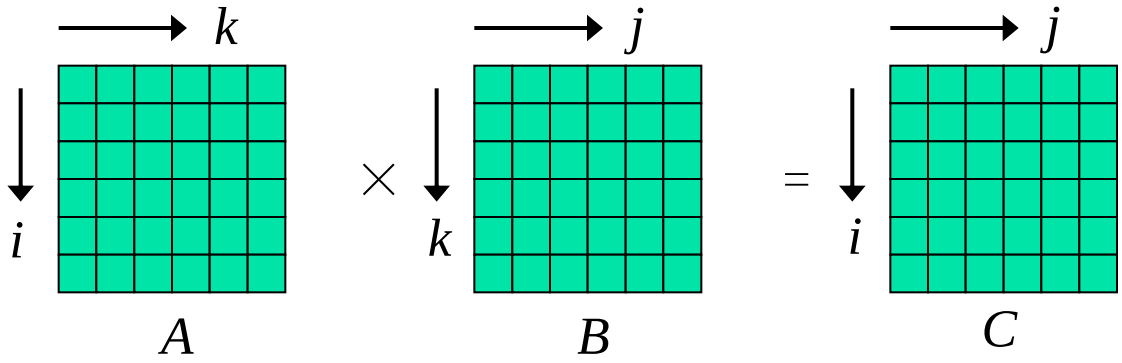
- 対称行列の場合
  - X は直交行列にとれるから，$\kappa(X) = 1$。よって上式は次のようになる

$$|\Delta\lambda| \leq \|\Delta A\|_2.$$

  - すなわち，固有値の誤差は後退誤差と同程度で非常に小さい

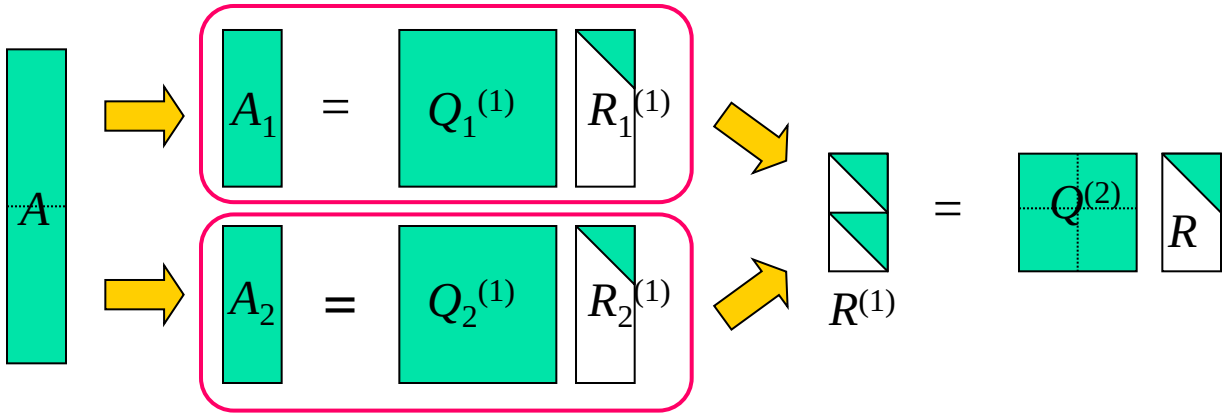
並列化と誤差

- 行列乗算の並列化
  - 行列乗算は i, j, k の3重ループ
  - ループ i, j については，並列化しても結果は（丸め誤差を含めて）変わらない
  - ループを k 並列化すると，演算順序が変わるため，丸め誤差のレベルで結果が変わる
  - ループを k 並列化すると，内積長が短くなるため，前進誤差／後退誤差の上界はむしろ小さくなる



TSQRアルゴリズムの誤差解析(1)

- TSQRアルゴリズム（Langou, 2007）
  - ハウスホルダー変換によるQR分解を大粒度並列で行う手法
  - 行列を上下に2分割し，それぞれをQR分解した後，結果を合成
  - 再帰的に適用することで，2^L 個のプロセッサで並列化可能
  - ハウスホルダー変換における内積の長さは短くなる



大粒度の並列性
（2PUで独立に実行）

第1ステップに比べ演算量少
（1PUで実行）

TSQRアルゴリズムの誤差解析(2)

- 後退誤差解析の結果
  - 分割数が増えるほど，最初のQR分解の誤差は減少
  - ただし，合成部分からの誤差が新たに生じる

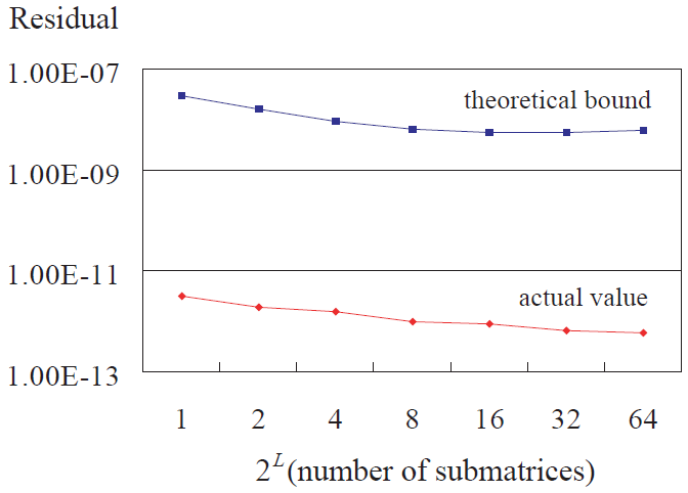
$$A + \Delta A = Q\tilde{R},$$

$$\|\Delta A\|_F \simeq n\gamma_{c\cdot\frac{m}{2^L}} + Ln\gamma_{c\cdot 2n}.$$

（後退安定）

最初のQR分解　合成部分

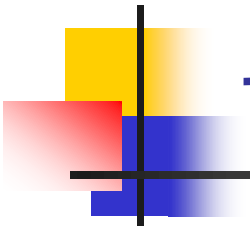
- 数値実験
  - $m = 6400$, $n = 100$ のランダム行列
  - 分割数（並列数）は 1 ～ 64
  - この範囲では，分割数が増えるにつれ，誤差はむしろ減少



参考：MPI_Reduce における演算順序

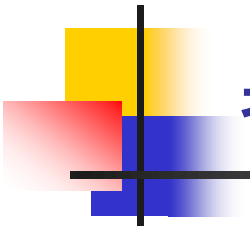
- リダクション操作に関する指定
  - リダクションの標準的評価順序はグループ内のプロセスのランクによって決まる。しかし、実装では結合則、または結合則と可換則を利用して、評価順を変更することもできる。浮動小数点数加算など厳密には結合的、可換的でない操作についてはこの操作によってリダクションの結果が変わることもある。

- 実装者へのアドバイス
  - MPI_REDUCEを実装する場合には、同じ順序で現れる同じ引数で関数を適用する際には、必ず同じ結果が得られるようにすることが強く求められる。このためプロセッサの物理的配置を利用する最適化はできない場合があることに注意されたい。



まとめ

- 丸め誤差の性質，後退誤差解析，摂動論と条件数など，誤差解析の基本について紹介した
- 内積，ハウスホルダー変換などの基本的な線形計算について，後退誤差解析の結果を紹介し，応用として，固有値計算の誤差について述べた
- 並列化と誤差の振る舞いについて，大粒度並列性を持つQR分解法であるTSQRアルゴリズムを例にとって紹介した



参考文献

- N. J. Higham: “Accuracy and Stability of Numerical Algorithms”, 2nd Ed. SIAM, 2002.
- G. Golub and C. F. van Loan: “Matrix Computations”, 4th Ed., The Johns Hopkins Univ. Press, 2012.
- MPI:メッセージ通信インターフェース標準(日本語訳ドラフト) http://phase.hpcc.jp/phase/mpi-j/ml/mpi-j-html/chap0.html