多倍長計算手法

平成26年度第4四半期

目次

1.はじめに

ファインマンループ積分のmassless計算に関してその発端から数学的根拠などをまとめました。
これに関連して,端点と内部に同時に特異点を持つ積分に対してHadamardの有限部分をsinc求積法と今までのε−算法で実行した時の結果の精度の比較をしています。

また、非対称行列反復計算の調査で作成したメモをまとめて記述しています。
(1) 整数演算での減算の桁落ち計算
(2) 仮数部のビット毎の乗算
またDD,D (longdouble) ,DQソースのインライン化,並列化のためのメモをまとめています。

2.massless計算

2.1 vtx-1

inf ra vtx **の計算において**

$$\int_0^1 \int_0^{1-x} \frac{1}{D} dydx \qquad D = -sxy + (x+y)^2 m^2 + (1-x-y)\lambda^2$$

を$s = 500^2, m = 0.0005$**で計算する場合**, $D-> D - i\varepsilon$**として計算しますが**λ**の値が小さくなると**, **反復回数が多くなり計 算に時間がかかったり**, **計算できない** ($\lambda = 10^{-200}$ **など倍精度**, DD**形式の**4**倍精度など**) **事が発生するので**, **以下の式で計算する事 があります。この場合 は**

$$\int_0^1 \int_0^{1-x} \frac{1}{D^{1-\eta}} dydx \qquad D = -sxy + (x+y)^2 m^2 \quad D-> D - i\varepsilon, \eta = \frac{1}{\log(\frac{1}{\lambda^2})}$$

の値が充分小さくないと結果の精度が良くない事があります。

実 測値と解析式の相対誤 差

λ	**実数 部**	**虚数部**
10^{-15}	6.23×10^{-3}	1.38×10^{-2}
10^{-150}	6.30×10^{-5}	1.49×10^{-4}
10^{-1500}	6.68×10^{-7}	1.59×10^{-6}
10^{-2500}	2.26×10^{-7}	5.58×10^{-7}
10^{-15000}	6.42×10^{-9}	1.61×10^{-8}

数学的には以下のようになります。

$$I=\int_0^1\int_0^{1-x}\frac{1}{-sxy+(x+y)^2m^2+(1-x-y)\lambda^2}dydx \qquad s<0$$

$$=\int_0^1\int_0^1\frac{p}{-sp^2q(1-q)+p^2m^2+(1-p)\lambda^2}\quad dpdq=A+B$$

$$A=\frac{1}{2}\int_0^1\frac{\log\dfrac{-sq(1-q)+m^2}{\lambda^2}}{-sq(1-q)+m^2}$$

$$B=\int_0^1\frac{\lambda^2}{2a}\frac{2}{\sqrt{4ac-b^2}}\arctan\frac{\sqrt{4ac-b^2}}{b+2c}$$

$$a=-sq(1-q)+m^2\ge m^2,b=-\lambda^2,c=\lambda^2,4ac-b^2\ge 4m^2\lambda^2$$

$$|B|\le\frac{\lambda}{2m^3}\frac{\pi}{2}=\frac{\lambda\pi}{4m^3}$$

$$J=\int_0^1\int_0^{1-x}\frac{1}{(-sxy+(x+y)^2m^2)^{1-\eta}}\qquad \eta=\frac{1}{\log(\frac{1}{\lambda^2})}$$

$$=\int_0^1\int_0^1\frac{p}{p^{2-2\eta}}\frac{1}{(-sq(1-q)+m^2)^{1-\eta}}dpdq=\frac{1}{2\eta}\int_0^1\frac{(-sq(1-q)+m^2)^{\eta}}{-sq(1-q)+m^2}dq$$

$$=\frac{1}{2}\int_0^1[\log(\frac{1}{\lambda^2})\frac{1}{-sq(1-q)+m^2}+\frac{\log(-sq(1-q)+m^2)}{-sq(1-q)+m^2}]dq$$

$$=\frac{1}{2}\int_0^1\frac{\log\dfrac{-sq(1-q)+m^2}{\lambda^2}}{-sq(1-q)+m^2}=A$$

$I \fallingdotseq J$　　**誤差** $=\dfrac{\lambda\pi}{4m^3}$

Aの計算から$s<0$の場合の解析近似解は以下の様にして導く事ができます。

$$A=\frac{1}{2}\int_0^1 \frac{\log\frac{-sq(1-q)+m^2}{\lambda^2}}{-sq(1-q)+m^2}dq$$

$$=\frac{1}{2}\log(\frac{m^2}{\lambda^2})\int_0^1 \frac{1}{-sq(1-q)+m^2}dq$$

$$+\frac{1}{2}\int_0^1 \frac{\log\frac{-sq(1-q)+m^2}{m^2}}{-sq(1-q)+m^2}dq$$

$$=\frac{1}{-s}\log(\frac{m^2}{\lambda^2})\log(\frac{-s}{m^2})+\frac{1}{2}\frac{1}{-s}[\log^2(\frac{m^2}{-s})-\frac{\pi^2}{3}]$$

$$=\frac{1}{-s}[\log(\frac{m^2}{\lambda^2})\log(\frac{-s}{m^2})+\frac{1}{2}\log^2(\frac{m^2}{-s})-\frac{\pi^2}{6}]$$

S<0では

inf ra vtx

$$s = -500^2, m = 0.0005$$

$$\lambda = 10^{-n}, \eta = \frac{1}{\log(\frac{1}{\lambda^2})}$$

n	実測値	解析近似解
20000	0.10179484611990185 2D+02	0.1017948457346719 59D+02
30000	0.15269306785192238 6D+02	0.1526930675952645 98D+02
40000	0.20359128964817937 7D+02	0.2035912894558572 01D+02
50000	0.25448951147013122 7D+02	0.2544895113164498 21D+02
60000	0.30538773330493054 2D+02	0.3053877331770424 42D+02
70000	0.35628595514707129 2D+02	0.3562859550376350 63D+02
80000	0.40718417699380047 7D+02	0.4071841768982277 19D+02
90000	0.45808239884358862 1D+02	0.4580823987588203 04D+02
100000	0.50898062069551805 6D+02	0.5089806206194129 60D+02
110000	0.55987884254900471 6D+02	0.5598788424800055 45D+02
120000	0.61077706440365936 7D+02	0.6107770643405982 01D+02
130000	0.66167528625921249 8D+02	0.6616752862011908 58D+02
140000	0.71257350811547155 4D+02	0.7125735080617833 72D+02
150000	0.76347172997229520 8D+02	0.7634717299223760 28D+02
160000	0.81436995182957787 2D+02	0.8143699517829686 84D+02
170000	0.86526817368723826 0D+02	0.8652681736435613 40D+02
180000	0.91616639554521356 1D+02	0.9161663955041538 54D+02
190000	0.96706461740345417 9D+02	0.9670646173647465 10D+02
200000	0.10179628392619200 4D+03	0.1017962839225339 17D+03
210000	0.10688610611205791 7D+03	0.1068861061085931 82D+03
220000	0.11197592829794051 3D+03	0.1119759282946524 34D+03
230000	0.11706575048383761 8D+03	0.1170657504807116 99D+03
240000	0.12215557266974742 8D+03	0.1221555726667709 65D+03
250000	0.12724539485566839 4D+03	0.1272453948528302 31D+03
260000	0.13233521704159926 5D+03	0.1323352170388894 82D+03
270000	0.13742503922753888 9D+03	0.1374250392249487 62D+03
280000	0.14251486141348638 6D+03	0.1425148614110080 13D+03
290000	0.14760468359944090 4D+03	0.1476046835970672 65D+03
300000	0.15269450578540175 9D+03	0.1526945057831265 44D+03

と非常に良い結果が得られています。

2.2 vtx-2

さらに, $\int_0^1 \int_0^{1-x} \frac{1}{D^{1-\eta}} dydx \quad D = xy \quad \eta = \frac{1}{\log(\frac{1}{\lambda^2})}$ の場合を

計算すると,

$$I = \int_0^1 \int_0^{1-x} \frac{1}{D^{1-\eta}} dydx = \int_0^1 \frac{(1-x)^{\eta}}{\eta x^{1-\eta}} dx = \frac{1}{\eta} B(\eta, \eta+1) = \frac{\Gamma(\eta)\Gamma(1+\eta)}{\eta\Gamma(1+2\eta)}$$

$$= \frac{1}{\eta^2} \frac{(\Gamma(1+\eta))^2}{\Gamma(1+2\eta)}$$ となり, 二重指数関数型積分で容易に計算出来ています。結果は積分変数の変換区間(0,1)での最小値に依存し, 10^{-150} と 10^{-2465} (指数部のビット数が11か15かに依存)では $\eta = 0.01$ ($\lambda = 10^{-21.7147241}$ に相等)でその精度は10進1桁と10進24桁(反復法を使用せず)の差があり, 前者では40回の反復により10進8桁となっています。このケースでは $D- > D - i\varepsilon$ とする操作は必要ありませんでした。

2.3 box

少し複雑なケース

$$\eta = \frac{1}{\log(\frac{1}{\lambda^2})}$$

$$I = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} \frac{1}{D^{2-\eta}} dz dy dx$$

$$D = xz + y(1-x-y-z)$$

では,

$$I = \frac{1}{1-\eta}\frac{2}{\eta}(\frac{2}{\eta} - \frac{\pi^2}{4}2^{1-\eta}\eta)\frac{(\Gamma(1+\eta))^2}{\Gamma(1+2\eta)}$$

$$\frac{(\Gamma(1+\eta))^2}{\Gamma(1+2\eta)} = \frac{1+2\eta}{(1+\eta)^2}\exp[\sum_{k=2}^{\infty}(-1)^k(2-2^k)(\frac{\varsigma(k)-1}{k})\eta^k]$$

$$\log\Gamma(1+x) = -\log(1+x) + (1-\gamma)x + \sum_{k=2}^{\infty}(-1)^k(\frac{\varsigma(k)-1}{k})x^k$$ **より。**

$$I = \frac{a_{-2}}{\eta^2} + \frac{a_{-1}}{\eta} + a_0$$ **とすると,** $$a_{-2} = 4, a_{-1} = 4, a_0 = -4(\frac{5}{12}\pi^2 - 1) = -\frac{5}{3}\pi^2 + 4$$

これは以下の事からきています。

$$I = [\int_0^1 \frac{(1-q)^{1-\eta} - q^{1-\eta}}{(1-2q)q^{1-\eta}(1-q)^{1-\eta}} dq] \times \frac{1}{1-\eta}[\int_0^1 \frac{1}{p^{1-\eta}(1-p)^{1-\eta}} dp]$$

から $$\frac{1}{1-\eta}[\int_0^1 \frac{1}{p^{1-\eta}(1-p)^{1-\eta}} dp] = \frac{1}{1-\eta}\frac{2}{\eta}\frac{(\Gamma(1+\eta))^2}{\Gamma(1+2\eta)}$$ **。**

$$\int_0^1 \frac{(1-q)^{1-\eta} - q^{1-\eta}}{(1-2q)q^{1-\eta}(1-q)^{1-\eta}} dq = \frac{2}{\eta} - \frac{\pi^2}{4}2^{1-\eta}\eta$$ **となるがより詳細には**

次ぎの様になります。

$$\int_0^1 \frac{(1-q)^{1-\eta}-q^{1-\eta}}{(1-2q)q^{1-\eta}(1-q)^{1-\eta}}dq = \int_0^1 [\frac{q^{\eta}}{(1-2q)q} - \frac{(1-q)^{\eta}}{(1-2q)(1-q)}]dq$$

$$= 2\int_0^1 \frac{q^{\eta}-(1-q)^{\eta}}{1-2q}dq + \frac{2}{\eta} = \frac{2}{\eta} + 2^{1-\eta}\int_0^1 \frac{(1-t)^{\eta}-(1+t)^{\eta}}{t}dt$$

$$= \frac{2}{\eta} - 2^{1-\eta}2\eta(1+\sum_{k=1}^{\infty}\frac{(\eta-1)(\eta-2).....(\eta-2k)}{(2k+1)!(2k+1)})$$

(　)の定数項は $-2(1+\sum_{k=1}^{\infty}\frac{1}{(2k+1)^2}) = -2(\sum_{k=1}^{\infty}\frac{1}{(2k-1)^2})$

$$= -2\frac{3}{4}\varsigma(2) = -\frac{\pi^2}{4}$$

(　)のηの一次の項は,負の数の和の絶対値で計算すると,

$$\sum_{k=1}^{\infty}\frac{\text{分子の項}}{(2k+1)!(2k+1)}.$$

分子の項 $= (2k)!\sum_{r=1}^{2k}\frac{1}{r} < (2k)![\log(2k+1)+\gamma]$

$\gamma = 0.5772156649$　(**オイラーの定数**)

$$\sum_{k=1}^{\infty}\frac{(2k)![\log(2k+1)+\gamma]}{(2k+1)!(2k+1)} = \sum_{k=1}^{\infty}\frac{\log(2k+1)}{(2k+1)^2} + \gamma\sum_{k=1}^{\infty}\frac{1}{(2k+1)^2}$$

$$\sum_{k=1}^{\infty}\frac{\log(2k+1)}{(2k+1)^2} = \pi^2[\frac{3}{2}\log A - \frac{1}{6}\log(2) - \frac{1}{8}\log(\pi) - \frac{\gamma}{8}]$$

$$= 0.4181157384$$

$$A = \lim_{n->\infty}\frac{1\times 2^2\times 3^3\times .....\times n^n}{n^{n^2/2+n/2+1/12}}e^{\frac{n^2}{4}} = 1.282427130$$

$$\log A = 0.2487544703$$

$$\gamma\sum_{k=1}^{\infty}\frac{1}{(2k+1)^2} = \gamma[\sum_{k=1}^{\infty}\frac{1}{(2k-1)^2} - 1] = \gamma(\frac{\pi^2}{8} - 1) = 0.1348956184$$

これから(　)内のηの一次の項の係数は -0.5530113568

全体では $-2^{1-\eta}\times 2\eta$**をかけて** $2.212\eta^2$**となりこれが一次式まで取った場合の誤差となります。**

実測例としては以下のものがあります。

解析近似式を使用して，$\lambda = 10^{-1000000001} \sim 10^{-1000000045}$

で計算しました。**ただしgamma関数はサポートされていない機種もあり,展開式より計算しています。**

a2= 4.000000000000000650941944549950 95
a1= 4.000000000000011177566605145637 09
a0= -12.449340667611871206644530645411 1

変数変換区間SR16000$[10^{-100},1-10^{-100}]$, x5570**と**x2670**では**
$[10^{-200},1-2^{-113}]$**の4倍精度では**
epsilon**算法の収束具合は**
$\lambda = 10^{-150}$**で**

	値	**誤差**
x5570	0.191143398445070834D + 07	0.14063331231312 7879D - 05
x2670	0.191143398444577885D + 07	0.21027092538834 4072D - 07
sr16000	0.191143398444582948Q + 07	0.15665822292190 9210Q - 05

$\lambda = 10^{-1500}$**で**

x5570	0.190895950695064375D + 09	0.91950002332009 1177D - 01
x2670	0.190895950598091269D + 09	0.47781492985379 9862D - 01
sr16000	0.190895950686589134Q + 09	0.47081222571149 9829Q - 01

$\lambda = 10^{-15000}$**で**

x5570	0.190871021209837876D + 11	0.771928148538333458D + 04
x2670	0.190871096344010433D + 11	0.183847060803184605D + 03
sr16000	0.190871111864473018Q + 11	0.190871121149625 21Q + 11

(**注**)SR16000**では他に変化の小さい要素がある。**

となっています。

あとSR16000でのあるλの範囲でのa_{-2}, a_{-1}, a_0は以下の様になっています。

ramda=10^{-16}~10~[-30}
a2= 4.000003447246776458440340467574 16
a1= 3.998614076193875546687457655832 65
a0= -12.263945340240600000000000000200 20

ramda=10^{-31}~10~[-45}
a2= 4.000000995778690564344849305288 36
a1= 3.999395092335383959734798438551 65
a0= -12.326990512798999999999999975916 2

ramda=10^{-46}~10~[-60}

a2= 4.000000412642012798006451400647 05
a1= 3.999663959451037435790618527291 21
a0= -12.358196281822999999999999967960 46

ramda=10^{-61}~10~[-75}

a2= 4.000000196766259218742464194527 51
a1= 3.999794900923913886687933816573 74
a0= -12.378147480874000000000000184813 6

ramda=10^{-76}~10~[-90}

a2= 4.000000144598692200929075901625 62
a1= 3.999832409009722650036422490660 11
a0= -12.384869002304000000000000160380 0

rambda=10^{-91}~10^{-105}

a2= 4.000000036947829850173391166644 84
a1= 3.999964948924225079755979080362 44
a0= -12.415458713321752600000000000000 01

λの値が小さくなるほど

a_0は$-\frac{5\pi^2}{3}+4$に近付いています。

3sin c**求積法による**Hadamard**有限部分計算**

端点と領域内部で同 時に特異点をもつ

場合の計算を，数理解析研究所講究録

第791**巻**1992**年**206 - 219

Hadamard**有限部分積分に関する**DE**公式**

緒方秀教，杉原正顯，森正武

(東京大学工学部物理工 学科)

の資料から以下の問題を測定しました。

$$\text{f.p.}\int_{-1}^{1}\frac{F(x)}{(x-\lambda)^2}dx, F(z)=(1-z)^{\frac{1}{4}}(1+z)^{\frac{-1}{4}}$$

厳密値 $=-\frac{\pi}{2}(1-\lambda)^{\frac{-3}{4}}(1+\lambda)^{\frac{-5}{4}}$

公式 $I=h\sum_{k=-N_1}^{N_2}\frac{F(x_k)/\phi'(x_k)}{(x_k-\lambda)^2}$

$$+\{\cot[\frac{\pi}{h}\phi(\lambda)]F'(\lambda)-\frac{\pi}{h}\frac{\phi'(\lambda)}{\sin^2[\pi\phi(\lambda)/h]}F(\lambda)\}$$

3.1 SE,DEの計算方法

$$DE: x = \tanh\left(\frac{\pi}{2}\sinh(t)\right) \qquad \phi(x) = \sinh^{-1}\left(\frac{1}{\pi}\log\left(\frac{1+x}{1-x}\right)\right)$$

$$s = \frac{1}{\pi}\log\left(\frac{1+x}{1-x}\right), \phi(x) = \log\left(s + \sqrt{s^2+1}\right)$$

$$\phi'(x) = \frac{\frac{1}{\pi}\frac{2}{(1-x^2)}}{\sqrt{\left(\frac{1}{\pi}\log\left(\frac{1+x}{1-x}\right)\right)^2+1}}$$

$$SE: x = \tanh\left(\frac{t}{2}\right) \qquad \phi(x) = \log\left(\frac{1+x}{1-x}\right) \qquad \phi'(x) = \frac{1}{1-x^2}$$

$$F'(z) = -\frac{1}{2(1-z)^{\frac{3}{4}}(1+z)^{\frac{5}{4}}}$$

$x = 1-\varepsilon, -1+\varepsilon$**となる**$t$

$$DE \qquad t = \log\left(\frac{2y}{\pi} + \sqrt{\left(\frac{2y}{\pi}\right)^2+1}\right) \qquad 2y = \log\left(\frac{1}{\varepsilon}\right)$$

$$SE \qquad t = \ln\left(\frac{2}{\varepsilon}\right)$$

$\varepsilon = 2^{-106}$**の場合**, $DE: t = 3.855, SE: t = 74.1667$

$h = 0.125$**で測定**

ここでsinhの逆関数はサポートされていませんので，logを使用した式を使っています。

3.2　ε－算法

$$f.p\int_{-1}^{1}\frac{F(x)}{(x-\lambda)^2}dx=\lim_{\varepsilon->0}\int_{-1}^{1}\frac{F(x)}{(x-\lambda+i\varepsilon)^2}dx$$

で計算しています。

$h=0.5^8,\varepsilon_0=1.0,\varepsilon_n=(\frac{1}{1.1})^{n-1}$**で**$n=50$**まで計算**

今回は領域を分割しない場合と，領域を

$[-1,\lambda],[\lambda,1]$**と分割した場合を計算しました。**

変数変換区間$[-1+\varepsilon,1-\varepsilon],\varepsilon=2^{-106}$**の場合，**

領域分割なしでは，$N=1976$**ですが，領域分割する**

場合，$\lambda=\pm0.9$**でオーバーフローを防ぐ処理が必要**

になります。

$(x=x30(i)*cnt0+cnt1$**で発生する丸め誤差のため**$)$

$N=1976$**の場合，**$x+1.0q0>0.0,1.0q0-x>0.0$**の場合のみ**

計算するとします。また$\varepsilon=2^{-104.5}$**として**$N=1962$**とする**

方法もあります。

3.3 計算結果精度比較

4つの方法で精度を検証した結果は以下の様になっています。

変数変換区間は全て$[-1+\varepsilon,1-\varepsilon]$

$\varepsilon = 2^{-106}$としています。

相対誤算一覧表

λ	DE(分割なし)	DE(分割あり)	sinc (DE)	sinc (SE)
-0.9	0.1400D-04	0.1382Q-06	0.8469D-21	0.1825D-21
-0.8	0.6763D-06	0.5489Q-11	0.6245D-24	0.4563D-22
-0.7	0.5165D-06	0.3596Q-14	0.1757D-24	0.2028D-22
-0.6	0.2143D-06	0.2125Q-15	0.1855D-25	0.1143D-22
-0.5	0.6756D-07	0.1301Q-15	0.9166D-26	0.7301D-23
-0.4	0.3159D-07	0.4492Q-17	0.6154D-26	0.5070D-23
-0.3	0.1333D-07	0.1588Q-16	0.4696D-26	0.3714D-23
-0.2	0.3104D-07	0.2240Q-16	0.5978D-26	0.2852D-23
-0.1	0.1650D-07	0.7873Q-17	0.2845D-26	0.2253D-23
0.1	0.2247D-08	0.4837Q-20	0.1905D-26	0.1508D-23
0.2	0.1024D-07	0.1085Q-17	0.1217D-25	0.1268D-23
0.3	0.1762D-06	0.1065Q-17	0.1360D-26	0.1050D-23
0.4	0.7024D-06	0.2245Q-17	0.1172D-26	0.9314D-24
0.5	0.9061D-07	0.1208Q-16	0.1010D-26	0.8110D-24
0.6	0.7005D-06	0.1356Q-15	0.1099D-26	0.7211D-24
0.7	0.1890D-06	0.5910Q-14	0.5852D-25	0.6299D-24
0.8	0.2144D-06	0.3936Q-11	0.2826D-24	0.5632D-24
0.9	0.5647D-06	0.1723Q-06	0.1583D-21	0.5057D-24

今回の結果ではsinc (DE) 分点数62で最も適しているという結果になりました。

4.多倍長減算桁落ちメモ

整数演算方式では加減算において内部処理が減算の場合の桁落ちと丸め処理が特に問題となりますので,作成時使用しましたメモを紹介します。

多倍長整数減算$IZ = IX - IY \quad (IX > IY > 0)$**の結果が**64**ビット整数変数の**60**ビットのみ使用するとします。**

4.1 3倍精度

配列数 2

$IZ(2) \geq 2^{22}$ $IZ(2) = IZ(2) - 2^{22}$ **桁落ちなし.**

丸め位置$IZ(1)(1:1)$

$2^{21} \leq IZ(2) < 2^{22}$ $IZ(2) = IZ(2) - 2^{21}$ **丸め位置**$IZ(1)(0:0)$

丸め後 $IZ(2) = 2^{22}$ **桁落ちなし.**

$IZ(2) \neq 2^{22}$ **桁落ち**1

$2^{20} \leq IZ(2) < 2^{21}$ $IZ(2) = IZ(2) - 2^{20}$ **桁落ち**2 **丸めなし**

$IZ(2) < 2^{20}$ $IZ(ikk) = IZ(ikk) - 2^{ik}$

桁落ち $82 - (ikk - 1) \times 60 - ik$

丸めなし。 $(1 \leq ikk \leq 2, 0 \leq ik \leq 59)$

4.2 5倍精度

配列数 3

$IZ(3) \geq 2^{26}$ $IZ(3) = IZ(3) - 2^{26}$ **桁落ちなし.**

丸め位置$IZ(1)(1:1)$

$2^{25} \leq IZ(3) < 2^{26}$ $IZ(3) = IZ(3) - 2^{25}$ **丸め位置**$IZ(1)(0:0)$

丸め後 $IZ(3) = 2^{26}$ **桁落ちなし**

.$IZ(3) \neq 2^{26}$ **桁落ち**1

$2^{24} \leq IZ(3) < 2^{25}$ $IZ(3) = IZ(3) - 2^{24}$ **桁落ち**2 **丸めなし**

$IZ(3) < 2^{24}$ $IZ(ikk) = IZ(ikk) - 2^{ik}$

桁落ち $146 - (ikk - 1) \times 60 - ik$

丸めなし。 $(1 \leq ikk \leq 3, 0 \leq ik \leq 59)$

4.3　6倍精度

配列数　3

$IZ(3) \geq 2^{58}$　　$IZ(3) = IZ(3) - 2^{58}$　　**桁落ちなし.**

丸め位置 $IZ(1)(1:1)$

$2^{57} \leq IZ(3) < 2^{58}$　　$IZ(3) = IZ(3) - 2^{57}$　　**丸め位置** $IZ(1)(0:0)$

丸め 後　$IZ(3) = 2^{58}$　**桁落ちなし.**

$IZ(3) \neq 2^{58}$　**桁落ち**1

$2^{56} \leq IZ(3) < 2^{57}$　　$IZ(3) = IZ(3) - 2^{56}$　　**桁落ち**2　**丸めなし**

$IZ(3) < 2^{56}$　　$IZ(ikk) = IZ(ikk) - 2^{ik}$

桁落ち　$178 - (ikk - 1) \times 60 - ik$

丸めなし。　$(1 \leq ikk \leq 3, 0 \leq ik \leq 59)$

4.4　7倍精度

配列数　4

$IZ(4) \geq 2^{30}$　　$IZ(4) = IZ(4) - 2^{30}$　　**桁落ちなし.**

丸め位置 $IZ(1)(1:1)$

$2^{29} \leq IZ(4) < 2^{30}$　　$IZ(4) = IZ(4) - 2^{29}$　　**丸め位置** $IZ(1)(0:0)$

丸め 後　$IZ(4) = 2^{30}$　**桁落ちなし**

$.IZ(4) \neq 2^{30}$　**桁落ち**1

$2^{28} \leq IZ(4) < 2^{29}$　　$IZ(4) = IZ(4) - 2^{28}$　　**桁落ち**2　**丸めなし**

$IZ(4) < 2^{28}$　　$IZ(ikk) = IZ(ikk) - 2^{ik}$

桁 落ち　$210 - (ikk - 1) \times 60 - ik$

丸 めなし。　$(1 \leq ikk \leq 4, 0 \leq ik \leq 59)$

5. 仮数部のビット数と乗算のメモ
非対称行列の反復解法で使用した仮数部のビット数を調整した乗算のメモを紹介します。

仮数部のビット数に対する乗算の場合の桁上がりと丸め
多倍長整数乗算$Z = IX \times IY \quad (IX > 0, IY > 0)$**の結果**
が32**ビット整数変数の**30**ビットのみ使用するとします。**
符号部1**ビット,指数部**15**ビットを想定。**

(1)60**ビット**

配列数 5

$IZ(5) \geq 2$ **桁上がりあり** **丸め位置**IZ(3)(0 : 0)

$IZ(5) < 2$ **桁上がりなし** **丸め位置**IZ(2)(29 : 29)

(2)64**ビット**

配列数 5

$IZ(5) \geq 2^9$ **桁上がりあり** **丸め位置**IZ(3)(4 : 4)

$IZ(5) < 2^9$ **桁上がりなし** **丸め位置**IZ(3)(3 : 3)

(3)68**ビット**

配列数 5

$IZ(5) \geq 2^{17}$ **桁上がりあり** **丸め位置**IZ(3)(8 : 8)

$IZ(5) < 2^{17}$ **桁上がりなし** **丸め位置**IZ(3)(7 : 7)

(4)72**ビット**

配列数 5

$IZ(5) \geq 2^{25}$ **桁上がりあり** **丸め位置**IZ(3)(12 : 12)

$IZ(5) < 2^{25}$ **桁上がりなし** **丸め位置**IZ(3)(11 : 11)

(5)76**ビット**

配列数 6

$IZ(6) \geq 2^3$ **桁上がりあり** **丸め位置**IZ(3)(16 : 16)

$IZ(6) < 2^3$ **桁上がりなし** **丸め位置**IZ(3)(15 : 15)

6 DD,D (long double) ,DQのインライン化メモ

E5-2670,E5-2660,Phi5110P等で2つの変数の和で表す演算を使用した時,並列化効果を出すために作成したソースのメモを紹介します。

6.1 FORTRAN

(1) 加算 c=a+b

```
p1=a(1)
p2=a(2)
q1=b(1)
q2=b(2)
t1=p1+q1
t2=t1-p1
t3=(p1-(t1-t2))+(q1-t2)
t4=p2+q2
t5=t4-p2
t6=(p2-(t4-t5))+(q2-t5)
t7=t3+t4+t6
t8=t1+t7
t9=t8-t1
t10=(t1-(t8-t9))+(t7-t9)
c(1)=t8
c(2)=t10
```

(2)減算 c=a-b

```
p1=a(1)
p2=a(2)
q1=b(1)
q2=b(2)
t1=p1-q1
t2=t1-p1
t3=(p1-(t1-t2))-(q1+t2)
t4=p2-q2
t5=t4-p2
t6=(p2-(t4-t5))-(q2+t5)
t7=t3+t4+t6
t8=t1+t7
t9=t8-t1
t10=(t1-(t8-t9))+(t7-t9)
c(1)=t8
c(2)=t10
```

(3)乘算 c=a*b

p1=a(1)
p2=a(2)
q1=b(1)
q2=b(2)
t1=p1*q1
t2=p1*r
a1=t2-(t2-p1)
a2=p1-a1
t3=q1*r
b1=t3-(t3-q1)
b2=q1-b1
t4=((a1*b1-t1)+a1*b2+a2*b1)+a2*b2
t5=t4+p1*q2
t6=t5+p2*q1
t7=t6+p2*q2
t8=t1+t7
t9=t8-t1
t10=(t1-(t8-t9))+(t7-t9)
c(1)=t8
c(2)=t10

$$r = 134217729 = 2^{27} + 1(\textbf{倍精度})$$

$$r = 2^{57} + 1(4\textbf{倍精度})$$

(4)除算　c=a/b

```
p3=a(1)
p4=a(2)
q3=b(1)
q4=b(2)
ts=p3/q3
q1=ts*q3
q2=ts*q4
p1=p3
p2=p4
 t1=p1-q1
 t2=t1-p1
 t3=(p1-(t1-t2))-(q1+t2)
 t4=p2-q2
 t5=t4-p2
 t6=(p2-(t4-t5))-(q2+t5)
 t7=t3+t4+t6
 t8=t1+t7
 t1=ts
 t7=t8/q3
 t8=t1+t7
 t9=t8-t1
 t10=(t1-(t8-t9))+(t7-t9)
 c(1)=t8
 c(2)=t10
```

6.2 C

(1) 加算 c1=work11+work12

```
p1=work11[0] ;
p2=work11[1] ;
q1=work12[0] ;
q2=work12[1] ;
t1=p1+q1 ;
t2=t1-p1 ;
t3=(p1-(t1-t2))+(q1-t2) ;
t4=p2+q2 ;
t5=t4-p2 ;
t6=(p2-(t4-t5))+(q2-t5) ;
t7=t3+t4+t6 ;
t8=t1+t7 ;
t9=t8-t1 ;
t10=(t1-(t8-t9))+(t7-t9) ;
c1[0]=t8 ;
c1[1]=t10 ;
```

(2)減算 work1=one-xx

```
p1=one[0] ;
p2=one[1] ;
q1=xx[0] ;
q2=xx[1] ;
t1=p1-q1 ;
t2=t1-p1 ;
t3=(p1-(t1-t2))-(q1+t2) ;
t4=p2-q2 ;
t5=t4-p2 ;
t6=(p2-(t4-t5))-(q2+t5) ;
t7=t3+t4+t6 ;
t8=t1+t7 ;
t9=t8-t1 ;
t10=(t1-(t8-t9))+(t7-t9) ;
work1[0]=t8 ;
work1[1]=t10 ;
```

(3)乗算 work3=work1*work2

```
p1=work1[0] ;
p2=work1[1] ;
q1=work2[0] ;
q2=work2[1] ;
t1=p1*q1 ;
t2=p1*r  ;
a1=t2-(t2-p1) ;
a2=p1-a1  ;
t3=q1*r   ;
b1=t3-(t3-q1) ;
b2=q1-b1  ;
t4=((a1*b1-t1)+a1*b2+a2*b1)+a2*b2 ;
t5=t4+p1*q2 ;
t6=t5+p2*q1 ;
t7=t6+p2*q2 ;
t8=t1+t7 ;
t9=t8-t1 ;
t10=(t1-(t8-t9))+(t7-t9) ;
work3[0]=t8  ;
work3[1]=t10 ;
```

$r = 134217729 = 2^{27} + 1$(**倍精度**)

$r = 8589934593 = 2^{33} + 1$(**拡張倍精度**)

(4)除算 w3=work8/work2

```
p3=work8[0] ;
p4=work8[1] ;
q3=work2[0] ;
q4=work2[1] ;
ts=p3/q3 ;
q1=ts*q3 ;
q2=ts*q4 ;
p1=p3 ;
p2=p4 ;
 t1=p1-q1 ;
 t2=t1-p1 ;
 t3=(p1-(t1-t2))-(q1+t2)  ;
 t4=p2-q2 ;
 t5=t4-p2 ;
 t6=(p2-(t4-t5))-(q2+t5)  ;
 t7=t3+t4+t6 ;
 t8=t1+t7 ;
 t1=ts ;
 t7=t8/q3 ;
 t8=t1+t7 ;
 t9=t8-t1 ;
 t10=(t1-(t8-t9))+(t7-t9) ;
 w3[0]=t8  ;
 w3[1]=t10  ;
```