

# 規模 100 倍の並列処理の実行時間を予測するモデル構築方法

折居茂夫<sup>†1</sup> 山本義郎<sup>†2</sup>

問題の規模 100 倍のスケールの並列処理時間を予測するモデル化方法を提案する。この方法は、残差の連立不等式をモデルパラメータに非負の条件を付けて解く。線形計画法のシンプレックス法を有理演算して解くと、モデル化に不要な項は厳密に零となり過剰適合が自動的に抑制される。更に、解として得られた最大残差を用いて揺らぎ由来の項を削除し、予測力を向上することができる。

## A Modeling Method to Predict Parallel Runtime for Hundredfold Growth in Scale of Computation

SHIGEO ORII<sup>†1</sup> YAMAMOTO YOSHIRO<sup>†2</sup>

We propose a novel method to Predict Parallel Runtime for Hundredfold Growth in Scale of Computation. The method solves simultaneous inequalities for a residual to obtain model parameters restricted in the non-negative range. The simplex method of linear programming with rational arithmetic minimizes the residual and gives model parameters as an exact solution of the simultaneous inequalities. Because the exact zero value becomes the solution for model parameters whose bases functions create large residuals than the minimum residual, over fitting can be suppressed automatically. And furthermore, we can improve predictive power of the model by removing the other over fitting terms using the maximum of residuals in the solution obtained.

### 1. はじめに

並列処理の実行時間を予測する回帰モデルを構築する時、プログラムの実行時間を記述するために基底関数を増やして最小二乗法でモデルパラメータを決定すると、測定値の揺らぎをもモデル化して本来モデル化に必要無い基底関数が使われる過剰適合がしばしば生じる。この本来不要な基底関数項がモデルの外挿の時に効きだし、構築したモデルを予測に使えない場合もしばしば経験する。

並列処理の処理時間をモデル化する場合、この過剰適合はモデルパラメータに非負の条件を付けることにより緩和することを提案してきた[1, 2]。同時に数式処理の一つ限定記号消去法 (QE) を用いてモデルパラメータを決定することを提案した。本来モデル化に必要無い基底関数項の係数即ちモデルパラメータが整数の零値として得られるため、その項が本当に過剰適合項であるかを確認作業（例えばその項が想定レンジで他項と比べて無視できるかどうかを確認する作業）無しに、用いた入力データに対して過剰適合項であると判定することができるようになった。またパラメータ決定時の計算過程で自動的に過剰適合項が排除できるようになった。

また[1, 2]では QE を用いることにより与えられた基底関数と入力データに対して厳密なモデルパラメータ値が得られるようになることも報告した。モデルパラメータに非負の条件を付けてその値を最小二乗法で決定しようとする時、数値計算は繰り返し法を用いることになる。このとき 2 変

数、3 変数モデルのモデルパラメータをこの方法で求めると残差平方和の最小計算がうまくできず、2 変数では外挿に十分な精度が得られず、3 変数では予測ばかりか内挿もできない場合があるが、QE を用いるとこれらの現象は解消できる。

モデルパラメータが整数の零値として得られるのは、有理数演算のため、回帰計算の入力データに対して厳密な解として整数の零値を出力する場合があるためである。また数値計算の繰り返し法で決定できないモデルパラメータを QE で決定できるのは解の候補のみを全て列挙してその中から解を選ぶという、QE のアルゴリズムに由来する。

このようにモデルパラメータに非負条件を付けることと、その計算を QE で行うことによりモデルの過剰適合を緩和し予測力の向上を図れることを確認してきたが、文献[1]でも報告したように揺らぎ由来の過剰適合項が全て削除できるとは限らず、揺らぎの状態によっては予測精度を大幅に下げる過剰適合項が残る場合がある課題がある。そこで確認データを用いず、文献[1, 2]でモデル化の時得られた残差の最大値(後述する  $e_{max}$ )を用いて課題をクリアし予測力を大幅に向上することを試みた。 $e_{max}$ を利用して揺らぎレベルを同定し、閾値以下のモデルパラメータの基底関数項を棄却する方法である。

本報告では更に QE の代替として線形計画法のシンプレックス法を用いる方法を提案する。シンプレックス法を有理演算することにより、線形基底関数問題で解が一意的の場合、QE と同じ解を得ることができる。QE は全ての解の候補を列挙するため、演算量は大変大きなものになり計算時間もかかる。またモデルパラメータ数や入力データ数が増

<sup>†1</sup> GDE00740@nifty.com  
<sup>†2</sup> 東海大学  
Tokai University

すと必要なメモリ量が増え、パラメータの決定に要する時間やメモリ量の資源が非現実的なものになる。シンプレックス法を用いることによりこの制限が著しく緩和できた。

2章では文献[2]で報告した並列処理時間の予測モデルを構築する課題を整理する。3章では基本的アイデアとそれをシンプレックス法で解く方法を提案する。4章で  $e_{\max}$  を閾値として揺らぎ由来のモデル項を削除する方法を提案する。5章では得られたモデルから揺らぎを付けた入力データを作成し提案方法を確認し、6章でまとめを行う。

## 2. 並列処理時間の予測モデルを構築する課題

ここでは最小二乗法(LSM)で予測モデルを作る時の課題について述べる。目的変数  $y$  を説明変数  $x$  の  $M$  個の基底関数  $f_k(x)$  で説明する回帰モデル  $y=a_0+\sum a_k f_k(x)$  を考える。係数  $a_k$  はモデルパラメータである。このパラメータを  $(x_i, y_i)$  ( $i=1, \dots, N$ ) の観測データで推定する。LSMでは式(1)の残差平方和  $E$  を最小にする  $a_k$  を推定する。式を簡単にするため  $f_0(x_i)=1$  として  $k=0, \dots, M$  とすると  $E$  は式(1)となる。

$$E = \sum_{i=1}^N \left( y_i - \sum_{k=0}^M a_k f_k(x_i) \right)^2 \quad (1)$$

$E$  最小を得るため、両辺をモデルパラメータで次式のように偏微分し、 $\partial E / \partial a_k = 0$  として整理すると式(2)を得る。

$$\begin{aligned} \frac{\partial E}{\partial a_k} &= -2 \sum_{i=1}^N \left[ f_k(x_i) \left( y_i - \sum_{k=0}^M a_k f_k(x_i) \right) \right] \\ &= \sum_{i=1}^N [f_k(x_i) y_i - f_k(x_i) (a_0 f_0(x_i) + a_1 f_1(x_i) + \dots + a_M f_M(x_i))] = 0 \\ a_0 \sum_{i=1}^N f_k(x_i) f_0(x_i) + a_1 \sum_{i=1}^N f_k(x_i) f_1(x_i) + \dots + a_M \sum_{i=1}^N f_k(x_i) f_M(x_i) \\ &= \sum_{i=1}^N f_k(x_i) y_i \end{aligned} \quad (2)$$

ここで要素  $b_k = \sum_{i=1}^N f_k(x_i) y_i$  のベクトルを  $\mathbf{b}=(b_0, b_1, \dots, b_M)^T$ ,

要素  $g_{kj} = \sum_{i=1}^N f_k(x_i) f_j(x_i)$  ( $j=0, \dots, M$ ) の  $M \times (M+1)$  の行列を

$\mathbf{G}$  とし、 $\mathbf{a}=(a_0, a_1, \dots, a_M)^T$  とすると、モデルパラメータ  $a_i(i=0, \dots, M)$  は連立方程式  $\mathbf{Ga}=\mathbf{b}$  を解くことにより決定できる。

次に1番目の基底関数  $f_1$  がデータ  $y_i$  をモデル化するのに必要ない基底関数の場合を考え、そのような  $y_i$  を式(3)と仮定する。

$$y_i = a'_0 + a'_2 f_2(x_i) + \dots + a'_M f_M(x_i) + d_i \quad (3)$$

ここに  $d_i$  は回帰モデルの基底関数で表しきれないデータからの差異で、測定誤差等を総合したものである。この式(3)を式(2)の右辺に代入して整理すると式(4)を得る。

$$\begin{aligned} \sum_{i=1}^N f_k(x_i) f_0(x_i) (a_0 - a'_0) + \sum_{i=1}^N f_k(x_i) f_1(x_i) a_1 + \\ \sum_{i=1}^N f_k(x_i) f_2(x_i) (a_2 - a'_2) + \dots + \\ \sum_{i=1}^N f_k(x_i) f_M(x_i) (a_M - a'_M) = \sum_{i=1}^N f_k(x_i) d_i \end{aligned} \quad (4)$$

$$\text{ここで } b_k = \sum_{i=1}^N f_k(x_i) d_i,$$

$\boldsymbol{\alpha}=(a_0 - a'_0, a_1, a_2 - a'_2, \dots, a_{M-1} - a'_{M-1}, a_M - a'_M)^T$  として式(4)を  $\mathbf{G}\boldsymbol{\alpha}=\mathbf{b}$  で現わし、クラメル公式を適用すると

$\alpha_k = (\gamma_{1,k} b_1 + \gamma_{2,k} b_2 + \dots + \gamma_{M,k} b_M) / \det \mathbf{G}$  となる。ここに  $\det \mathbf{G}$  は  $\mathbf{G}$  の行列式、 $\gamma_{j,k}$  は  $\mathbf{G}$  の第  $(j, k)$  成分の余因子である。結果  $d_i=0$  の場合  $\alpha_k=0(k=1)$ ,  $\alpha_k=a'_k(k \neq 1)$  となるが、 $d_i \neq 0$  の場合  $\alpha_k \neq 0(k=1)$ ,  $\alpha_k - a'_k \neq 0(k \neq 1)$  となり  $f_1$  が過剰適合項となることがわかる。

このように LSM では入力データに揺らぎ等で生じる誤差  $d_i$  が存在すると、 $f_1$  のような不要な基底関数も  $E$  を最小化するために使われる。もしこの  $f_1$  が  $x$  の冪乗等の急激に増加する関数であれば、データ  $y_i$  の外挿領域においてその項の影響が大きくなり、結果モデルの予測力が低下する。従ってモデルパラメータの決定時にこの不要な基底関数項を見つけて削除することが、予測モデル構築における重要な課題の一つとなる。

## 3. 有理演算シンプレックス法による最大残差とモデルパラメータの決定

モデル化の対象を並列処理時間に限定すると入力データ  $y_i$  は非負となる。また一般に処理時間のモデル化において、処理時間は四則演算数、通信回数等に比例すると仮定してモデル化するため、各基底関数項も時間の単位を持つ。そこでこれらの項の係数即ちモデルパラメータが非負であるという仮定を設け、並列処理で過剰適合を引き起こす不要な基底関数をモデル構築時に見つけ出す方法を考えた。

本研究での不要な基底関数をモデル構築時に見つけ出すための情報源は、式(5)の各データに対する残差  $e_i$  である。

$$e_i = y_i - \sum_{k=0}^M a_k f_k(x_i) \quad (5)$$

全ての残差の情報を考慮するため、残差  $e_i$  の最大値以上のある残差  $e$  を用いて式(5)を式(6)のように連立残差不等式で表す。

$$e \geq \left| y_i - \sum_{k=0}^M a_k f_k(x_i) \right| \quad (6)$$

ここで式(6)を満足する  $e$  の最小値を求める。 $e$  の最小値が

つ  $e_i$  の最大値であるこれを、今後本論文では  $e_{\max}$  と記述する。  $e_{\max}$  は全ての  $e_i$  を基にして決定されているので、ある  $i$  の範囲の  $e_i$  だけを減少し、他の範囲では増加するような基底関数は  $e_{\max}$  を増すように働くと考えられ、そのような基底関数のモデルパラメータは零になると考えられる。

このような状態を具体的に把握するため、式(3)を式(6)に代入し、全ての  $i$  で成り立つ  $e$  の最小値  $e_{\max}$  を求めて代入すると式(7)を得る。

$$e_{\max} \geq |-a_1 f_1(x_i) + d_i + \varphi(x_i)| \quad (7)$$

ここに

$$\varphi(x_i) = (a'_0 - a_0) + (a'_2 - a_2) f_2(x_i) + \dots + (a'_M - a_M) f_M(x_i)$$

である。

$-a_1 f_1(x_i) + d_i$  が無ければ  $e_{\max}$  は零なので、これらの項が  $e_{\max}$  を決定している。従って  $e_{\max}$  は、 $\text{Max}(|d_i|)$  か或いはそれより小さい値となる可能性を持つ  $-a_1 f_1(x_i) + d_i$  と考えられる。ここでもしモデルパラメータ  $a_1$  が正負どちらの値をも取れるとすると、 $f_1(x_i)$  と  $d_i$  の和で  $\text{Max}(|d_i|)$  より小さい値となる可能性があり、その場合  $a_1 \neq 0$  となる。一方  $a_1 \geq 0$  とすると、 $\text{Max}(|d_i|) < |-a_1 f_1(x_i) + d_i|$  にならないために  $a_1 = 0$  であることが必要な場合が生じる。

このようにモデル化対象が時間であることを利用し、モデルパラメータに否負の条件  $a_k \geq 0$  を付けて  $e_{\max}$  を求め、それを基にモデルパラメータを決定し、 $e_{\max}$  より大きな残差を生じるモデルパラメータが零になるよう仕向けて過剰適合の緩和を図ることが、提案する方法の基本的アイデアである。

数値計算で式(2)の連立方程式  $\mathbf{Ga}=\mathbf{b}$  を解く場合、直接法を使ってモデルパラメータ  $a_i$  を得ることができる。一方非負条件  $a_k \geq 0$  を付けて解く場合は繰り返し法を用いることになり、変数の数が 2, 3 と増えるに従いパラメータを決定する際の残差平方和 SSD が増る。

この問題は式(5)の  $e_{\max}$  の実行可能解の候補を全て列挙してその中から解を選ぶことにより解決できる。我々は数式処理のアルゴリズムの 1 つ限量記号消去法(QE)を用いることによりこれを行う方法を提案してきた[1,2]が、本論文では QE より計算速度が速くより多くの変数や入力データを扱える線形計画法の解法の一つシンプレックス法を用いることを提案する。シンプレックス法は問題が作る凸空間の全ての端点を解の候補とするので、解の候補を全て列挙するのと同様な効果を得ることができる。本論文では更にシンプレックス法の計算を有理数演算で行うことを提案する。これにより  $f_k(x)$  中で、モデル化に必要な(最小残差  $e_{\max}$  より残差を大きくする)項のモデルパラメータの値が整数で零値となり、確実に過剰適合項を判別することが可能になる。また零値により過剰適合項の自動的削除が可能になる。

まず 2 種類の  $N$  個の入力データ  $y_i, x_i (i=1, \dots, N)$  と  $M$  個の

基底関数  $f_k(x) (k=0, \dots, M)$  を用いて  $e_{\max}$  とモデルパラメータを決定する。式(6)にモデルパラメータに正の条件を付けて表わすと式(8)となる。ベクトル  $\mathbf{x}_i$  は  $y_i$  が複数の変数の場合もあることを意味する。

$$\begin{aligned} -e \leq y_i - \sum_{k=0}^M a_k f_k(\mathbf{x}_i) \leq e \\ a_k \geq 0 \end{aligned} \quad (8)$$

次に  $\mathbf{x}_i$  に対する基底関数  $f_k(\mathbf{x}_i)$  の値  $F_{i,k}$  を使い、式(8)を線形化し式(9)とする。

$$\begin{aligned} -e \leq y_i - \sum_{k=0}^M F_{i,k} a_k \leq e \\ a_k \geq 0 \end{aligned} \quad (9)$$

式(7)のように全ての不等式で成り立つ  $e$  の最小値  $e_{\max}$  と否負のモデルパラメータ  $a_k$  は、式(9)を式(10)の線形計画法の問題に書き直し、式(10)を有理数演算のシンプレックス法で解くことにより、過剰適合項のモデルパラメータ値が整数の零値となる厳密な解として得ることができる。

$$\begin{aligned} e + \sum_{k=0}^M F_{i,k} a_k &\geq y_i \\ e - \sum_{k=0}^M F_{i,k} a_k &\geq -y_i \\ e_{\max} = e &\rightarrow \text{Min} \end{aligned} \quad (10)$$

#### 4. $e_{\max}$ を閾値とした揺らぎ由来項の削除

非負条件を付加したことにより揺らぎ等に起因する過剰適合は式(10)で緩和できる。一方入力データに人為的に揺らぎを持たせたデータを作成し、モデル化を 100 回試みるとモデルの外挿を困難にする過剰適合項が残る現象が観測される場合があった[2]。予測精度を向上し、100 倍の規模の計算の処理時間を予測するモデルを作るためには、このような現象を抑制する必要がある。そこで本報告では式(10)で得られた  $e_{\max}$  を基準にしてモデルの基底関数項の取捨選択を行う方法を提案する。

図 1 に過剰適合の抑制方法を示す。まず式(10)を有理数演算のシンプレックス法(LP)で解き、 $e_{\max}$  より大きな残差を生じる基底関数モデルのモデルパラメータ  $a_k$  が整数の零となる。このように有理数演算 LP を用いると、一部の過剰適合は自動的に抑制される。

次に自動的に抑制されない過剰適合を得られた  $e_{\max}$  を用いて判別する。まずモデルから  $a_i = 0$  の項を削除する。次に残った項を 1 項ずつ削除しその  $(e'_{\max})_j$  を算出する。判別は、この値を  $e_{\max}$  で除した式(11)の  $\lambda_j$  を閾値  $\lambda_{th}$  と比較して行う。

$$\lambda_j = \frac{(e'_{\max})_j}{e_{\max}} \quad (11)$$

予めモデルから  $a_1=0$  の項を削除して  $(e'_{\max})_j$  を計算することにより、 $\lambda_j$  の値を大きくすることが重要である。閾値  $\lambda_{th}$  より小さい  $\lambda_j$  の項をモデルから削除することにより、過剰適合をより緩和してモデルの予測力を向上することができる。この閾値、本論文では  $\lambda_{th}=2$  とした。

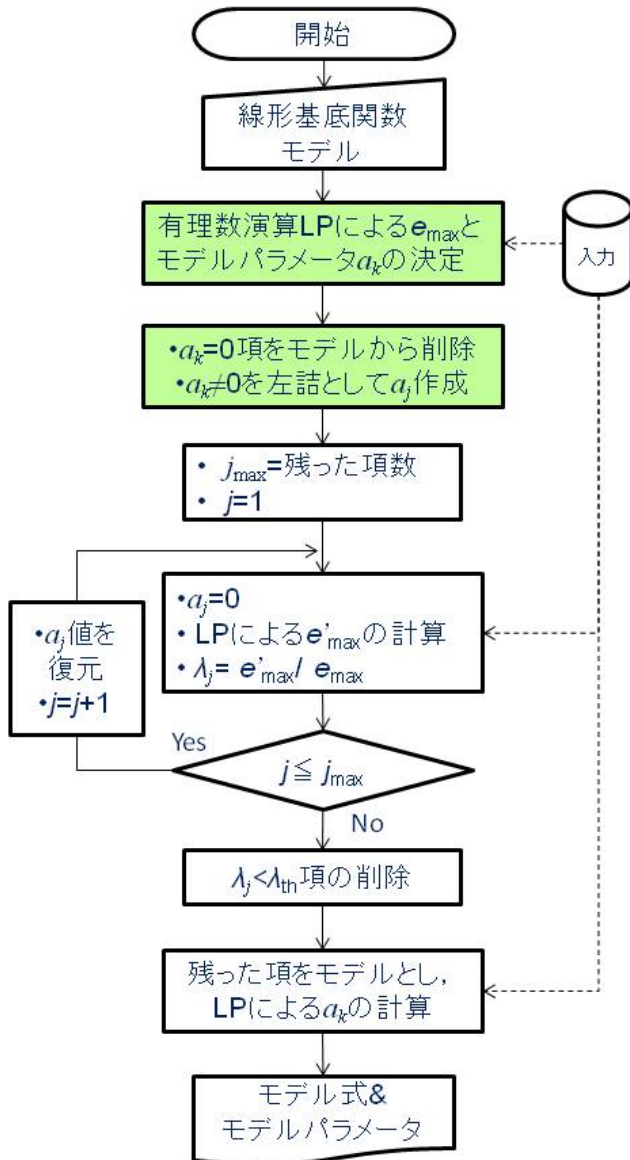


図1 LPの解  $a_k=0$  と  $e_{\max}$  を用いた過剰適合の抑制

## 5. 適用事例

モデルパラメータが正という条件を付けることにより、モデルの予測力が向上すること、LPを有理数で計算することによりモデルパラメータが零になる場合が生じて過剰適合を自動的に抑制できること、この計算で得られた  $e_{\max}$  を用いて予測に必要なモデル項を見分けること。これらを4つの事例で確認する。

LPオペレーションを行うツールとして Mathematica [3] バージョン 8 の LinearProgramming 関数を用いた。この関

数は入力を有理数とすると厳密な有理数または整数の結果を得ることができる。

### 5.1 $y=x^2$ のモデル化と過剰適合の抑制

$y(x)=x^2$  を基に生成したデータ  $y_i$  を基底関数  $f(x)=a_0+a_1 \cdot x+a_2 \cdot x^2+a_3 \cdot x^3+a_4 \cdot x^4+a_5 \cdot x^5$  ( $M=5$ ) でモデル化した。 $y_i$  は、1.1 から 1.85 を等間隔に 8 点で分割した  $x_i$  に対する  $y(x_i)$  で標準偏差  $\sigma=0.01$  のばらつきを持たせた。比較のため最小二乗法 (LSM) の結果を列挙する。LSM のツールとしては Mathematica の FindFit を用いた。この関数はモデルパラメータの正負の条件付けをすることができる。またモデルパラメータの初期値を設定することができる。

表1はこの入出力値を用い、 $F_{1,k}(a_k)=f(x_i)$  ( $k=0, \dots, 5$ ) として LSM で得られたモデルパラメータで、図2のように過剰適合により正解の  $a_2=1$ 、他は零とは大きく異なる。

表1 モデルパラメータ ( $\sigma=0.01$ )

	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$
LSM	-63.3	224.	-314.	220.	-76.4	10.5

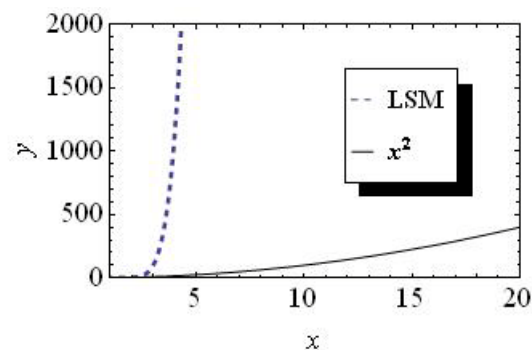


図2 表1-1による  $f(x)$  の外挿

次に  $a_k \geq 0$  の条件を付け、式(10)により LP で決定したモデルパラメータと LSM で得られたものを表2に示す。両者において、モデルパラメータの大きな変動が抑制されることが確認できる。また LP では  $a_0, a_3, a_4$  のモデルパラメータを整数の零値として得られ、過剰適合が自動的に抑制されたことが確認できる。

表2 モデルパラメータ ( $\sigma=0.01, a_k \geq 0$ )

	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$
LP	0	0.0498	0.962	0	0	0.00126
LSM	2.83	4.09	0.932	1.29	2.20	7.11
	$\cdot 10^{-2}$	$\cdot 10^{-2}$		$\cdot 10^{-2}$	$\cdot 10^{-3}$	$\cdot 10^{-4}$

次に図1を適用してLPでモデルパラメータ選択を行った場合を示す。入力の揺らぎにより結果が変わることが予想されるため、同じ  $y(x)$  と  $\sigma$  に対して揺らぎが異なる入力を100ケース用意して確認した。図3(a)は各モデルパラメ

ータを削除した時の  $\lambda$  の値である。  $a_2$  項の  $\lambda$  が赤線の閾値  $\lambda_{th}=2$  より十分大きくなり、他の項の  $\lambda$  は大半が 2 以下であることが確認できる。この選択処理により 95 ケースの入力に対して  $a_2$  項のみを選択できることを確認した。尚  $\lambda_{th}$  を用いなかった場合、100 ケースにおいて  $a_2$  項以外が全て零となることは無かった。

図 3 (b)は LP の代わりに LSM を用い、  $e_{max}$  の代わりに残差平方和 RSS を使い、数値計算した場合の  $\lambda_{RSS}=(RSS')_j/RSS$  を示す。  $a_2$  項の  $\lambda$  が赤線の閾値  $\lambda_{th}=2$  より大きくなり、他の項の  $\lambda$  は大半が  $\lambda_{th}$  より十分小さいことが確認できる。

これらのことからモデルパラメータに非負条件を付けることにより、モデルの予測力が向上することが確認できた。

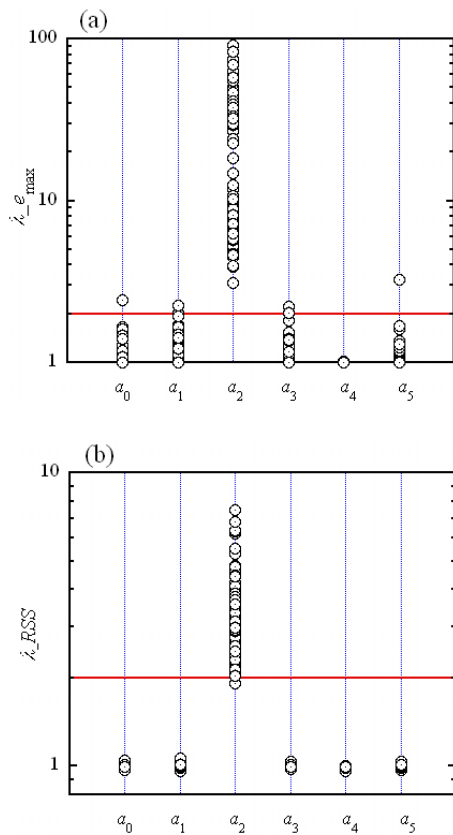


図 3  $\lambda$  を用いたモデル項の選別

### 5.2 $y=x^2+\log_2(x)$ のモデル化と過剰適合の抑制

5.1 節では 1 つの関数から作った入力に対するモデル化だったが、現実の並列処理時間のモデルでは、は複数の関数の重ね合わせが必要となる。このような場合に対しても 5.1 節と同様に過剰適合が緩和される場合がある。

これを確認するために、  $x^2$  に加えて並列処理の通信時間等のモデル化で良く使われる  $\log_2(x)$  を用いて  $y(x)=x^2+\log_2(x)$  のデータを作り、それに対するモデル

$$f(x)= a_0+a_1x+a_2x^2+a_3x^3+a_4x^4+a_5x^5+a_6\log_2(x)+a_7 \cdot \sqrt{x} \quad (M=7)$$

を用いた。データは 5.1 節と同じ 1.1 から 1.85 を等間隔に 8 点で分割した  $x_i$  に対する  $y(x_i)$  で、  $\sigma=0.01$  のばらつきを持たせた。ここで 8 つのパラメータを  $a_k \geq 0$  とし、100 ケースのデータを用いて確認を実施した。

表 3 に LP と LSM のモデルパラメータの 1 例を示す。

LP では  $a_0, a_1, a_4, a_5, a_7$  が零となり、入力データのモデル化にこれらの項が必要ないことがわかる。

表 3 モデルパラメータ ( $\sigma=0.01$ )

	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$
LP	0	0	0.970	0.983· 10 <sup>-2</sup>	0	0	1.07	0
LSM	1.32· 10 <sup>-2</sup>	2.22· 10 <sup>-5</sup>	0.963	1.14· 10 <sup>-2</sup>	7.57· 10 <sup>-7</sup>	1.09· 10 <sup>-7</sup>	1.06	3.22· 10 <sup>-5</sup>

5.1 節と同様に、図 1 を適用してモデルパラメータ選択を行った場合を入力 100 ケースに対して確認した。図 4 (a) は各モデルパラメータを削除した時の  $\lambda$  の値である。  $a_2$  と  $a_6$  項の  $\lambda$  が赤線の閾値  $\lambda_{th}=2$  より十分大きくなり、他の項の大半の  $\lambda$  は 2 以下であることが確認できる。この選択処理により 60 ケースの入力に対して正解の  $a_2$  と  $a_6$  項のみを選択したことを確認した。

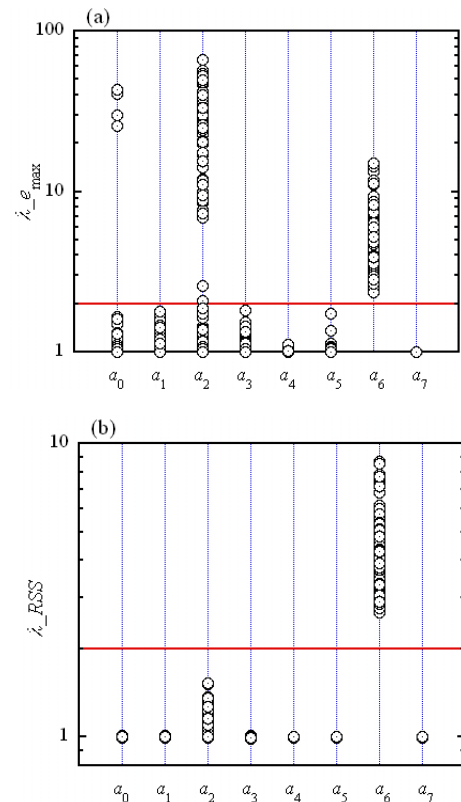


図 4  $\lambda$  を用いた過剰適合項の選別

図 4 (b)に  $e_{max}$  の代わりに残差平方和 RSS を使った  $\lambda$  を示す。100 ケース全ての  $a_6$  項の  $\lambda$  が赤線の閾値  $\lambda_{th}=2$  より大きくなる。また  $a_0, a_3$  項の  $\lambda$  は 1 に近く、モデル化に必要な無



いと判断できる. 一方全ての  $a_2$  項の  $\lambda$  は 2 より小さく 1 に近いものもあるので, 単純に選択できる  $a_2$  項の数を数えることはできない.  $a_2$  と  $a_6$  項両方のみが選ばれる率は LP で 60%, LSM では 0% である.

これらのことから, 関数に対数が入っても, モデルパラメータに非負条件を付けて LP でモデルパラメータを決定すると,  $\lambda$  によって入力データの揺らぎ由来でない項を判別できる場合があることが確認できる.

### 5.3 分子動力学計算の並列処理時間の予測

分子動力学プログラムを並列計算機 IBM SP2 で実行した時の処理時間モデルを使い, 提案した方法により過剰適合緩和と予測力向上を確認する. 並列処理時間モデル化に対するデマンドとして次の 3 つを考慮した.

- 1) 日々の計算で得られる処理時間が入力
- 2) なるべく少ないデータ数でのモデル構築
- 3) 精度の良い予測モデル

1) として, 入力データは数値シミュレーション等の処理を投入してから終了するまでの時間, ウォールクロックタイムとした. 1), 2) を想定し, また本提案が有効だった 6.1, 6.2 節の入力データ数を考慮し, 入力データ  $y_i$  はプロセッサ数  $p=\{2, 4, 6, 8, 10\}$  と, 問題の大きさ  $n=\{7200, 12800, 20000, 39200\}$  の組み合わせで 20 通りの測定した実行時間からランダムに 7 点を抽出したものを 5 ケース用いた. 7 つ日々の計算の中でモデル構築が可能となる勘定である. また予測精度を確認するため, 入力データより大きな最大  $p=48, n=96800$  規模の計算の処理時間を測定して確認を行った.

#### (1) 処理時間モデル

モデルの基底関数はプログラムから抽出した. モデル化した分子動力学プログラムは, ファンデルワールス力と重力の影響下で時間発展するアルゴン原子の巨視的振舞いをシミュレーションする分子動力学プログラムである[4].

プログラム全体の処理時間を式(12)のように, 8 つのモデルパラメータを持つ 2 変数  $p, n$  のモデルとして記述することができる[2]. 式(12)の各項は時間項を表すので, モデルパラメータは全て非負と置けるものとした.

$$T_{MD}(p, n) = a_0 + a_1 \cdot n^2 + a_2 \cdot \frac{n^2}{p} + a_3 \cdot n + a_4 \cdot \frac{n}{p} + a_5 \cdot n \cdot p + \frac{a_6}{p} + a_7 \cdot p \quad (12)$$

#### (2) LP に過剰適合の抑制

モデル式(12)と前述した 5 つの入力データに図 1 の方法を適用した. 入力データは幾つかのモデルパラメータが零値になることを想定して 7 点として表 3 の結果を得た. 表は全ケースで  $a_6$  が零となり, 過剰適合項が  $a_6$  であることが分かる. 同時に他のモデルパラメータはこの過剰適合項を削除した時の値となる.

表 3 LP により決定した式(12)のモデルパラメータ

No	1	2	3	4	5
$a_0$	1.73	0	64.0	0	22.0
$a_1(10^{-7})$	1.73	0	0.888	0	4.80
$a_2(10^{-6})$	4.00	4.26	5.39	2.74	1.06
$a_3(10^{-2})$	3.51	3.97	2.66	4.18	3.77
$a_4$	0.294	0.290	0.2851	0.302	0.304
$a_5(10^{-4})$	7.02	0	10.0	0	0
$a_6$	0	0	0	0	0
$a_7$	2.12	4.01	0	0	1.30

この表のモデルパラメータを用いて問題の大きさ  $n=96800$  の処理時間の予測を確認した結果を図 5 に示す. 図はプロセッサ  $p$  に対する経過時間  $T_{MD}$  である. 黒の×印は予測確認用の測定値である. 緑線が LP を用いたモデルの予測である. 緑色のモデルによる予測は, データの組み合わせを変えても黒の×印の値と傾向を予測できることを示す. またデータとして用いた測定値を青色の×印で示す. モデル構築に使用したデータの時間域と予測される時間を比較するため, 5 ケース全てのデータをプロットした.

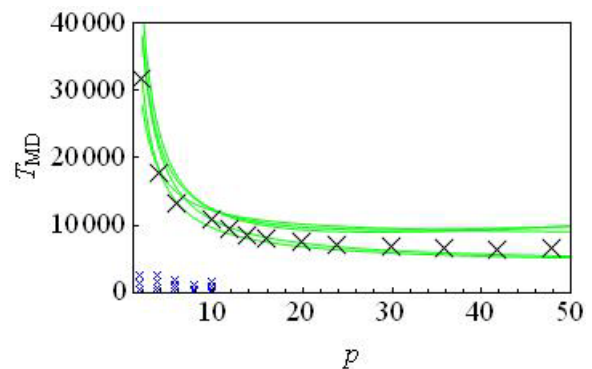


図 5 LP によるモデルの予測の確認 (MD)

図 1 を適用してモデルパラメータ選択を行った場合を入力 5 ケースに対して確認した. これを図 6 に示す.

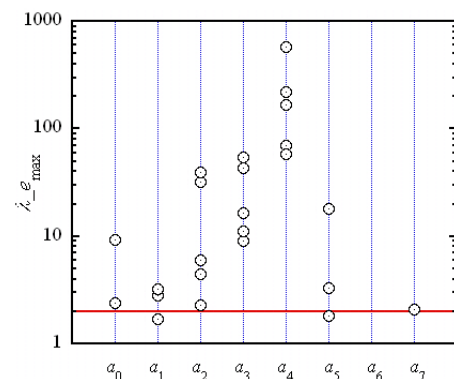


図 6 LP による MD コードの  $\lambda$

図 6 は  $\lambda_{th}=2$  を超えるモデルパラメータが  $a_2, a_3, a_4$  であることを示すそこで, これらの項のみで式 (13) のようにモ

デル化してモデルパラメータを決定した。これを表4に、予測の確認結果を図7に示す。図5に比べ予測の範囲が狭まりまた確認データと一致し、予測力が向上していることがわかる。

$$T_{MD}(p, n) = a_2 \cdot \frac{n^2}{p} + a_3 \cdot n + a_4 \cdot \frac{n}{p} \quad (13)$$

表4 LPにより決定した式(13)のモデルパラメータ

No.	1	2	3	4	5
$a_2(10^{-6})$	2.73	2.39	3.08	2.74	2.48
$a_3$	0.0416	0.0481	0.0448	0.0418	0.0476
$a_4$	0.303	0.283	0.292	0.302	0.285

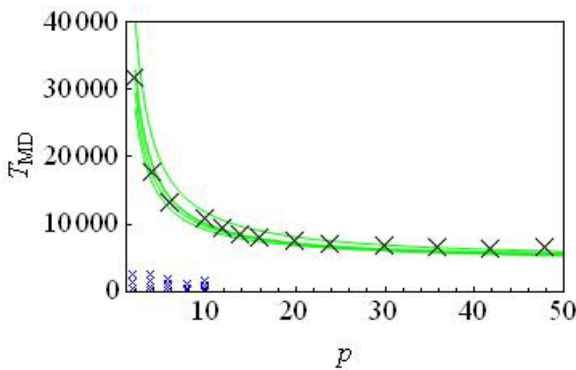


図7  $\lambda_0=2$  で構築したモデルの予測の確認テスト (MD)

図8はMDの処理時間のモデル化に対して図1の $e_{\max}$ の代わりに残差平方和RSSを使った $\lambda$ である。図は $a_4$ のみが閾値2を超えることを示す。従ってLSMでは、複雑な基底関数のMDには、図1の手法を用いて図7のように予測精度を改善できない。図に示されるように、得られる $\lambda$ は値が1より小さい場合がある。

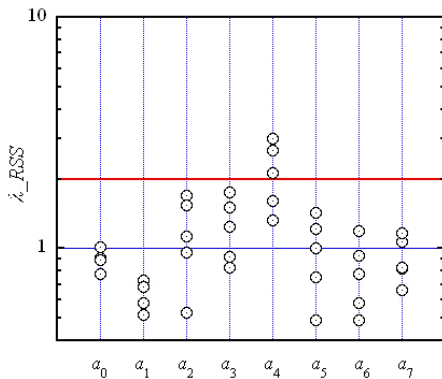


図8 LSMによるMDコードの $\lambda$

### 5.4 HPLの並列処理時間の予測

100倍規模の問題の処理時間予測を、HPL[5]を統計数理研究所のPRIMERGY RX200S5スーパーコンピュータシス

テムを利用して確認した。このシステムはインテル Xeon プロセッサ X5570 (2.93GHz) を搭載した 360 台の PRIMERGY RX200S5 から成り、ノード間は InfiniBand QDR で接続されている。CPU 数 (コア数) は 2(8) である。

HPLは次元数 $N$ の密行列計算を $P \times Q$ の2次元プロセス格子に分割し、並列処理するプログラムである。その処理時間モデルは計算量が処理時間に比例すると仮定して得られる[5]。このとき panel factorization と trailing sub matrix 計算で生じる通信のモデルパラメータが異なることを想定して各々の通信時間を  $t_{cf} = \alpha_f + \beta_f L$ ,  $t_{cm} = \alpha_m + \beta_m L$  とする。ここに  $L$  は通信量である。次にネットワークがで通信することを考慮し、立ち上がり時間  $\alpha_f$  と  $\alpha_m$  が  $P$  に比例すると仮定して  $a = aP + b$  として処理時間を求めると、処理時間モデルは式(13)のように  $P, Q, N$  の3変数7パラメータモデルとなる[2]。

$$T_{HPL}(P, Q, N) = \gamma \frac{2N^3}{3PQ} + \beta_f \frac{N^2}{2P} + \beta_m \frac{3N^2}{2Q} + a_f P \frac{N(1 + NB \log_2(P))}{NB} + a_m P \frac{N(P - 1 + \log_2(P))}{NB} + b_f \frac{N(1 + NB \log_2(P))}{NB} + b_m \frac{N(P - 1 + \log_2(P))}{NB} \quad (14)$$

ここに  $NB$  はプロセス格子のブロックサイズ。  $\gamma, \beta_f, \beta_m, a_f, a_m, b_f, b_m$  がモデルパラメータである。

入力データ即ちウォールクロックタイム  $y_i$  は、  $NB=100$  に対して問題の大きさ  $N = \{25000, 35000\}$  とプロセッサ数  $P=Q = \{10, 11, 12, 13, 14, 15, 16\}$  から各  $N$  に対してランダムに選んだ2点(計4点)と、  $N = \{20000, 30000\}$  とプロセッサ数  $Q=8, P = \{32, 40, 48, 56, 64, 80, 96, 112, 128\}$  から各  $N$  に対してランダムに選んだ2点(計4点)の総計8点を1ケースとし、前節と同様5ケースの入力を用意した。

#### (1) LPによるモデル化と予測力向上

式(10)を用いて式(13)のモデルパラメータを決定したものを表5のNo.1~5に示す。表で値が零の項は、過剰適合項が自動的に削除されたことを示す。図8はこのモデルパラメータを用い、最大630倍の規模の計算の予測の確認結果である。(モデル化に使用した  $N$  の最大値は 35000 ゆえ  $630 \sim (300000/35000)^3$ 。○と×印の各6点は予測確認のために測定した時間である。○は  $Q=8, P=128$  と固定、×は  $Q=64, P=16$  と固定し、 $N$  を変えて測定した。図中実線は○の予測、破線は×の予測である。図は5ケース全てのモデルパラメータが  $N=150000$  即ち約80倍まで一致した予測をしているが、それ以上の  $N$  ではばらつきが出ることを示す。このように100倍のレンジの予測は、LPによる過剰適合項の抑制だけでは不十分な場合がある。

表5 LPにより決定した式(14)のモデルパラメータ

No.	$\gamma$	$\beta_f$	$\beta_m$	$a_f$	$a_m$	$b_f$	$b_m$
1	$8.60 \cdot 10^{-11}$	0	$4.63 \cdot 10^{-8}$	0	0	$2.88 \cdot 10^{-5}$	$1.42 \cdot 10^{-4}$
2	$1.71 \cdot 10^{-11}$	$2.07 \cdot 10^{-7}$	$9.43 \cdot 10^{-8}$	0	$1.52 \cdot 10^{-6}$	0	0
3	$2.68 \cdot 10^{-11}$	$2.07 \cdot 10^{-7}$	$7.92 \cdot 10^{-8}$	$3.38 \cdot 10^{-7}$	0	0	0
4	$6.25 \cdot 10^{-11}$	$5.71 \cdot 10^{-8}$	$8.89 \cdot 10^{-8}$	$1.03 \cdot 10^{-7}$	$5.45 \cdot 10^{-7}$	0	0
5	$6.97 \cdot 10^{-11}$	$2.83 \cdot 10^{-8}$	$8.46 \cdot 10^{-8}$	0	$8.17 \cdot 10^{-7}$	$1.02 \cdot 10^{-5}$	0

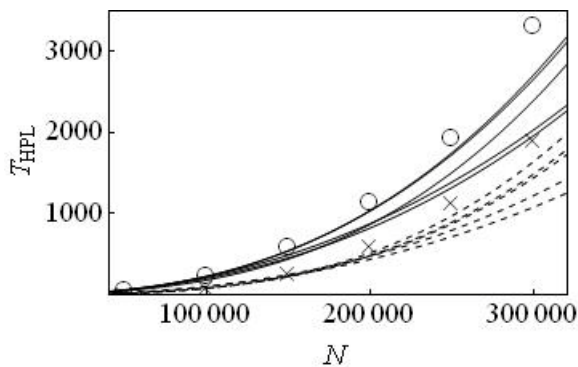


図8 HPL 処理時間モデルの確認テスト (HPL)

そこで4章で提案した  $e_{\max}$  から揺らぎに起因して決まる項を判別するための  $\lambda$  を計算した. これを表6に示す. 図1に沿って全て  $\lambda_{th}=2$  以下の  $b_f$  と  $b_m$  のモデル項を削除する. 2以上の値が1回しか出ない  $\beta_f$  と  $a_f$  も削除することになると, 式(15)を得る. この式と表5で用いた入力データを用いてモデルパラメータを決定すると表7を得る. ケースNo.1を除いてケース間のモデルパラメータ値のばらつきが少ない. 図9はこれを用いた予測の確認である. 図8に比べて各ケースの予測のばらつきは減少し, ○印の場合, 図の全レンジで予測ができてることが確認できる.

表6 式(14)と表(6)の  $\lambda$

No.	$\gamma$	$\beta_f$	$\beta_m$	$a_f$	$a_m$	$b_f$	$b_m$
1	4.43	-	1.48	-	-	1.48	1.08
2	1.10	3.00	12.0	-	2.36	-	-
3	1.66	1.34	4.55	2.67	-	-	-
4	2.24	1.12	8.73	1.15	1.01	-	-
5	8.00	1.00	1.64	-	2.08	1.52	-

$$T_{HPL}(P, Q, N) = \gamma \frac{2N^3}{3PQ} + \beta_m \frac{3N^2}{2Q} + a_m P \frac{N(P-1+\log_2(P))}{NB} \quad (15)$$

表7 LPにより決定した式(15)のモデルパラメータ

No.	$\gamma$	$\beta_m$	$a_m$
1	$5.64 \cdot 10^{-11}$	$1.18 \cdot 10^{-7}$	0
2	$7.29 \cdot 10^{-11}$	$1.08 \cdot 10^{-7}$	$2.05 \cdot 10^{-8}$
3	$9.42 \cdot 10^{-11}$	$7.47 \cdot 10^{-8}$	$2.40 \cdot 10^{-6}$
4	$8.37 \cdot 10^{-11}$	$8.92 \cdot 10^{-8}$	$1.12 \cdot 10^{-6}$
5	$7.62 \cdot 10^{-11}$	$9.49 \cdot 10^{-8}$	$9.89 \cdot 10^{-7}$

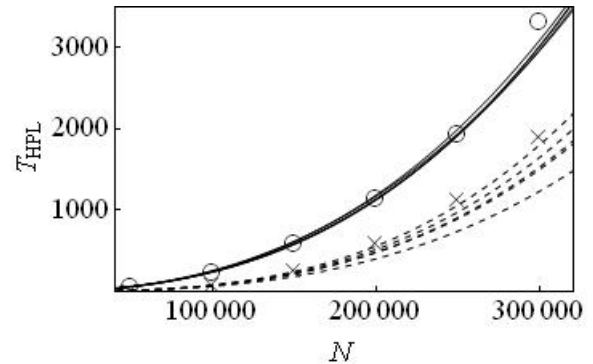


図9  $\lambda_{th}=2$  で構築したモデルの予測の確認 (HPL)

## 6. おわりに

非負のモデルパラメータを持つ並列処理時間の回帰モデルとモデルパラメータ決定に用いた入力データの残差の最大値  $e_{\max}$  を用い,  $e_{\max}$  値の2倍以下にしか寄与しない基底関数項を削除することによりモデルの予測力向上を図る方法を提案した. 揺らぎに起因する可能性がある基底関数項を削除することにより, 100倍規模の予測の可能性を例示した. なるべく実プログラムと実際の並列処理に近いデータを用いて本提案の有効性を確認していくことが今後の課題と考える.

## 参考文献

- 1) 折居茂夫: 限量記号消去法による時間モデルパラメータの決定, 情報処理学会研究報告, 2012-HPC-134, No.11 (2012).
- 2) 折居茂夫, 山本義郎: 限量記号消去法を用いた回帰モデルの予測力向上, 情報処理学会研究報告, 2013-HPC-139, No.1 (2013).
- 3) Mathematica, <http://reference.wolfram.com/mathematica/guide/Mathematica.en.html>
- 4) Watanabe, T. and Kaburaki, H.: Increase in chaotic motions of atoms in a large-scale self-organized motion, Phys. Rev. E, Vol.54, pp. 1504-1509 (1996).
- 5) A. Petitet, R. C. Whaley, J. Dongarra and A. Cleary: HPL - A Portable Implementation of the High-Performance Linpack Benchmark for Distributed-Memory Computers. <http://www.netlib.org/benchmark/hpl/>