ＩＳＳＮ 1341－6839

情報処理センター
研　究　報　告

The Bulletin of the Information Processing Center

第　28　号

（２００７.３）

岡 山 理 科 大 学

Okayama University of Science

岡 山 市 理 大 町 １ － １

℡（086）252－3012（直）

目　　次

「高速数値計算の実験」

総合情報学部情報科学科
河野敏行

1. 課題概要

センター実習室の複数の PC を利用して，数値計算を分散させ，大規模な問題を高速に解くことを目標として実験に取り組んだ．昨年度は，学生の利用していない時間帯に，クライアント PC を Linux 系 OS で立ち上げ，その上で，並列計算を試みる実験を考慮し準備を行っていた．この場合，夜間などの学生が利用していない時間を使用することとなるが，多くの問題があり，断念した．そこで，今回は実験室の PC にインストールされている Windows XP の環境で計算環境を構築し，数値計算を行う実験を行った．

全体の構想として，Web サーバを立ち上げ，サーバ上に Web コンテンツとして計算管理システム，計算プログラムを準備しておき，各クライアント PC からサーバに Web アクセスを行うことでブラウザを介して，その PC を 1 クラスタとしてサーバに登録する方法を考えた．管理を行う PC からは計算管理ページを開くことで，各クライアントの状態を管理し，各種の計算の実行・停止・結果の表示などの管理を行う．

計算システム自体は試作段階であるが，簡単な数値実験の結果を示す．

2. 問題設定について

現在，コンピュータの高速化が進み，より複雑な問題を取り扱うことが可能となった．問題が複雑となるに従い，その解析には多くの時間がかかり，1 台の計算機では，解くことが困難となる．特に差分法を用いて得られる次式で示されるような線型方程式を解くことに多くの時間が費やされる．

$$Ax = b$$

ここで，A は $n \times n$ 行列，x は n 次の未知ベクトル，b は n 次の既知ベクトルである．ベクトル b は，真の解 $x^* = (1,1,\cdots,1)^T$ を与え，$b = Ax^*$ を計算することで与え，初期近似ベクトルは $x^{(0)} = (0,0,\cdots,0)^T$ と設定して定常反復法を用いて計算することとする．定常反復法は Gauss-Seidel 法を用い，1CPU の場合とさらに PC クラスタを利用し分散計算をさせた場合とを比較した．一般に，差分法で得られる線型方程式の係数行列は疎な行列となる．すなわち，比較的零要素が多い問題である．今回はテスト問題として，以下に示す A のような行列を扱うこととする．すなわち対角成分をすべて 1.0，上下に 1 ないし 2 つの要素を持つ 3 重または 5 重対角行列とする．

$$A=\begin{bmatrix} 1.0 & p & q & 0 & \cdots & 0 \\ r & 1.0 & p & q & \ddots & \vdots \\ s & r & 1 & p & \ddots & 0 \\ 0 & s & r & 1 & \ddots & q \\ \vdots & \ddots & \ddots & \ddots & \ddots & p \\ 0 & \cdots & 0 & s & r & 1 \end{bmatrix}$$

ここで，p,q,r,s は任意に設定するとする．3 重対角行列の場合は $q=s=0$ とする．

数値実験において扱う係数行列は理論上取り扱いを容易にするために，p,q,r,s を非正とし，$|p|+|q|+|r|+|s|<1$ となるように設定することとする．すなわち，係数行列 A が狭義優対角な Z 行列となる．PC クラスタによる分散処理を行うために係数行列 A を PC クラスタの数によって行の数を分けることにする．そして，それぞれの PC クラスタで得られた近似解を反復ごとにサーバ側で結合し，再度，各 PC クラスタのデータとして与えることとした．すなわち，係数行列 A が n 行，PC クラスタの数が k 台ある場合，1 行目から n/k 行目を 1 番目の PC クラスタが計算することとする．また，割り切れない場合は四捨五入を行い，最後の PC クラスタにおいて計算する行を調整することとした．並列計算の方法としてはさまざまな方法がある[1]が，今回は複雑な処理はせず，毎回の計算で確実に処理できる方法を選んだ．この方法では，サーバと PC クラスタの通信が多く，効率的ではない．また，反復計算に前処理[2]を用いることによって高速化することも可能であるが，今回の実験では省略した．

3.　計算機環境

この研究では Windows 上で分散計算を行う．また，クライアント側に新たにアプリケーションを追加することなく，計算を行うこととしたい．したがって 1 台の Web サーバを立て，そこに Web アクセスする各 PC を PC クラスタとして登録し，分散計算を行い，その結果を管理 PC で集約し管理する．その構成を図 1 に示す．

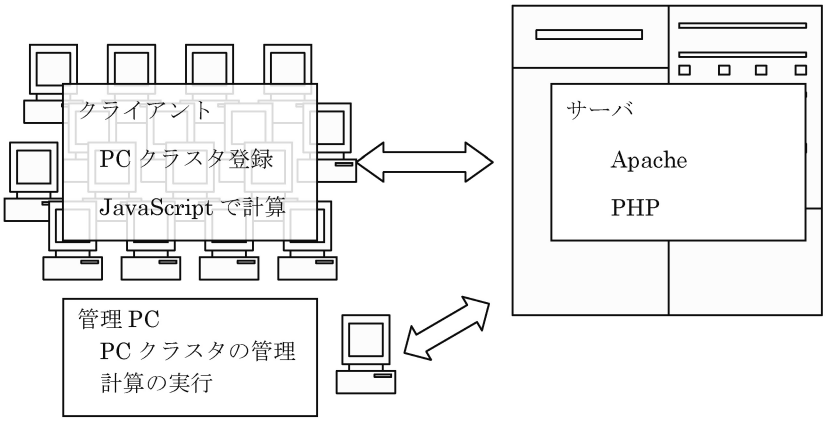

図 1　サーバとクライアント，管理 PC

図 1 において，サーバには Web サーバとして Apache を利用し PHP で各プログラムを作成した．クライアント側では JavaScript で記述されたプログラムで計算が行われるが，これらはサーバ側で PHP によって各 PC クラスタに合わせて生成されたプログラムである．通信は通常の Web 接続（80 番ポート）を利用し，サーバに保存された数値データを用いて計算を行っている．

以下に計算の流れを箇条書きで示す．

○サーバ及び管理 PC 上

- 全体のデータを処理
- アクセスした PC に自動的に番号を割り振る
- 計算プログラムを配信する
- ①各データを配信し，計算をさせる
- ②計算結果を集約し，データを更新する
- ①②を収束するまで繰り返す．
- 結果を表示

○　PC クラスタ

- サーバにアクセス
- 計算プログラムを受け取る
- データを受け取り，計算した結果を返す
- 終了まで繰り返す

サーバにおいて各 PC クラスタの計算結果を受け取る際に，各 PC の処理能力や通信のタイミングなどの影響で，割り当てた PC の順序ごとには最新の結果が得られないため，計算の済んだ PC から順に値を受け取るようにした．現在はデータの受け渡しのタイミングにおいて余裕を持たせるように処理に間隔を空けるための時間制御のパラメータを設定している．そのため，計算時間として余分にかかってしまうがデータへの無駄なアクセスを減らすことが期待できるはずである．このパラメータは環境によって，最適に設定することで，計算時間を短縮することが可能であるが，この設定は大変難しい問題であり，今回は適当な値を与えている．

4.　計算実験

11 号館 6 階第 1 実習室において，4 台の PC を利用し計算プログラムの実験を行った様子を図 2 に示す．図 2 は 4 台の PC において研究室にインストールされたサーバへアクセスし，クライアント登録を行って実験を行った様子である．左の PC では計算の管理画面を開き，管理 PC としても登録された状態である．拡大したものを図 3 に示す．

図 2　実験の試験段階の様子

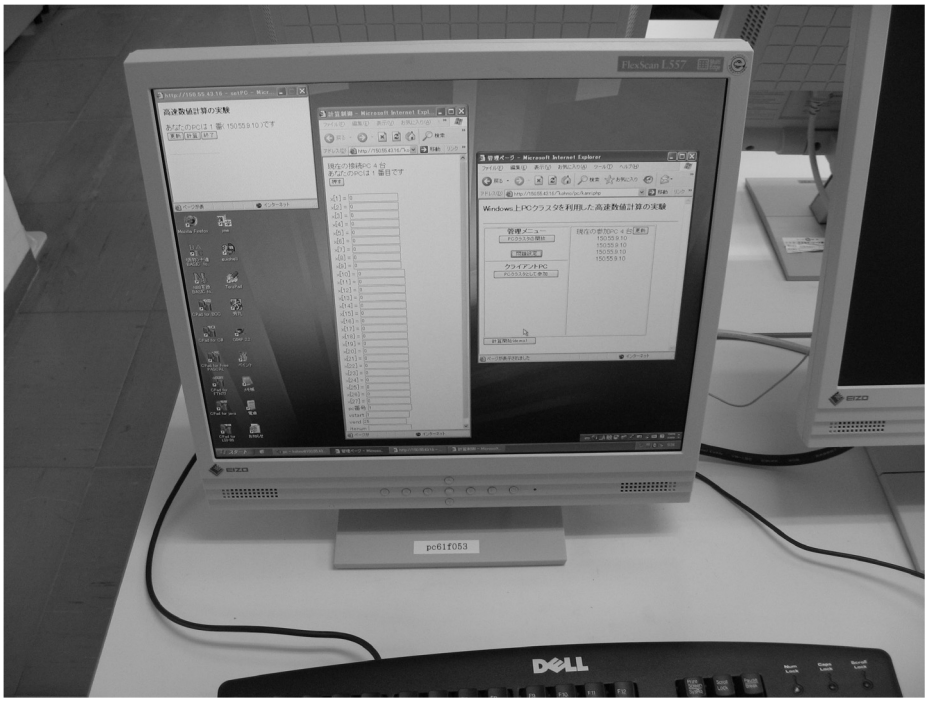

図 3　クライアントにおける PC クラスタ画面と管理画面

図 3 において中央に表示した縦長のウインドウは計算過程をチェックするために計算で得られた近似値を表示している様子である．左側に PC クラスタのプログラムが開いており，右側に管

理メニューを表示している．管理メニューでは，PC クラスタの開放，問題の設定，計算開始のボタンなどを用意している．このように，PC クラスタと管理プログラムを 1 台の CPU で起動することも可能である．

実験としては小さな問題しか扱っていないが，所属する情報科学科の実験室を用いて表 1 の環境の下，表 2 の結果を得た．

表 1　計算環境

実験問題	次数　100 3 重対角行列 $p=-0.4, q=0, r=-0.3, s=0$ 収束判定条件　$\left\|x^{*}-x^{(k)}\right\|_2 \le 0.0001$
計算 PC	Celeron 1.7GHz　Windows XP

表 2　計算結果

CPU 数	計算時間(秒)	備考
1	0.2124	サーバ上で PHP によりかかれたプログラムで計算
4	0.1848	PC クラスタ 4 台，管理 1 台をサーバにアクセス
5	0.1671	PC クラスタ 5 台，管理 1 台をサーバにアクセス

サーバ機には同スペックの PC に Apache_2.0.53　win32 版をインストールし，設定を行った．表 2 において，CPU 数 1 台の場合，PHP プログラムのみで実行されており，PC クラスタを利用した結果ではない．したがって，比較としては適切ではないが参考までに示した．次に，CPU 数を 4 から 5 に増やした場合，クライアントでは JavaScript で計算がされ，サーバ側で PHP において処理されるという同じ環境である．4 台の場合は 1 台あたり 25 行，5 台の場合は 1 台あたり 20 行の計算を各クライアントが計算している．この場合，計算時間が減少した結果が得られた．しかしながら，同様の実験を繰り返し行った場合に，計算時間は変化することがある．おそらくはネットワーク通信上の遅延が原因であると思われるが，より詳細な実験が必要であると考える．CPU の台数を増やした場合，通信エラーが起こすことがあり，プログラムの修正が必要であることがわかった．

5.　考察

今回の実験において，プログラムの試作段階まで行うことができた．プログラムは PHP と JavaScript で作成し，現在のところ Internet Explorer 6 上での動作確認をおこなった．PHP はサーバ側で動作し，JavaScript はクライアント側のブラウザ上で動作する．OS に依存する部分が少ない環境での分散計算環境が構築できた．さらに，最近ではガジェットの利用も増えている

ので，今後，ガジェットとしての開発も考えられる．より多くの PC に起動と同時にアプリケーションを登録しておくことで, サーバからの計算命令で各 PC の余剰している CPU パワーを利用して計算を行うことが可能であると考える．

OS に依存することなく，分散処理を行うプロジェクトとしては，BOINC が知られている[3]. BOINC とはボランティア・コンピューティングとデスクトップ・グリッド・コンピューティングのためのオープンソース・ソフトウェアである．データとしては膨大な量を扱うために，1 台や数台の PC でその PC 間の通信のロスが計算のロスにつながるというものでは無く，インターネット上に存在する膨大な数の PC，もちろん Windows，Mac などの OS に依存しない PC を利用して計算を行い，何百年もかかってしまう計算を短期間で行うものである．このような分散環境で計算を行うためには，解くべき問題を適切に処理できる形に最適化を行う必要がある．

今後の課題として，より複雑な問題を領域分割など行い簡略化しその部分ごとの計算を各クラスタが行うような計算環境を構築することが必要であるといえる．また，各計算を高速化するために，前処理の適用を今回は出来なかったが，前処理の導入と，さらにはこのような分散処理に合わせた前処理を導出することが今後の課題といえる．プログラムにおいては，データの更新においてデータ量を最小限に抑えることや，計算のタイミングの調整など多くの課題が残されている．

参考文献

[1] 小国力訳，J.J.Dongarra, I.S.Duff, D.C.Sorensen, H.A. van der Vorst，コンピュータによる連立一次方程式の解法，丸善株式会社,1993.

[2] 河野敏行，適応型前処理付反復法について，日本応用数理学会論文誌 15,3(2005)235-243.

[3] BOINC, http://boinc.berkeley.edu/

大規模問題高速求解のためのキャンパスグリッド環境構築

上嶋　明
岡山理科大学工学部情報工学科
E-mail: uejima@ice.ous.ac.jp

１．はじめに

大規模計算を行う方法として，ネットワークに接続された多数のコンピュータをまとめて仮想的な 1 台の高性能システムとして使用するグリッド・コンピューティングが広まっている。大学内には授業や研究のために使用する多数のパソコン（PC）がある。例えば，岡山理科大学情報処理センターには約 500 台[1]，工学部情報工学科の学生実験室と実習室には合計約 140 台の PC があり，さらに各研究室にも多くの PC がある。しかし，夜間や休日にはこれらの大部分は使用されていない。そこで，これらの遊休資源を利用してグリッド・コンピューティングを行う環境を構築する方法を検討する。各 PC にはその用途に応じて Linux や Windows などの OS がインストールされているが，これらの既存環境に一切変更を加えなくてもよいよう，ローカルの HDD を使用せずにネットワークから Linux をブートし，グリッド・コンピューティングのノードとして運用する方法を用いる。

２．グリッド・コンピューティング

２．１　概要

電力はどこでどのように発電されたのかを特に意識しなくとも，コンセントに電気機器を接続すれば必要な量を利用できる。グリッド・コンピューティングとは，必要なときに必要なだけ電力と同様にコンピュータ資源を利用できるような環境を目指したものである。名前は電力網を意味する「パワーグリッド」に由来する。ネットワークに接続された多数のコンピュータを接続して仮想的な 1 台のシステムとして使用し，ユーザは必要なとき必要なだけこのシステムから演算能力や記憶領域を取り出して利用する。グリッド・コンピューティングの分類として，演算処理を行う資源を仮想化し，共有するプロセッシング・グリッドや，膨大なデータを分散して保存し，共有するデータ・グリッドなどがある[2]。

２．２　Globus ツールキット

複数の異なった種類のコンピュータをネットワークで接続してグリッドを構成するためは，ライブラリやツール群からなるミドルウェアが必要となる。本研究では，Globus ツールキット 4.0 を用いる。Globus ツールキットは Globus Alliance[3]により開発されたミドルウェアであり，現在では事実上の標準として広く用いられている。Globus ツールキットは，グリッド・コンピューティングで必要となる認証，リソース管理，データ管理，データ転送などの機能を提供する。

２.３　MPICH-G2

MPI（Message Passing Interface）は分散メモリ型並列計算機で複数のプロセス間でメッセージ通信を行うための規格であり，MPI を用いて記述した並列プログラムは MPI に対応した様々な並列計算機上で実行可能であるという特長がある。MPI 1.1 の実装の一つである MPICH[4]において，通信デバイスとして Globus を使用するように設定したものを MPICH-G2 と呼んでいる。MPICH-G2 の導入により，MPI を用いて記述された並列プログラムをグリッド環境で実行することが可能となる。

３．ディスクレスによる Linux の運用

３.１　ブート方法

ローカルの HDD を使用せずに OS を起動する方法として，CD を用いる方法やネットワークを用いる方法などがある。前者は，1 枚の CD-ROM で OS を起動して使用できるようにした，いわゆるライブ CD と呼ばれるものを用いる。代表的なものに，CD-ROM（または DVD-ROM）1 枚で Linux を起動できる KNOPPIX がある。しかし，この方法では情報工学科の実習室や実験室の学生用 PC のように，CD ドライブが装備されていない PC を用いることができない。また，たとえ研究室の PC のように CD ドライブが装備されている場合でも，全 PC に CD-ROM を挿入して起動させる必要があり，多くの台数の PC を使用する際には大変な手間となる。そのため，本研究ではネットワークから OS をブートさせるネットワークブートと呼ばれる方法[1][5]を用いる。

３.２　ネットワークブート

各クライアント PC をネットワークから起動させるため，IP アドレスの発行，ブートローダやカーネルの転送，NFS によるルートディレクトリ公開などを担当するブートサーバを 1 台用意する。サーバは負荷を考慮し，NIC（Network Interface Card）には性能が高く高性能計算の分野で実績のある Intel Pro/1000 を，HDD には高転送速度／低シークタイムのエンタープライズ用のものを使用している。

各クライアント PC では，Wake On LAN によりリモートから電源 ON 後，PXE BIOS が動作，ブートサーバの DHCP により IP アドレスと起動ファイルの情報を取得して TFTP によりブートローダ PXELINUX を受信，OS(Fedora Core Linux)カーネルを受信，ルートディレクトリを NFS マウント，という順序でブートが行われる[1][5][6]。

表 1：ブートサーバの構成

CPU	Pentium 4 3.6GHz
メモリ	2GB(DDR2)
HDD	150GB(10,000RPM, SATA)
NIC	Intel Pro/1000(1000BASE-T)
OS	Fedora Core 5(カーネル 2.6.18)

３．３　NFS ルート動作

クライアント PC は，ルートディレクトリ以下のすべてについて，ブートサーバ上に置かれたファイルを NFS でマウントして使用する NFS ルート動作を行うため，ローカルの HDD を全く使用しない。この方法では，/bin や/usr/bin のコマンド群，/lib や/usr/lib のライブラリ群などが格納されているディレクトリを全クライアント PC で共有できるが，/etc ディレクトリ内にある各種設定ファイルや/var ディレクトリのスプール・ログファイル領域，/tmp の一時ファイル領域も共有してしまうことになる。特に Globus ツールキットにおいては，使用する各ノードにおいて固有のホスト証明書を/etc/grid-security 内に配置する必要があるために不都合が生じる。

そこで，本研究では/etc，/var，/tmp の三つのディレクトリについては，クライアント PC 毎に個別に保持する方法を用いた（図 1）。これにより，ディスクレスでありながら通常の PC と同じ動作が可能となる。初回ブート時にはクライアント共通用の三つのディレクトリの内容をブートサーバ中の各クライアント専用領域にコピーし，以降はこれらを各クライアント PC の/etc，/var，/tmp に NFS でマウントして使用する。これらのディレクトリ内に変更が加えられた場合には，ブートサーバ中のそのクライアント専用領域内のファイルが更新されることになる。この実現のため，クライアントの/etc/rc.d ディレクトリ内の rc.sysinit や set-hostname-network.sh などの OS ブートスクリプト複数個に変更を加えるとともに，専用のスクリプト rc.dlclient を新規に記述した。

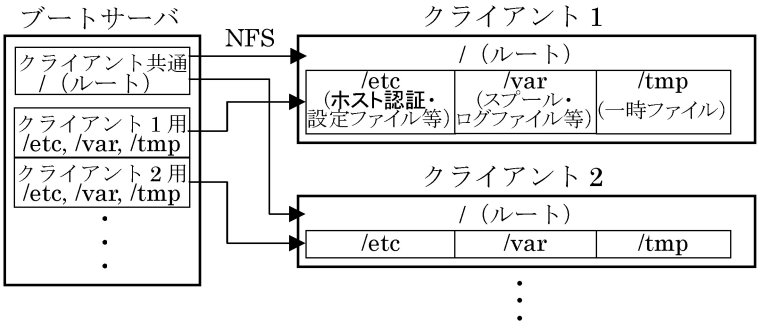

図 1：クライアントのディレクトリ構成

表 2：クライアントの構成

CPU	Pentium 4 3.0GHz, Pentium 4 3.2GHz, Pentium D 3.2GHz
メモリ	1GB
HDD	120～300GB（非使用）
NIC	Intel Pro/100(100BASE-T), Realtek RTL8169/8110(1000BASE-T), Broadcom BCM5788/5789(1000BASE-T)
OS	Windows XP（非使用）

４．実験と考察

今回は実習室や情報処理センターなどの多くの台数の PC を用いる前段階として，研究室

内の複数台の PC をクライアントとして用いて実験を行った[6]。各 PC の構成は表 2 に示すように CPU，HDD 容量，NIC など各々異なるが，HDD にはすべて Windows XP がインストールされている。

実験の結果，ネットワークより各 PC で Linux を起動させ，ホスト証明書の認証の後，GRAM（ジョブ実行），GridFTP（データ転送）などの動作確認を行い，いずれも正常に動作することを確認した。また, MPICH-G2 による MPI テストプログラムの動作を確認した。

クライアント専用領域に必要な容量を調べたところ，ログファイルなどのない初期状態で 1 台あたり約 140MB であった。クライアント PC 台数分の容量をブートサーバの HDD 上に専有するという問題があるが，クライアント PC 台数を増加させる場合にはこの容量よりも HDD やネットワークの転送速度の方が先にボトルネックとなる可能性が高いと考えられる。実際には/etc ディレクトリ内でクライアント専用にする必要がある部分は小容量であるため，必要な領域を削減できる可能性はある。

実際に大規模計算を行う際にディスクレスのノードを用いると，各クライアント PC ではページファイルを作成できないため，実行可能なプログラムのサイズが物理メモリ容量に依存する。これは，膨大なデータを扱う際に問題となる可能性がある。

５．おわりに

ネットワークブート PC により，ローカルの HDD に変更を加えることなくグリッド・コンピューティング環境を構築できることを確認した。今後は 1 台のブートサーバで対応できるクライアント PC 台数の上限の検証を進め，実習室の PC など多数のノードを用いた実験を行う予定である。また，実際に構築したグリッド・コンピューティング環境上で実行する大規模問題の選定を行っている。

謝辞

有益な情報を提供いただきました岡山理科大学情報処理センターの畠山唯達先生，長谷輝章氏，岡山理科大学総合情報学部の河野敏行先生に感謝致します。

参考文献

[1] 畠山唯達，長谷輝章，河野敏行：“既存の教育用計算機を用いたクラスタコンピューティング環境の構築： Ⅰ．起動に関する諸問題と解決”, 岡山理科大学情報処理センター研究報告，第 27 号，pp. 9-16(2006).

[2] 日本アイ・ビー・エム システムズ・エンジニアリング：“グリッド・コンピューティングとは何か”，ソフトバンクパブリッシング(2004).

[3] The Globus Alliance: http://www.globus.org/

[4] MPICH - A Portable Implementation of MPI: http://www-unix.mcs.anl.gov/mpi/mpich1/

[5] Fedora Core Linux ディスクレス化キット： http://wikiwiki.jp/disklessfun/?disklessfc

[6] 沼 美月，上嶋 明：“ネットワークブート PC によるグリッド環境構築の検討”，電気・情報関連学会中国支部第 58 回連合大会論文集，p. 315(2007).

あるハミルトン系に対する数値解法の数値実験による比較

古松和治（岡山理科大学大学院　総合情報研究科　情報科学専攻　Ｍ１）
榊原道夫（岡山理科大学大学院　総合情報研究科　情報科学専攻）

概要

ハミルトン系の問題に対する数値解法として近年、種々なシンプレクティック法が提案されている．シンプレクティック法は保存系の性質を残した離散化として注目される点があるが、実際のシミュレーションにおいて従来の非シンプレクティック法と比較して実用的であるかどうかの研究は十分存在しない．そこでわれわれは、単振動の方程式を用いていくつかのシンプレクティックおよび非シンプレクティック数値解法を適用し数値誤差をエネルギーの保存状況、フェーズのずれの観点より比較した．本レポートで，それらをまとめた結果を示す．数値解法としてはシンプレクティック・オイラー法，シュテルマー・ベルレの方法，Collatz と Nyström の４次と５次のルンゲ・クッタ法を用いる．

１．はじめに

ハミルトン系は，R^{2d} 値関数 $x(t)=[q(t)^T, p(t)^T]^T$ $(q(t), p(t) \in R^d)$ を未知変数とする方程式を考えるとき，R^{2d} 上で定義された C^2 級関数 $H(q,p)$ を用いて，

$$\frac{dq_i}{dt}=\frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt}=-\frac{\partial H}{\partial q_i} \quad (i=1,2,...,d)$$

と表される．また，関数 $H(q,p)$ をハミルトニアンといい，物理学におけるエネルギーを表す関数である[h1]．今回の研究では，このハミルトン系の問題に対してシンプレクティック・オイラー法，シュテルマー・ベルレの方法，Collatz と Nyström の４次と５次のルンゲ・クッタ法を適用し，数値解やフェーズのずれ，エネルギーの変化について調べる． 扱う例題として，関数

$$H(q,p)=\frac{1}{2}(p^2+\pi^2 q^2)$$

をハミルトニアンとするハミルトン系である単振動の方程式

$$\frac{dq}{dt}=p \quad , \quad \frac{dp}{dt}=-\pi^2 q$$

を用いる．初期値は $q_0=0$， $p_0=\pi$ とし，ステップ幅は $h=0.05$ とし，数値解を計算し比較する．

２．計算用いた解法について

シンプレクティック・オイラー法とシュテルマー・ベルレの方法

シンプレクティック・オイラー法とシュテルマー・ベルレの方法の式は,

シンプレクティック・オイラー法

$$\begin{cases} q_{n+1} = q_n + hp_n \\ p_{n+1} = p_n - h\pi^2 q_{n+1} \end{cases}$$

シュテルマー・ベルレの方法

$$\begin{cases} p_{n+\frac{1}{2}} = p_n - \dfrac{h}{2}\pi^2 q_n \\ q_{n+1} = q_n + hp_{n+\frac{1}{2}} \\ p_{n+1} = p_{n+\frac{1}{2}} - \dfrac{h}{2}\pi^2 q_{n+1} \end{cases}$$

となる[4].

Nyström と Collatz のルンゲ・クッタ法

Nyström と Collatz のルンゲ・クッタ法は２階微分方程式の初期値問題：

$$\frac{d^2y}{dt^2} = f(y), \qquad y(0) = \alpha, y'(0) = \beta$$

に対するルンゲ・クッタ法として用いられる．Nyström と Collatz のルンゲ・クッタ法の一般式は,

Nyström

$$k_i = f(x_n + c_i h, y_n + c_i hy'_n + h^2 \sum_j \bar{a}_{ij} k_j)$$

$$y_{n+1} = y_n + hy'_n + h^2 \sum_i \bar{b}_i k_i$$

$$y'_{n+1} = y'_n + h\sum_i b_i k_i$$

Collatz

$$k_i = h^2 f\left(x_n + c_i h, y_n + c_i hy'_n + \sum_j a_{ij} k_j \right)$$

$$y_{n+1} = y_n + hy'_n + \sum_i b_i k_i$$

$$hy'_{n+1} = hy'_n + \sum_i b'_i k_i$$

となる．また，４次と５次の Nyström と Collatz のルンゲ・クッタ法をテーブル表記する

と表１，表２，表３，表４のようになる[1][2][3].

表１．Nyström, order 4

c_i	$\bar{a}_{ij}$		
0			
$\frac{1}{2}$	$\frac{1}{8}$		
1	0	$\frac{1}{2}$	
$\bar{b}_i$	$\frac{1}{6}$	$\frac{1}{3}$	0
b_i	$\frac{1}{6}$	$\frac{4}{6}$	$\frac{1}{6}$

表２．Collatz, order 4

c_i	a_{ij}		
0			
$\frac{1}{2}$	$\frac{1}{8}$		
1	0	$\frac{1}{2}$	
b_i'	$\frac{1}{6}$	$\frac{2}{6}$	
b_i	$\frac{1}{6}$	$\frac{4}{6}$	$\frac{1}{6}$

表３．Nyström, order 5

c_i	$\bar{a}_{ij}$			
0				
$\frac{1}{5}$	$\frac{1}{50}$			
$\frac{2}{3}$	$-\frac{1}{27}$	$\frac{7}{27}$		
1	$\frac{3}{10}$	$-\frac{2}{35}$	$\frac{9}{35}$	
$\bar{b}_i$	$\frac{14}{336}$	$\frac{100}{336}$	$\frac{54}{336}$	0
b_i	$\frac{14}{336}$	$\frac{125}{336}$	$\frac{162}{336}$	$\frac{35}{336}$

表４．Collatz, order 5

c_i	a_{ij}			
0	0			
$\frac{1}{4}$	$\frac{1}{32}$			
$\frac{7}{10}$	$-\frac{7}{1000}$	$\frac{63}{250}$		
1	$\frac{2}{7}$	0	$\frac{3}{14}$	
b_i'	$\frac{1}{14}$	$\frac{8}{27}$	$\frac{25}{189}$	
b_i	$\frac{1}{14}$	$\frac{32}{81}$	$\frac{250}{567}$	$\frac{5}{54}$

３．数値解の比較

シンプレクティック・オイラー法とシュテルマー・ベルレの方法

数値解法を例題に適用し，数値解を比較した．双方のある地点での数値解を表５に示した．また$t=485$付近のグラフを図１，図２，$t=485$付近のグラフを図３，図４に示した．$t=485$のときにフェーズが４分の１ずれ，$t=970$のときに完全に反転していることが表５とグラフからわかる．

表５．シンプレクティック・オイラー法とシュテルマー・ベルレの方法の
t = 0, 485, 970 での数値解と解析解との比較

t	解析解	シンプレクティック･オイラー法	シュテルマー･ベルレの方法
0	0	0	0
485	0	1.003098593540	1.003098593510
970	0	0.000061932703	0.000061932786

$t = n * h$

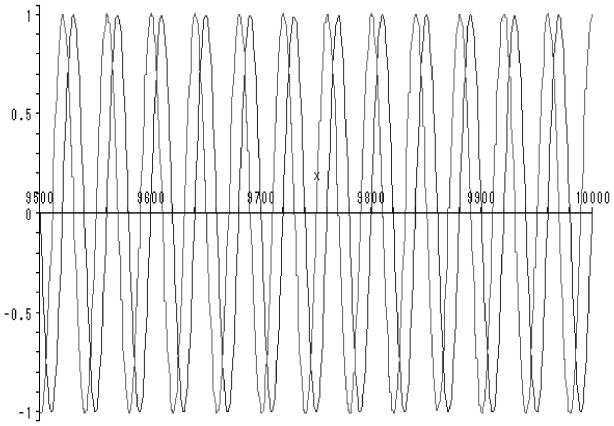

図１．シンプレクティック・オイラー法
(t=475～500)

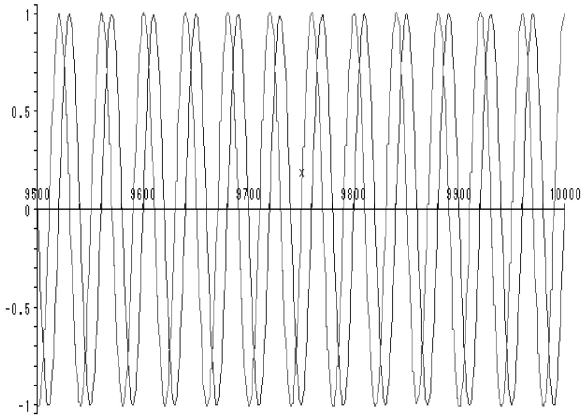

図２．シュテルマー・ベルレの方法
(t=475～500)

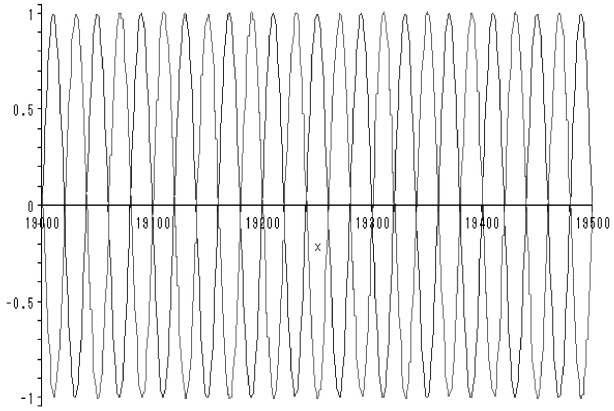

図３．シンプレクティック・オイラー法
(t=950～975)

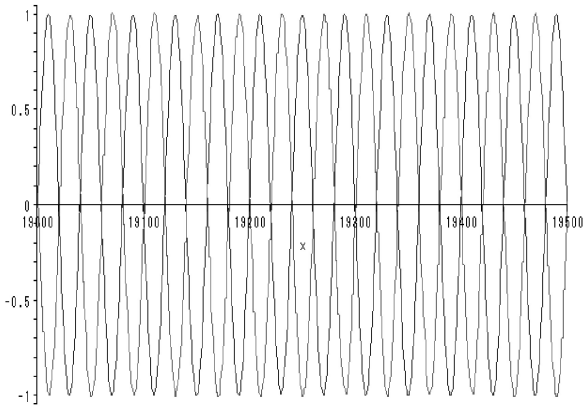

図４．シュテルマー・ベルレの方法
(t=950～975)

シンプレクティック・オイラー法とシュテルマー・ベルレの方法のハミルトニアンの比較

シンプレクティック・オイラー法とシュテルマー・ベルレの方法のハミルトニアンの値について調べ，$t = 1000$ までの値を表６に示した．また，そのグラフを図５，図６に示した．シンプレクティック・オイラー法とシュテルマー・ベルレの方法のどちらもハミルトニアンの値のグラフは振動しており，いくつかの値の平均をとると厳密解に近づくことが

わかった．また，ハミルトニアンの値の振動はシュテルマー・ベルレの方法の方が緩やかであった．

表６．シンプレクティック・オイラー法とシュテルマー・ベルレの方法のハミルトニアンの数値解

t	シンプレクティック・オイラー法	シュテルマー・ベルレの方法
0	4.93480220054	4.93480220054
250	4.57859977088	4.95086103797
500	5.03365616377	4.96535915062
750	5.34563256356	4.94789126955
1000	4.86025580064	4.93509082588
平均	4.95058929988	4.94680089691

$t = n * h$

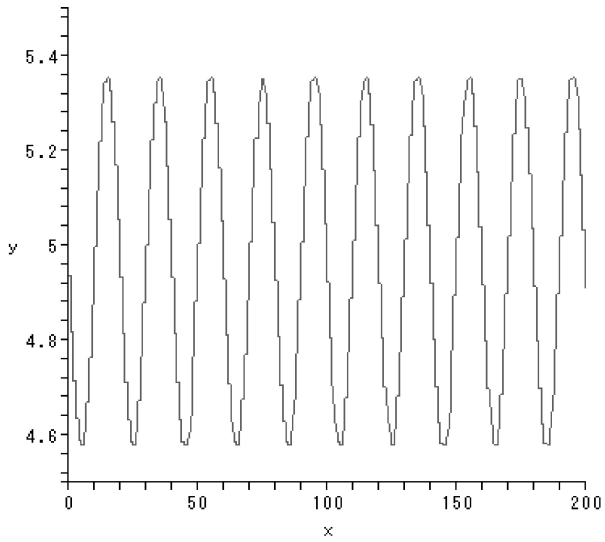

図５．シンプレクティック・オイラー法

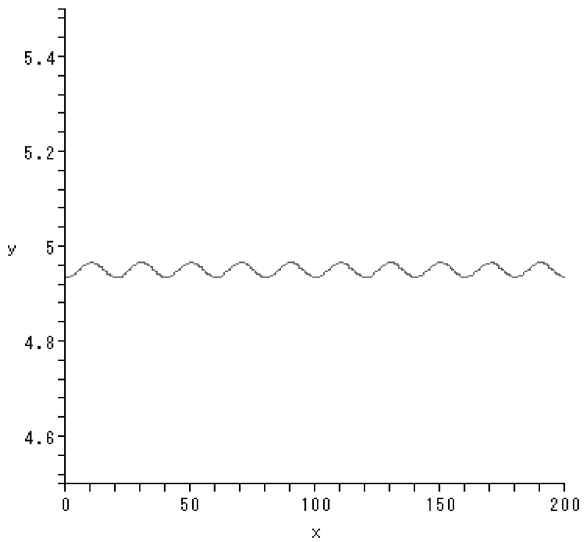

図６．シュテルマー・ベルレの方法

Nyström と Collatz のルンゲ・クッタ法

Nyström と Collatz の４次のルンゲ・クッタ法を例題に適用し，数値解を比較した．その数値解を表７に示した．

表７．Nyström, order 4 & Collatz, order 4 の数値解

t(n*h)	Nystrom-4	Collatz-4
0	0	0
250	0.00148964492113	0.00148964500324
500	0.00297889807939	0.00297889816764
750	0.00446775621240	0.00446775631480
1000	0.00595621610813	0.00595621638473
2000	0.01190600929730	0.01190600929710
3000	0.01784917320200	0.01784917303500
4000	0.02378550232200	0.02378550245540
5000	0.02971479119780	0.02971479126550
6000	0.03563683535090	0.03563683496840
7000	0.04155143013890	0.04155142929500
8000	0.04745837185870	0.04745837147940
9000	0.05335745749010	0.05335745690690
10000	0.05924848414760	0.05924848366730

式には違いがあるがかかる重みは変わらず，それぞれの数値解はほとんど変わらない値を示した．

Nyström と Collatz の４次のルンゲ・クッタ法のハミルトニアンの比較

Nyström と Collatz の４次のルンゲ・クッタ法のハミルトニアンについて比較した.

表８．Nyström,order 4 の
ハミルトニアンの数値解

t	数値解
0	4.93480220054
1000	4.92965702136
2000	4.92451720874
3000	4.91938275143
4000	4.91425364799
5000	4.90912989206
6000	4.90401147346
7000	4.89889839056
8000	4.89379063488
9000	4.88868820312
10000	4.88359108937

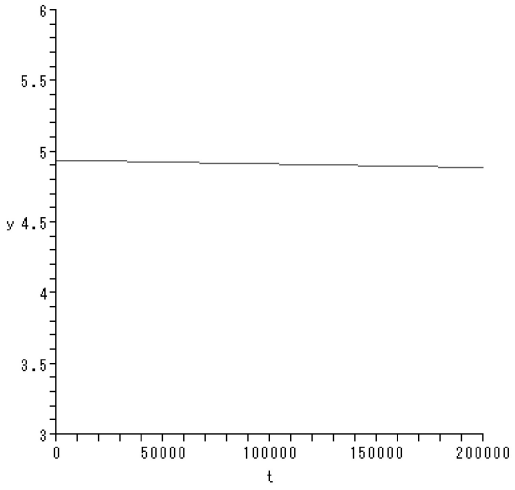

図９．Nyström,order 4 のハミルトニアンのグラフ

表９．Collatz,order 4 のハミルトニアンの数値解

t	数値解
0	4.93480220054
1000	4.92965702141
2000	4.92451721006
3000	4.91938275169
4000	4.91425364789
5000	4.90912988826
6000	4.90401147451
7000	4.89889839189
8000	4.89379064127
9000	4.88868821309
10000	4.88359109761

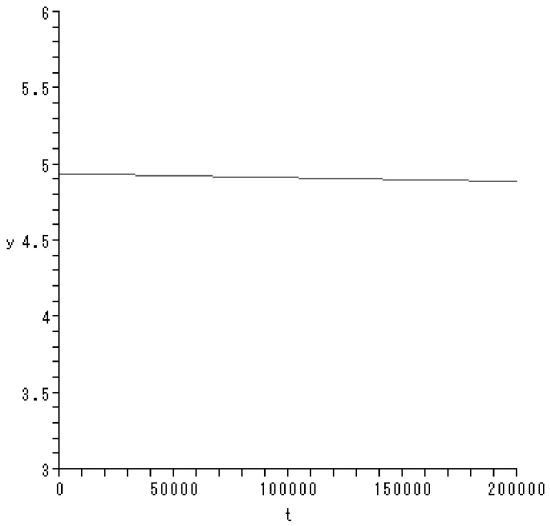

図１０．Collatz,order 4 のハミルトニアンのグラフ

ハミルトニアンの値も数値解のときと同様にほとんど差のない値が得られた．厳密解では一定であるハミルトニアンだが，Nyström の４次と Collatz の４次の両方とも $t=10000$ まで計算すると，図９，図１０のようにグラフで少々の傾きが目に見えた.

Nyström と Collatz の５次のルンゲ・クッタ法の数値解の比較

Nyström と Collatz の５次のルンゲ・クッタ法を例題に適用し，数値解を比較した．その数値解を表１０に示した．

表１０．Nyström, order 5 & Collatz, order 5 の誤差

t(n*h)	Nystrom-5	Collatz-5
0	0	0
250	0.00000175457764	0.000001447822009
500	0.00000350916316	0.000002895455977
750	0.00000526391211	0.000004343326990
1000	0.00000701877693	0.000005791121751
2000	0.00001403892271	0.000011583189676
3000	0.00002106007847	0.000017376018641
4000	0.00002808266773	0.000023169585297
5000	0.00003510644782	0.000028964086636
6000	0.00004213111383	0.000034759721379
7000	0.00004915700999	0.000040555999330
8000	0.00005618419941	0.000046352733912
9000	0.00006321254689	0.000052150649420
10000	0.00007024196392	0.000057949040370

４次の式とは違いそれぞれにかかる重みに違いがある．数値解は４次の時ほど両方の値が似たような値を得られたわけではないが，どちらも４次の時の式よりも精度のよい値を得ることができた．また，この問題では Collatz の５次の方が精度はよい．

Nyström と Collatz の５次のルンゲ・クッタ法のハミルトニアンの比較

４次の時と同様に５次の時のハミルトニアンについて比較した.

表１１．Nyström,order 5 のハミルトニアンの数値解

t	数値解
0	4.93480220054
1000	4.93562338525
2000	4.93644470723
3000	4.93726616418
4000	4.93808775837
5000	4.93890948690
6000	4.93973135518
7000	4.94055335935
8000	4.94137550046
9000	4.94219777746
10000	4.94302019491

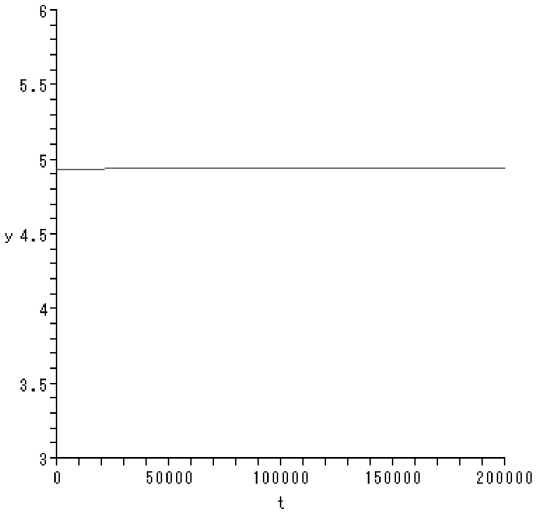

図１３．Nyström,order 5 のハミルトニアンのグラフ

表１２．Collatz,order 5 のハミルトニアンの数値解

t	数値解
0	4.93480220054
1000	4.93552105801
2000	4.93624000424
3000	4.93695906188
4000	4.93767821231
5000	4.93839747585
6000	4.93911685088
7000	4.93983632390
8000	4.94055589396
9000	4.94127559026
10000	4.94199537559

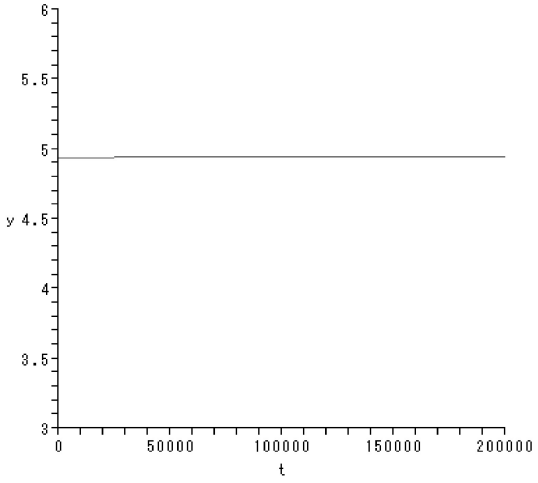

図１４．Collatz,order 5 のハミルトニアンのグラフ

Nyström と Collatz の５次のときのハミルトニアンの値は，一定である厳密解のハミルトニアンに対し，数値解の誤差により４次のときは減少したが，５次の場合は，増加していることがわかった．しかし５次の場合，その誤差も僅かであり双方ともかなり精度のよい値を得ることができた.

４．数値解の考察

単振動問題に対し，数値例を計算しその振る舞いについて比較検討した．シンプレクティック法は，ハミルトニアンの値は有界な振動で変動し，誤差の振幅は時間ステップを小さくするほど小さくなる．しかし，時間が進むにつれ，位相のずれが大きくなる．シンプレクティック・オイラー法とシュテルマー・ベルレの方法はそれぞれ１次および２次精度であるため，ハミルトニアンの誤差の増大の変わりに誤差の蓄積が位相のズレに出ていると考えられる．比較的長い時間の計算においてもそれらの数値解は厳密解とそれほど違わず，ハミルトニアンおよび位相ともに誤差の成長が緩慢である．ハミルトン系は物理的な背景より導出されているため，位相のズレはシミュレーションの結果に深刻な誤解を生む場合が考えられる．一方，ここで取り上げたルンゲ・クッタ法は２階の微分方程式に対するもので，非シンプレクティックであるが高次アルゴリズムが提案されている．数値結果よりシンプレクティックでないとしても，誤差の少ないアルゴリズムの方がよい結果であるといえる．

参考文献

[1]L. Collatz, The Numerical Treatment of Differential Equations, Berlin : Springer Verlag, p. 61, 1960.

[2]J.R. Dormand, Numerical Methods for Differential Equations A Computational Approach, 1996.

[3]R. E. Scraton, The numerical solution of second-order differential equations not containing the first derivative explicitly, The Conputer J., vol. 6, pp. 368-370, 1963-1964.

[4]三井 斌友, 小藤 俊幸, 齊藤 善弘, 微分方程式による計算科学入門, 共立出版株式会社, 2004 年

[h1] http://ja.wikipedia.org/wiki/

Estimation of Difficulty of Multidimensional Knapsack Problems with Demand Constraints

Hirokazu OHTAGAKI

Okayama University of Science

1 Introduction

The development of computational complexity theory led to a fascinating insight into the inherent difficulty of computational optimization problems. Optimization problem of knapsack type involves various applications in engineering, science and economics, it is significant to solve various kinds of knapsack problems. Recently, an improved surrogate constraints method is proposed to solve multidimensional nonlinear knapsack problems of large size, witch have surrogate gaps. However the knapsack problems are of NP-hard. Furthermore, the knapsack problems with demand constraints are frequently quite difficult to obtain high quality solution because of loss of monotonicity of constraints. To obtain even feasible solution is extremely difficult. Application of the algorithms to such a problem directly is not effective.

In this report, by using entropy, a method for evaluating difficulty of multidimensional 0-1 knapsack problems with demand constraints is proposed. It is shown in chapter 5 that solving the problems with demand constraint is rather difficult than solving that with ordinary problems with no demand constraint. Computational experiments show that the proposed method is effective to evaluate the computational difficulty of the problems.

2 Formulation of Problems

Multidimensional Knapsack Problems with Demand Constraints can be formulated as follows;

$$[P] \quad \text{maximize} \quad f(\boldsymbol{x}) = \sum_{n=1}^{N} a_n x_n \tag{1.a}$$

$$\text{subject to} \quad g_m(\boldsymbol{x}) = \sum_{n=1}^{N} c_{mn} x_n \le b_m, \tag{1.b}$$

$$m = 1, 2, \ldots, m1,$$

$$g_m(\boldsymbol{x}) = \sum_{n=1}^{N} c_{mn} x_n \le b_m,$$

$$m = m1+1, m1+2, \ldots, M \tag{1.b}$$

$$\boldsymbol{x}(= (x_1, x_2, \ldots, x_N)^T) \in \mathbf{X} = \Pi_{n=1}^{N}\{0,1\} \subseteq \mathbf{R}^N, \tag{1.c}$$

where, $\boldsymbol{x}$ denotes an N dimensional 0-1 valued decision variable, $f(\boldsymbol{x})$ denotes N dimensional vector-valued objective function $(g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \ldots, g_{m1}(\boldsymbol{x}))^T$ and $(g_{m1+1}(\boldsymbol{x}), g_{m1+2}(\boldsymbol{x}), \ldots, g_M(\boldsymbol{x}))^T$ denote $m1$ and $M - m1$ dimensional vector-valued constraint functions, respectively. The vector $\boldsymbol{b} = (b_1, b_2, \ldots, b_{m1}, b_{m1+1}, \ldots, b_M)^T$ denotes available amounts of resources. The constraint (1.b) is of ordinary type, and the constraint (1.c) is of demand type.

By using a surrogate multiplier $\boldsymbol{u} = (u_1, u_2, \ldots, u_M)^T$, the problem $[P]$ is translated into the surrogate demand constraints problem;

$$[P^{\mathrm{S}}(\boldsymbol{u})] \quad \text{maximize} \quad f(\boldsymbol{x}) \tag{2.a}$$

$$\text{subject to} \quad \boldsymbol{u}^T \boldsymbol{h}(\boldsymbol{x}) \le 0, \tag{2.b}$$

$$\boldsymbol{x} \in \mathbf{X}, \tag{2.c}$$

where,

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{b}, \tag{2.d}$$

$$\boldsymbol{u} \in \mathbf{U} = \{\boldsymbol{u} \in \mathbf{R}^M | \sum_{m=1}^{M} u_m \le 1, \quad \boldsymbol{u} \ge 0\} \tag{2.e}$$

Inequality $(2.b)$ is called a surrogate demand constraints equation.

A surrogate dual $[P^{\mathrm{SD}}]$ to the original problem $[P]$ is written as follows;

$$[P^{\mathrm{SD}}] \quad \min\{\mathrm{opt}[P^{\mathrm{S}}(\boldsymbol{u})] : \quad \boldsymbol{u} \in \mathbf{U}\} \tag{3}$$

where, opt[$P^{\mathrm{S}}(\boldsymbol{u})$] denotes an optimal value of the objective function of the problem P^{S}.

3 Estimation of Difficulty of Problems

Let p_n be a probability such that

$$f^{\mathrm{UB}}(\boldsymbol{x})|_{x_n=0} \geq f^{\mathrm{UB}}(\boldsymbol{x})|_{x_n=1}$$

or

$$f^{\mathrm{UB}}(\boldsymbol{x})|_{x_n=1} < f^{\mathrm{UB}}(\boldsymbol{x})|_{x_n=0}$$

To estimate the probality p_n, the normalized difference of the upper bound;

$$\frac{d_n}{f^{\mathrm{REAL}}(\boldsymbol{x}) - f^{\mathrm{NEAR}}(\boldsymbol{x})}$$

$$d_n = |f^{\mathrm{UB}}(\boldsymbol{x})|_{x_n=0} - f^{\mathrm{UB}}(\boldsymbol{x})|_{x_n=1}|$$

The following entropy is used to estimate the difficulty of the problem;

$$H = \sum_{n=1}^{N} h_n$$

$$h_n = -p_n \log_2(p_n) - (1-p_n)\log_2(1-p_n)$$

4 Computational Experiments

The results obtained from computational experiments to 60 problems shown in the references Chu et.al.(12). The problems tested are generated by using random number generator such that the objective functions and constraint functions are correlated to mutually.

From the obtained results, the probability p_n in the problems with demand constraints may be estimated approximately as

$$2^{-\alpha d_n - \beta},$$

where the values of parameters α and β are -12.2 and 1.0, respectively.

In order to study effectiveness of the estimation method presented, the value of PGC(Percent Gap Closure) is introduced;

$$\frac{f^{\mathrm{REAL}} - f^{\mathrm{OPT}}}{f^{\mathrm{REAL}} - f^{\mathrm{OPT}}} \times 100$$

Table 1. Difference of upper bounds, probability p_n and its approximation

Difference	Problem				
δ_n	00∼09	10∼19	20∼29	00∼29	p_n
0.0∼	0.5469	0.5517	0.4098	0.5027	0.503
0.01∼	0.2800	0.2937	0.3085	0.2938	0.234
0.1∼	0.1161	0.1008	0.1221	0.1129	0.100
0.2∼	0.0421	0.0382	0.0490	0.0427	0.043
0.3∼	0.0265	0.0167	0.0187	0.0206	0.018
0.4∼	0.0074	0.0000	0.0000	0.0027	0.008
0.5∼	0.0000	0.0000	0.0000	0.0000	0.003
0.6∼	0.0000	0.0000	0.0000	0.0000	0.001
0.7∼	0.0000	0.0000	0.0000	0.0000	0.001
0.8∼	0.0000	0.0000	0.0000	0.0000	0
0.9∼	0.0000	0.0000	0.0000	0.0000	0
1.0∼	0.0000	0.0000	0.0000	0.0000	0

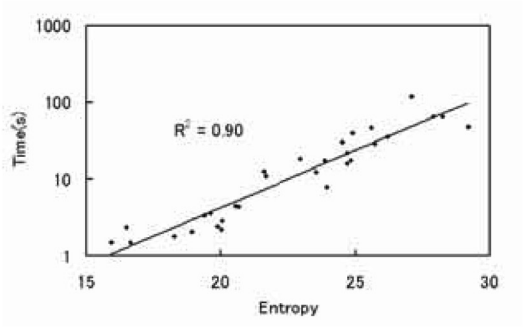

Fig.1. Entropy vs Computational Time ($M = m1 = 5, N = 250$)

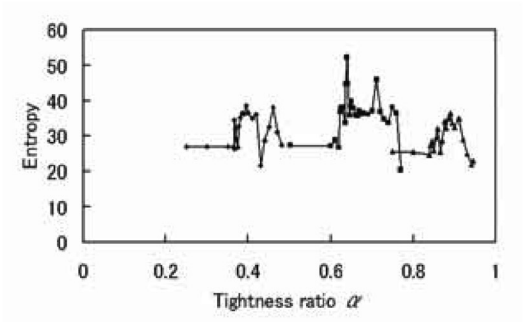

Fig.2. Entropy vs Tightness Ratio ($m1 = 4, M = 5, N = 500$: Single Demand Constraint)

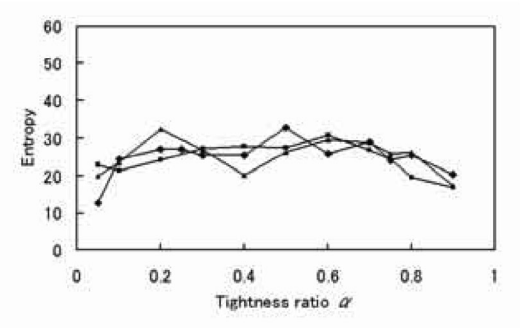

Fig.3. Entropy vs Tightness Ratio

($m1 = M = 4, N = 500$)

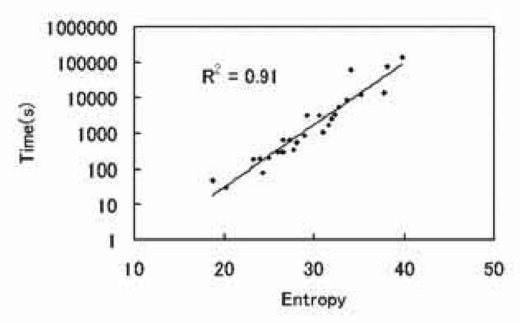

Fig.4. Computational Time vs Entropy

($m1 = M = 5, N = 500$)

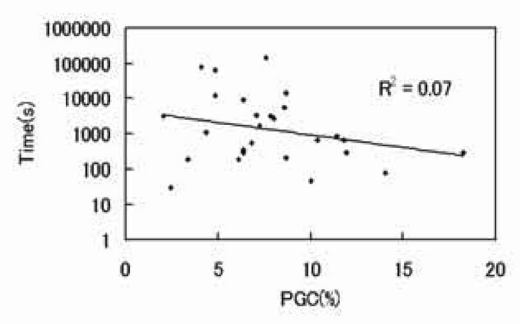

Fig.5. Computational Time vs PGC

($m1 = M = 5, N = 500$)

5 Conclusions

Estimation method for difficulties of multidimensional 0-1 knapsack problems with multiple constraints including demand constraints is presented. From computational experiments, the probablity of p_n such that

$$f^{\mathrm{UB}}(\boldsymbol{x})|_{x_n=0} \geq f^{\mathrm{UB}}(\boldsymbol{x})|_{x_n=1}$$

can be approximately expreed as

$$2^{-\alpha d_n - \beta}$$

and the obtained computational results show that the approximated expression for the probability p_n is effective to estimate the entropy of difficulties of the tested problems.

6 Acknowledgments

I am grateful to thanks to Professor Y. Nakagawa, Kansai University, A. Iwasaki of Information Processing Center and Mr. K. Uda of Graduate School, Okayama University of Science.

逆問題におけるデータ数と散らばりの限度について
On distribution and number of data points in damped inverse problems

*畠山唯達
岡山理科大学情報処理センター
Tadahiro Hatakeyama
Information Processing Center, Okayama University of Science,
Ridai-cho, Okayama 700-0005, Japan

1 はじめに

データ数が少ないながらも、数多くのモデルパラメータを決定しなければいけない問題が存在する。通常、このような場合、モデルパラメータ数 (M) がデータ数 (N) を超えてしまうと逆問題が列決定になってしまい。ユニークな最小二乗解を得ることは出来なくなる (Menke, 1989 など)。このような場合に解を求める有効な方法のひとつとして、モデルに対して「先験的情報」あるいは「制約条件」を与え、最小二乗的なフィッティングと先験的情報との間に重み ($\alpha^2 > 0$) つきの線形和をとって最小化すべき関数 (評価関数) を作る方法があり Stochastic Inversion (Gubbins, 1983) だとか、Bayesian Modelling (Jackson, 1979) などと呼ばれている。この 2 つは思想的に大きく異なるものであるが、最終的な定式化はそっくりであるので、ここでは統一して議論する。

Stochastic Inversion (または Baysian Modelling) で用いられる、モデルパラメータ ($\boldsymbol{x}$) を選択して最小化すべき評価関数は、

$$\begin{aligned} S(\boldsymbol{x}) &= \left(\boldsymbol{y^o} - \mathsf{G}\boldsymbol{x}\right)^T \mathsf{C}_\mathsf{e}^{-1} \left(\boldsymbol{y^o} - \mathsf{G}\boldsymbol{x}\right) \\ &\quad + \alpha^2 \left(\boldsymbol{x} - \boldsymbol{x}^\#\right)^T \mathsf{C}_\mathsf{m}^{-1} \left(\boldsymbol{x} - \boldsymbol{x}^\#\right), \end{aligned} \quad (1)$$

のようなものである。ここで、$\boldsymbol{y} = \mathsf{G}\boldsymbol{x}$ が線形な観測方程式 (非線形の場合は $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ のようになる) で、$\boldsymbol{y^o}$ はデータベクトル、C_e はデータの誤差行列を表す。右辺第 1 項は、単純な最小二乗法を表す。また、第 2 項の C_m は二次形式を形作るモデルへの先験的情報を表す行列 (制約など) であるが、もし、モデルのしかるべき中心が零ベクトルでない場合、第 2 項は、$\alpha^2 \left(\boldsymbol{x} - \boldsymbol{x}^\#\right)^T \mathsf{C}_\mathsf{m}^{-1} \left(\boldsymbol{x} - \boldsymbol{x}^\#\right)$, のようになる。

ここで α^2 は個々の逆問題を解く前に与えておかなければならに定数であるが、当然のことながら解を大きく支配する。畠山 (2006) では、ABIC 法 (Akaike, 1980) を利用して

$$\begin{aligned} ABIC(\alpha^2) &\equiv ABIC1 + ABIC2 + ABIC3 \\ &\equiv N \ln(S(\hat{\boldsymbol{x}})) - M \ln(\alpha^2) \\ &\quad + \ln \left|\left| \mathsf{J}^T \mathsf{C}_\mathsf{e}^{-1} \mathsf{J} + \alpha^2 \mathsf{C}_\mathsf{m}^{-1} \right|\right| \end{aligned} \quad (2)$$

という量を最小化するように α^2 を探しながら逆問題を解く場合での典型的な $\alpha^2 - ABIC$ 曲線を提示し、解法の際の注意を与えた。ただし (2) で、行列 J は線形の場合 G と同一、非線形の場合は解の周りでのヤコビ行列 $\mathsf{J}_{ij} = \partial f_i / \partial x_j$ である。本稿では畠山 (2006) の考察に基づき、合成データを使った具体的な問題について、ABIC 法を用いた Stochastic Inversion の可能性について考える。

2 問題設定

本稿で使用する問題は、世界中に散らばるサイトから得られた地磁気方位 (伏角・偏角) データから、地磁気を球面調和関数展開した係数であるガウス係数

*hatake@center.ous.ac.jp

を求める問題である。地磁気の 3 成分—X(北向き)、Y(東向き)、および Z(下向き)—は強度を含み、ガウス係数に対して線形な関数である。しかし、伏角 $I = \tan^{-1} Z/\sqrt{X^2+Y^2}$ や偏角 $D = \tan^{-1} Y/X$ はガウス係数の非線形関数である。現在の地球磁場に関して X, Y, Z を観測することは容易である。しかし残念ながら、過去の地球磁場の記録である古地磁気データの場合は、強度 $F = \sqrt{X^2+Y^2+Z^2}$ を求めることが困難である場合が多いんで、ほとんどの場合方位データしか使えない。また、サイトの分布は、測定可能な岩石が存在する場所の分布であるので、地球上に満遍なく、十分な数あるわけでもない。

そんな中でもどうにかして意味のある過去の地球磁場モデル (ガウス係数) を求めるために Stochastic Inversion を用いる。本稿では、実際に存在可能なサイトの分布のほかにいくつかの理想的なサイト分布を用い、誤差を含む合成データを使用して Stochastic Inversion に ABIC 法を加えて逆問題を解けるかどうか検証する。たとえサイト数が少ないデータセットでも、ちゃんと ABIC 最小が観察できて、最もよい α^2 を探し出し解を見つけることができるならば、その逆問題解法はうまく行ったと言えるだろう。

基準モデルと合成データの作成

対象モデル (答え) として国際標準磁場 2005 年分点 (IGRF 10th; Maus et al., 2005) を使った。IGRF はのっぺりした平均的な地球磁場からは大きく離れたある一時の「スナップショット」の良い例だと考えられるので、このような問題には適当である。以下のようにして、各サイト分布に対して IGRF モデルより誤差を含む合成データを作成した。

はじめに、リファレンス場 (IGRF) から各観測点での直行成分 (X, Y and Z) を作る。次に、正規分布に従うランダムな合成誤差 ($1\sigma = 1000$ [nT]) をそれぞれの成分に加えた上で、そこから伏角 (I) と偏角 (D) を計算し、「誤差つきの方位データ」とした。各成分に対する 1000 [nT] の観測エラーは、地表の観測点において大体 1.5〜3 度の方位のエラーを表す。

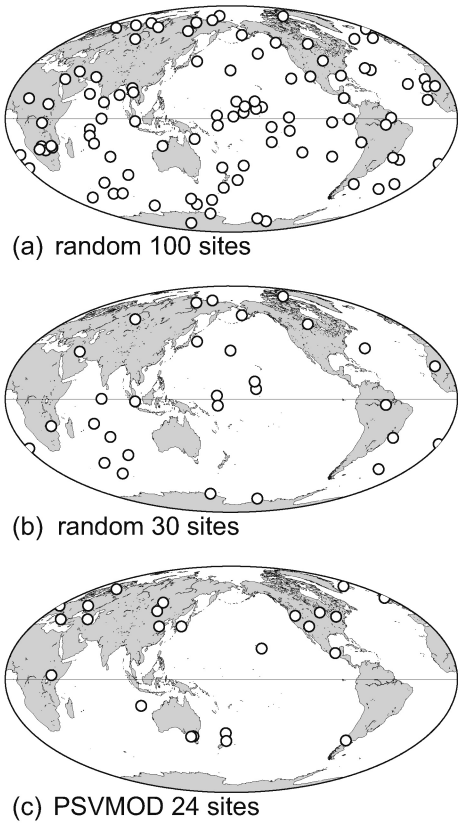

Figure 1: 試験したサイト分布、3 パターン

すべての計算において、打ち切り次数 $\ell_{\max}$ は 10 とした。これは地表での観測量から逆問題を解く場合には十分な次数である。また、すべての計算は倍精度で行った。非線形逆問題で解を見つけるためには反復計算が必要だが、この計算はすべて地心軸双極子 (GAD) 場 $\boldsymbol{x}_{\text{start}} = g_1^0$ から始めた。実際、非線形性はあまり強くなく、この近辺から計算をスタートすれば常に同じ正しい解へ落ちるようである。

サイト分布

サイト分布について 3 つのケースを考えた (Fig. 1)

(1) ランダムな 100 サイト

(2) ランダムな 30 サイト

(3) 実在の古地磁気データの 24 サイト from PSVMOD1.0 (Constable et al., 2000)

サイト分布 (1) と (2) は地表にランダムに分布する。この 2 つから、球面関数的には理想的な場合、どれ

くらいのサイト数があればよいかを判別した。サイト配置は、均等 (規則的) であると、特定の球面調和関数の成分が求まらなくなるので、くランダムである方が好ましい (Mochizuki et al., 1997)。(3) は 3000 年の連続地球磁場モデルを求めるために実際に使用された古地磁気データベース (PSVMOD 1.0) と同一のサイト分布で、17 は北半球、6 サイトは南半球、1 つはほぼ赤道上に位置する。この 3 つのサイトパターンは、それぞれ (1) 十分に数がある (2) 数は少ないが、ある程度、世界中にサイトがある (3) 数が少なく、しかも、地理的偏りがある (現在行われている解析の実際) を想定したものである。

モデルに与える先験的情報

モデルの先験的情報 (拘束条件) として以下のような 2 つの条件を扱った。

(i) コア表面で磁場がスムースになるような Giant Gaussian Process (GGP) (Constable and Parker, 1988) モデルに基づいた PSV モデル。ある瞬間の地球磁場を表すガウス係数の各々の成分 x_j が平均 $x_j^{\#}$ と分散 σ_j をもつ正規分布に従うものとする (Hatakeyama and Kono, 2002)。$x_j^{\#}$ はベクトル $\boldsymbol{x}^{\#}$ の j 番目の要素で σ_j^2 は行列 $\mathsf{C_m}$ の j 番目の対角要素である。これはオリジナルのベイジアンモデリングと stochastic アプローチの哲学に基づいている (Jackson, 1979)。

(ii) コア表面で電流による加熱を最小化する制約条件 (Gubbins, 1983; Korte and Constable, 2003; etc..)。この観点からだと、式 (1) の第 2 項はモデルパラメータの確率密度でなく、物理的制約の表現である。これは高い次数 ℓ にとっては (i) よりもちょっと強い制約である。

(1) 式の $\mathsf{C_m}$ の対角項の逆数 σ_j^2 の大きさはそれぞれ (i) $\sigma_j^2 \sim \mathcal{O}(\ell^{-2}(R_{core}/R_{earth})^{2\ell})$ と (ii) $\sigma_j^2 \sim \mathcal{O}(\ell^{-3}(R_{core}/R_{earth})^{2\ell})$ である。また、$\ell = 1$ と $\ell = 10$ にかかる制約の比は、$[\sigma_{10}/\sigma_1]_{(\mathrm{ii})}/[\sigma_{10}/\sigma_1)_{(\mathrm{i})}] \sim 2$ となる。そのため、(ii) の場合は (i) のケースよりも、高次の項はちょっとだけつぶされて、解はスムーズになることが期待される。

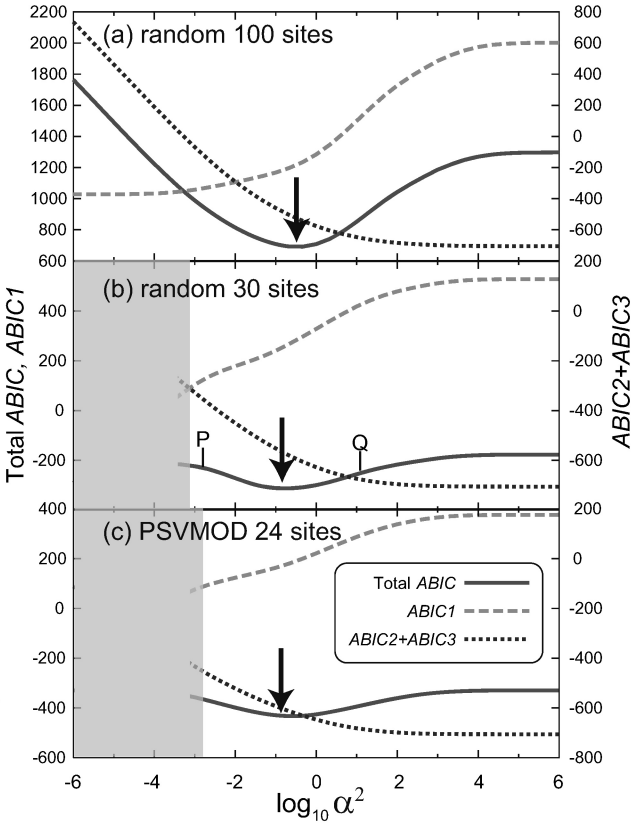

Figure 2: 超パラメータ α^2 を変化させたときの $ABIC$ の変化。詳細な議論は畠山 (2006) を参照のこと。

3 結果

まず、ABIC 最小化法と Stochastic Inversion で超パラメータ α^2 の最良の値について紹介する。

3 つのサイト分布それぞれについて、α^2 を変化させながら各 α^2 で Stochastic Inversion によって解を求め、(3) 式に基づいて $ABIC$ を計算すると Fig. 2 のようになった。ABIC 法は最良の超パラメータはうまく決定でき、その α^2 を用いて適当な解を得ることができている。実際、ABIC 最小の点に比べて大きな α^2 および小さな α^2 を用いた場合、それぞれ解として荒い (エラーを過小評価している) ものと鈍っている (過大評価した) ものが得られている (Fig. 3)。

(1) ランダムな 100 サイトの場合、十分小さな α^2 まで安定な解が求まる。これはデータの持つ情報量

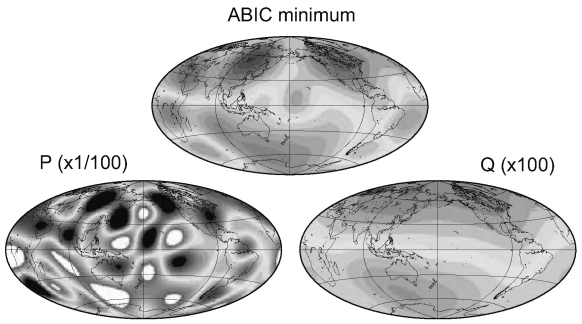

Figure 3: ABIC 法が決める最適な α^2 およびそれよりも 2 桁小さな α^2(Fig.2 の P 点) と 2 桁小さな α^2(同 Q 点) を採用したときの解の違い。解はコア-マントル境界 (CMB: $R_{core}/R_{earth} = 0.544$) での鉛直上向きの磁場で表現している。以下 Fig.4, 6 でも同様。求めたい正解 (IGRF2005) の磁場は Fig.4(a)。

がそれなりに大きいからである。一方、サイトの個数が少ない (2) と (3) のケースでも、$ABIC$ 最小は求まっていて、適当な α^2 の選択をしていると考えられる。しかしながら、最適と診断される α^2 よりもう少し小さい α^2 を採用したとき、評価関数 $\S(\boldsymbol{x})$ を最小化するための逆計算行列の正則性が崩れてしまい、倍精度の計算では安定な解は良く決まらなくなる (Fig.2(b) と (c) のハッチ領域)。もう少しデータが悪くなると、最適の α^2 がこの領域に入ってしまうかもしれない。こういった時、最適な超パラメータとして計算可能な最小の α^2 を採用し解を求めてしまいがちであるが、それは間違っている。「データが質・量ともに不足しているので計算不能。モデルは求まらない」とすべきである。

次に、サイト分布パターンと先験的情報のタイプについて、最良の α^2 が ABIC 最小化法で決定されたときの解を比較する。

Fig. 4 はコアーマントル境界 (CMB) における、(a) 答えのモデル (IGRF2005)、および (b) ランダム 100 サイト、(c) ランダム 30 サイト、(d) PSVMOD 1.0 24 サイトにデータを与えて、ABIC 法を使用して Stochastic Inversion で求めた磁場の鉛直成分 Br である。すべての結果は、磁場の同様の基本的様相がおおよそ決まっていることを示している。解の詳細は異なり、それは主に高次の項の成分に起因している。(b)〜(d) の場合での解の Br は答え (a) と比べてスムーズである。その原因は、逆問題は単純な最小二乗ではなく、stochastic なアプローチで、モデルの先験的情報が使われて、それが高次の項の成分についてより鈍った解を決定しているからである。さらに、サイト分布が貧弱であればあるほど、解はスムーズになる。その理由はおそらく、サイトの数と配置に依存する情報量が各成分に再配分されるが、先験的情報 (モデルへの制約) のためにそれらの大部分は低 ℓ の項へ配分されるので、微細な空間成分である高次の項がぶされてしまうからであろう。しかし、CMB においてガウス係数で定義されるこれらの高次の成分は地表では顕著ではなくなる。なぜなら、CMB における磁場の小さなスケールの構造は $(R_{core}/R_{earth})^{\ell+1}$ のオーダーに従って減衰されてしまうからである (Langel and Estes, 1982)。そのため、古地磁気における観測値は高次の要素に対してそんなに敏感ではなく、その差はわれわれ人間にとって致命的なものではなくなる。

(1) ランダムな 100 サイト、および (2) ランダムな 30 サイトのデータを用いた計算では、次数 ℓ が (1) の場合 5 ないし 6、(2) の場合 4 位までの要素はよく求まっている。モデルの決定度を表す共分散行列の範囲内で期待される IGRF の値に近い。一方、(3)PSVMOD の 24 サイトの場合に対しては、2 次や 3 次といった低い次数でさえ、たとえば g_2^0, g_2^1, g_3^1... というような幾つかの成分が良く解けていない。(Table. 1) この様相は、「サイトの数 (情報量) だけでなく配置も解を求めるために必要だ」と言うことを表している。正解の値から大きく離れた低次の要素は、逆問題の進行上このサイト配置からは「見えない (カーネルに近い)」要素だと推測される。

同様に、モデル解像度行列

$$\mathsf{R} = \left(\mathsf{J}^T\mathsf{C}_\mathsf{e}^{-1}\mathsf{J} + \alpha^2\mathsf{C}_\mathsf{m}^{-1}\right)^{-1}\hat{\mathsf{J}}^T\mathsf{C}_\mathsf{e}^{-1}\hat{\mathsf{J}}, \qquad (3)$$

も幾つかの不確定要素を示している (Fig. 5)。R の対角要素はモデルパラメータの再生産性と信頼性を示している (Gubbins and Bloxham, 1985)。一般的に、サイト分布の偏りが小さい (1) や (2) の場合では、それ

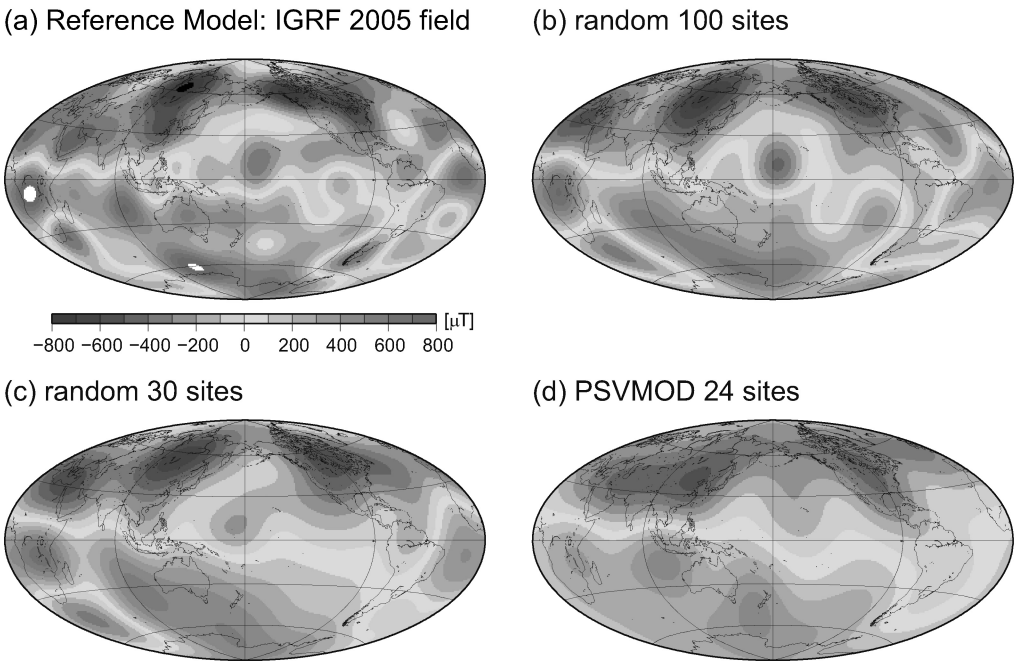

Figure 4: 答えとなる IGRF2005 場 (a) および 3 つのサイト分布に対してデータを与えて逆問題を解いた解 (b)〜(d)

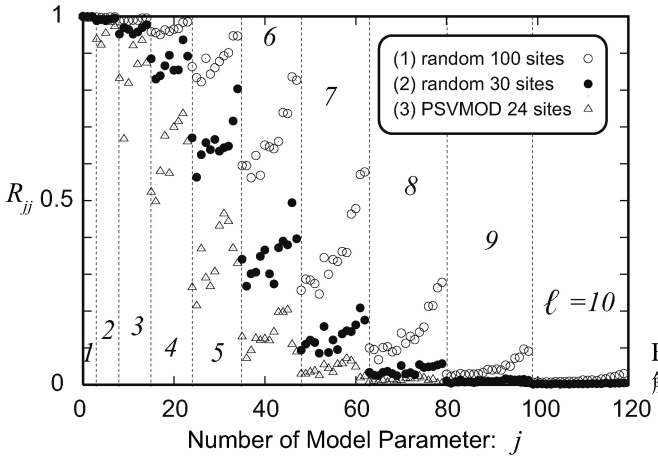

Figure 5: モデル解像度行列 R の対角項。1 に近いほどモデル (答え) の再生産性が良い。

ぞれの次数 ℓ について高い位数 m の項の方が低い m の項よりもよりよい解像度を持っている (Hatakeyama and Kono, 2002)。しかし、(3)PSVMOD 24 サイト分布の場合 (三角で表される) では、低次でも g_3^1 など幾つかの要素が顕著に低い解像度である。貧弱で不均質なサイトの配置 (Fig.1(c)) は球面調和関数展開で定まらない要素をつくるようだ。

もう一つの先験的情報 (ii) を用いて制限された解

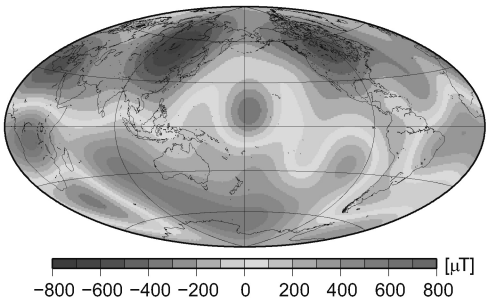

Figure 6: もう一つの先験的情報 (ii) を与えたときの解。比較対象は Fig.4(b)。

が Fig. 6 に表されている。この様子は条件 (i) によるもの (Fig.4 (b)) と大きな違いがない。従って、モデルの低次の成分に関して条件の違いは解に顕著に現れないと言える。

4 まとめ

上記の結果から、Stochastic インバージョンにおける ABIC 最小化法は、最良の超パラメータ α^2 を決め良い解を得るために非常に有効であることが示された。しかし、データやサイトの数が極端に少なかったり、サイト配置が均質でない場合、$ABIC$ を最小にする

α^2 を探すのが困難となる。さらに、不均質なサイト配置は、解の中でも本来求めやすいと考えられている球面調和関数で低次の要素にさえも影響を及ぼす。また、使用した先験的情報による解の違いは小さいことがわかった。

本研究での合成データセットと同じくらいの誤差を含むデータを使用したとしても、Stochastic Inversion を使用して良い解を得るためには、信頼できるデータが少なくとも世界中から30サイトほどは必要である。さもなくば、不十分なデータセットからは低次の項すらもなまった結果を見ることになる。これは先験的情報により縛られたモデルだ。また、なるべく広い地域から、データを得ることが肝心である。特に実際問題として、南半球のデータセットが少ないことは大きな問題である。しかし、こと深海性堆積物に関しては、今後の掘削研究の進展によって南半球でのデータを増やす必要があるだろう。

さらに重大な問題がもうひとつある。もしコアの回転のために信頼性のない偏角データしか得られない堆積物データなどの伏角データだけしか使えない場合、よりたくさんのサイトからデータを得る必要がある。不幸なことに、過去一万年以上の地磁気永年変化を探るためにはいささかサイト数が足りないのかもしれない (Johnson et al., 2003)。

References

- Akaike H., Likelihood and the Bayes procedure, in *Bayesian statistics*, eds. Bernardo, J. M., DeGroot, M. H., Lindley, D. V. and A. F. M. Smith, University Press, Valencia, 143–166, 1980.
- Constable C. G. and R. L. Parker, Statistics of the geomagnetic secular variation for the past 5 m.y., *J. Geophys. Res.*, **93**, 11569–11581, 1988.
- Constable, C. G., Johnson, C. L. and S. P. Lund, Global geomagnetic field models for the past 3000 years: transient or permanent flux lobes?, *Phil. Trans. R. Soc. Lond. A*, **358**, 991–1008, 2000.
- Gubbins, D., Geomagnetic field analysis – I. stochastic inversion, *Geophys. J. R. astr. Soc.*, **73**, 641–652, 1983.
- Gubbins, D. and J. Bloxham, Geomagnetic field analysis – III. Magnetic fields on the core-mantle boundary, *Geophys. J. R. astr. Soc.*, **80**, 695–713, 1985.
- Hatakeyama T. and M. Kono, Geomagnetic field model for the last 5 My: time-averaged field and secular variation, *Phys. Earth Planet. Inter.*, **133**, 181–215, 2002.
- 畠山唯達, ダンピングなどを用いた逆問題における最適解の存在点について, 岡山理科大学情報処理センター研究報告, **27**, 17–20, 2006.
- Hatakeyama T., Effects of the site distribution and the prior information on the inverted geomagnetic field model: a case study applying the ABIC method to the synthetic datasets, *Earth Planet. Space*, **59**, 703–709, 2007.
- Jackson D. D., The use of a priori data to resolve non-uniqueness in linear inversion, *Geophys. J. R. astr. Soc.*, **57**, 137–157, 1979.
- Johnson C. L., Constable C. G. and L. Tauxe, Mapping long-term changes in Earth's magnetic field., *Science*, **300**, 2044–2045, doi:20.1126.science.1032007, 2003.
- Korte M. and C. G. Constable, Continuous global geomagnetic field models for the past 3000 years, *Phys. Earth Planet. Inter.*, **140**, 73–89, doi:10.1016/j.pepi.2003.07.013, 2003.
- Langel, R. A. and R. H. Estes, A geomagnetic field spectrum, *Geophys. R. Lett.*, **9**, 250–253, 1982.
- Maus S., Macmillan S., Chernova T., Choi S., Dater D., Golovkov V., Lesur V., Lowes F., Luhr

	IGRF	Random 100 sites		Random 30 sites		PSVMOD 24 sites	
	(Model)	Sol	Std	Sol	Std	Sol	Std
g_1^0	-29556.8	-29556.8	—	-29556.8	—	-29556.8	—
g_1^1	-1671.8	-1649.0	(109.3)	-1237.3	(218.3)	-1263.8	(461.4)
h_1^1	5080.0	4825.4	(114.6)	5224.8	(272.6)	5354.1	(359.4)
g_2^0	-2340.5	-2245.0	(110.0)	-2066.9	(288.6)	-1401.0	(484.4)
g_2^1	3047.0	2994.6	(140.7)	2686.6	(250.0)	1755.9	(544.0)
h_2^1	-2594.9	-2622.4	(150.9)	-2376.9	(290.1)	-3102.5	(417.0)
g_2^2	1656.9	1602.9	(80.6)	1757.9	(234.4)	1207.2	(244.6)
h_2^2	-516.7	-624.7	(80.8)	-793.0	(184.5)	366.3	(322.7)
g_3^0	1335.7	973.1	(113.7)	1286.2	(235.8)	1075.1	(319.8)
g_3^1	-2305.3	-2241.4	(104.8)	-2468.6	(187.7)	-691.4	(449.9)
h_3^1	-200.4	-315.9	(111.0)	-456.2	(202.7)	-565.4	(332.2)
g_3^2	1246.8	1214.8	(100.4)	1267.7	(235.4)	954.5	(220.1)
h_3^2	269.3	358.1	(103.5)	455.9	(221.6)	-409.4	(281.3)
g_3^3	674.4	580.9	(65.2)	562.1	(188.4)	382.9	(198.9)
h_3^3	-524.5	-521.2	(65.5)	-579.8	(161.9)	191.2	(277.2)

Table 1: IGRF2005 場および各逆問題の解のガウス係数。単位は [nT]。“Sol” は答えの値、“Std” はモデル共分散行列の対角項の平方根で答えの信頼性を表す。計算は $\ell \leq 10$ の範囲でやっているがここでは $\ell = 3$ までを表示する。g_1^0(地心軸双極子) は固定している。

H., Mai W., McLean S., Olsen N., Rother M., Sabaka T., Thomson A., Zvereva T. and International Association of Geomagnetism, Aeronomy (IAGA), Division V, Working Group VMOD, The 10th generation international geomagnetic reference field *Phys. Earth Planet. Inter.*, **151**, 320–322, 2005.

- Menke W., Geophysical data analysis: discrete inverse theory, *Rev. Ed.*, 289pp, Academic Press, 1989, (W. メンケ, 離散インバース理論—逆問題とデータ解析, 古今書院, pp294, 1997)

- Mochizuki E., Yokoyama Y., Shimizu I. and Y. Hamano, Spherical harmonic analysis in terms of unevenly distributed observation points and its applications to geomagnetic data, *J. Geomag. Geoelectr.*, **49**, 1013-1033, 1997.

2007年度情報処理センター研究員報告

「超増加性を持たない非線形ナップザック暗号」

情報処理センター

岩崎彰典

工学研究科電子工学専攻

宇田浩司

１． まえがき

Merkle と Hellman によって提唱された 0－1 ナップザック暗号（MH 暗号）は特殊な 0－1 ナップザック問題である部分和問題に基づいた公開鍵暗号である．部分和問題は NP 困難な問題であるが，秘密鍵として超増加数列を用いて，一意かつ容易な復号を可能にしている．公開鍵は秘密鍵にモジュラー乗算を行い，見かけ上解くことが困難な問題にしている．Shamir は公開鍵から超増加数列を導き MH 暗号を解読した．Lagarias と Odlyzko は LLL アルゴリズム(以下 LLL と略記)を用いた低密度攻撃によって MH 暗号が解読できることを示した．MH 暗号に冗長性を持たせ，鍵数列の密度を高める研究が小林ら[1]によってなされている．

一方で，ナップザック暗号は，最近の組合せ最適化問題の解法の進歩にも注意しなくてはならない．仲川ら[2]は改良代理制約法（ISC）を開発した．この方法は不等号制約条件を持つ非線形ナップザック問題を解くために開発されたものであり，部分和問題を解く能力は未知数であるが，これらの最適化技法への耐性も調べておく必要がある．

本発表では，非線形ナップザック暗号の LLL 及び ISC への耐性を調べた結果を報告する．

２．暗号化の手順

(1)暗号に用いる非線形ナップザック問題の定式化

まず，$(n\times k)$ 個の要素を持つ行列 B を考える．

$$B\equiv\left[b_{ij}\right],\quad \left(i=1,2,\mathrm{K}\ ,n,\ j=1,2,\mathrm{K}\ ,k\right)$$

の行を変数，列を各変数の取る案と考え，i 番目の変数に対し k_i 番目の案のみを取ることにする．このとき，非線形ナップザック問題を，B とある整数 c が与えられたとき，

$$c=\sum_{i=1}^{n}b_{ik_i}$$

を満たす k_i の組合せを求める問題と定式化する．

(2)鍵の生成

n 文字をブロックとする暗号系を考える．$l \geq 2$ 個の1から次の関係を満たすビットマスク列 $s_i, (i = 1,2,\mathrm{K}\ n)$ を作る．

$$s_1 \,\&\, s_2 \,\&\, \Lambda \ \& \, s_n = 0$$
$$s_1 \oplus s_2 \oplus \Lambda \ \oplus s_n = 1$$

但し，&，⊕はそれぞれビットマスク列の論理積，論理和を表す．必要なビットマスク列のビット数は $l \times n$ となる．この条件は復合化の際一意に復号化をするために必要である．例を $n = 3, l = 3$ の場合で示す．

$$s_1 = (001110000) = 112$$
$$s_2 = (010001100) = 140$$
$$s_3 = (100000011) = 259$$

次に各ビットマスクの1をビット位置が重複しないようランダムに $(l - m)$ 個だけ0に変える．但し $l - 2 \geq m \geq 1$ とする．これは後で述べるモジュラー乗算のパラメータを逆算されないためである．これにより各ビットマスク s_i から 2^{l-m} 個の案 $a_{ij}, (j = 1,2,\mathrm{K}\ ,2^{l-m})$ を作ることができる．s_1 から生成される案の例を示す．

$$a_{11} = (001000000) = 64$$
$$a_{12} = (001010000) = 80$$
$$a_{13} = (001100000) = 96$$
$$a_{14} = (001110000) = 112$$

他のビットマスク列に対し同様の操作を行い，これを秘密鍵 A とする．先のビットマスクから作られた例を10進数で示す．

$$A = \begin{bmatrix} 64 & 80 & 96 & 112 \\ 8 & 136 & 140 & 12 \\ 259 & 1 & 3 & 257 \end{bmatrix}$$

ここで，ビットマスク列 s_i は超増加性を持つが，ビットマスクから生成された案の要素列は超増加性を持たないことに注意されたい．

A に次のモジュラー乗算を行い，公開鍵 B とする．

$$B = A \times w \bmod p$$

但し，$p > 2^{nl} - 1$とし，wとpは互いに素とする．

(3)暗号化

平文を$e_i, \left(1 \le e_i \le 2^{l-m}, i = 1,2,\mathrm{K}\ ,n\right)$とすれば，暗号$c$は，公開鍵$B$の要素を$b_{ij}$として，

$$c = \sum_{i=1}^{n} b_{ie_i}$$

となる．公開鍵Bと暗号cから平文の組み合わせを求めることは非線形ナップザック問題を解くことと等価であり NP 困難である．cを暗号文として受信者に送信する．

(4)復号

復号には$w \times w^{-1} \bmod p = 1$なる$w^{-1}$を用いて，

$$c' = c \times w^{-1} \bmod p$$

を求める．

c'とビットマスクs_iの論理積をe'_iとする．

$$e_i' = c' \& s_i, \quad \left(i = 1,2,\mathrm{K}\ ,n\right)$$

e'と秘密鍵Aの要素を比較し，

$$e_i' = a_{ik_i}, \quad \left(i = 1,2,\mathrm{K}\ ,n\right)$$

となるk_iを求めることにより復号することができる．

３．非線形ナップザック暗号の密度

非線形ナップザック暗号の公開鍵行列の行要素数は2^{l-m}個となるので，暗号化するブロックの文字数をnとすれば公開鍵行列の要素数は$n \cdot 2^{l-m}$個となる．従って，LLL の対象となる暗号ベクトルの要素数は$n \cdot 2^{l-m}$個となるので，この暗号ベクトルの密度dは，

$$d = \frac{n \cdot 2^{l-m}}{\log_2 \max\left(b_{ij}\right)}$$

である．

公開鍵からモジュラー乗数のパラメータpを逆算されないためには，mが大きくなければならないが，$l \approx m$では LLL により，$l >> m$では ISC により解読される可能性がある．

計算機実験では，$m = 1$とし，nとlを変化させて LLL と ISC に対する暗号の耐性を調

べた．

４．LLL と ISC に対する耐性

実際に暗号化の手順から作成した公開鍵に対して LLL と ISC よる解読を行った．表 1 に LLL による解読率を示す．表 1 から文字数 n とビットマスクの1の数 l の増加により，暗号の解読率は飛躍的に下がることがわかる．次に1の数 l を 6 に，文字数 n を 6，8 とし，b_{ij} を変えて d を変化させた実験を行った．実験結果を図 1，2 に示す．図中の▲が LLL，■が ISC による解読率を表す．縦の実線は本実験での非線形ナップザック暗号の密度である．図 1 の結果では，暗号の解読率が LLL より ISC の方が解読率に優れている．しかし，図 2 から ISC では解読されず，LLL にも耐性を持つ密度の範囲があることがわかる．

５．まとめ

今回提案した方式は非線形問題を最も単純な形で暗号化に利用したもので，「荷物を入れるか入れないか」によって暗号化していたこれまでのナップザック暗号系に対して，「どの荷物を入れるか」で暗号化を行うという新しい方法の有効性を示した．秘密鍵から公開鍵への変換方法や関数の値の決定方法などについては，さらに検討する余地もあると思われ，それらの研究を進めていくことで非線形離散最適化問題を応用した，安全性の高い暗号技術の研究が発展していくことが期待される．

表 1　LLL による解読率（m=1）

l=	4	5	6
d=	2	3.2	5.3
n=4	5.430E-01	8.935E-02	3.964E-03
n=5	3.130E-01	1.936E-02	2.046E-04
n=6	1.630E-01	3.644E-03	1.975E-05
n=7	8.825E-02	5.135E-04	2.000E-06
n=8	3.190E-02	9.440E-05	<1.000E-07

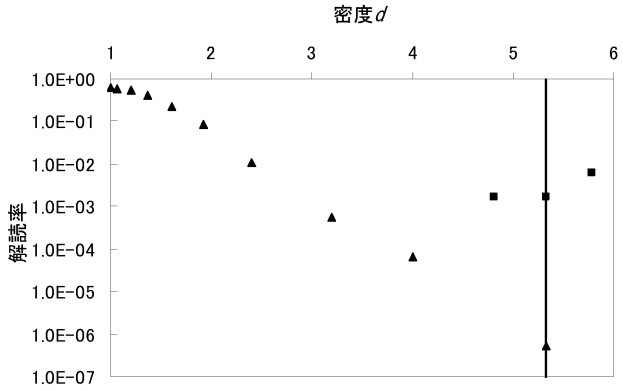

図 1　密度と LLL，ISC による解読率（l=6, n=6）

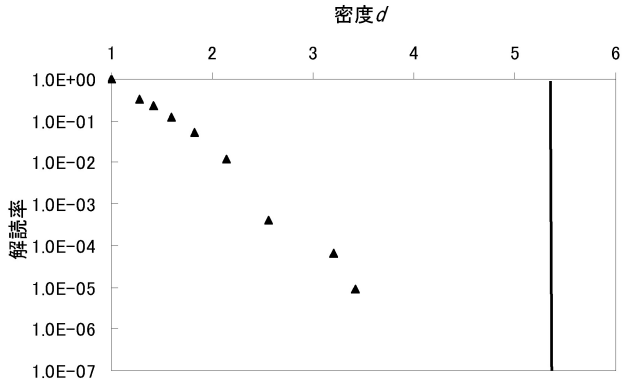

図 2　密度と LLL，ISC による解読率（l=6, n=8）

文献

[1]小林邦勝，“ナップザック暗号の安全性向上に関する一考察，”　信学論（A），vol.J79-A，no.8，pp.1339-1343，August.　1996.

[2]Y. Nakagawa, “An improved surrogate constraints method for separable nonlinear integer programming,” J. Oper. Res. Soc. Jpn., vol.46, no2, pp.145-163, June. 2003.

ＬＭＳ「MOMOTARO」における教育の質保証

大西荘一*,** 北川文夫* 榊原道夫* 河野敏行* 山本敏弘* 荒川智昭*
西崎書彦*,*** 田坂仁昭**

＊岡山理科大学大学院 総合情報研究科 情報科学専攻
＊＊岡山理科大学 情報処理センター ＊＊＊加計教育コンソーシアム

１．はじめに

筆者らの e-Learning の取り組みは情報処理センター研究報告の第 22 号，23 号，24 号，26 号，27 号[1][2][3][4][5]，及び学外誌[6][7][8],学会発表[9][10][11][12][13][14][15][16][17]に報告してきた．今回は，e-Learning（遠隔授業）における教育の質保証について LMS「MOMOTARO」では，どのような考慮をしているかを報告する．

マルチメディア技術の進歩とブロードバンドインターネットの普及により e-Learning は実用段階にきている．2001 年３月の大学設置基準の改定で，教育改革の一手法として注目され，多くの教育機関で単位認定を伴う e-Learning が実施されてきている．e-Learning が普及するとともに，近年その教育の質が問われ始めている[18]．特に，VOD による非同期双方向授業において，対面でないが故の利点がある一方，課題もあり，教育の質を落とさないための工夫が要求される．筆者らは，LMS の機能は教育の質保障のために重要なポイントであることを認識し，LMS「MOMOTARO」の改良を続けている．

２．加計グループサイバーキャンパスの履修状況

加計グループサイバーキャンパスの経緯は参考文献[5]に述べている． 表 1 は平成 17 年度～１９年度の MOMOTARO で管理されている科目数と受講登録者数である．単位互換受講登録者数は他大学の科目を単位互換で受講した大学生の人数である．

表 1. MOMOTARO に登録された科目数と延べ受講登録者数（単位:名）

年度	科目数	受講登録者数（大学生）	受講登録者数（高校生）	受講登録者延べ総数	単位互換延べ受講登録者数
17	7	735	118	853	415
18	26	3079	73	3152	2128
19	34	4436 570(通信生)	93	5099	3836

平成 19 年度は平成 18 年度に比べ，科目数は 8 科目増え，受講登録者延べ総数は 1947 名増えて 5000 名を超えた．平成 19 年度前期で，単位互換受講登録者数のうち出席回数が規定の 9 回を満たした者は科目により大きなバラつきがあり，16%～71%であった．全科目の平均は約 60%であった．

表２は平成 18 年度～19 年度の所属別の受講登録者の実数（通信生を除く）である．平成 19 年度の受講登録者実数は，全大学で平成 18 年度よりも増加している．岡山理科大学は在学生の約 17%，倉敷芸科大は約 14%の学生がサイバーキャンパスを利用しており，新しい教育形態として定着した感がある．

表2．平成 18 年度～19 年度　所属別受講登録者の実数（単位：名）

	岡山 理科大	倉敷 芸科大	千葉 科技大	吉備 国際大	九州 保福大	高校	合計
平成 18 年度	553	122	21	173	40	46	９５５
平成 19 年度	832	231	58	222	54	73	１４７０
平成 19 年度 在学生	4910	1600	1665	3720	2260	—	１４１５５
在学生比率%	17%	14%	3%	6%	2%	—	１０%

３．ＬＭＳ「MOMOTARO」の開発履歴

表３はＬＭＳ「MOMOTARO」の開発履歴である．平成１８年度から平成１９年度にかけて教育の質保証に関する機能と MOMOTARO の管理機能を主に研究開発した．

表3．LMS「MOMOTARO」の開発履歴

運用時期	バージョン	主な特徴
平成 15 年度後期 ～平成 16 年度前期	Ver1.0	Perl+ACCESS ４段階の権限，アンケート提出集計機能 回答自動集計機能
平成 16 年度後期	Ver2.0	PHP+MySQL 環境への移行 RDBMS によるデータベースリアルタイム処理
平成 17 年度前期 ～同年度後期	Ver3.0	高大連携と大大連携の統合 複数科目開講への対応 管理機能の強化
平成 18 年度前期	Ver4.0	通信生ユーザへの対応 科目メイン画面へ受講機能の集約
平成 18 年度後期 ～平成 19 年度前期	Ver4.1	教育の質保障機能の強化
平成 19 年度後期	Ver5.0	管理者モードの追加 管理の効率化

４．教育の質に影響する要因

e-Learning において教育の質に影響する要因には次のことが考えられる．

（１）授業コンテンツの内容
（２）教授力（授業の巧拙）
（３）通信の品質（速度，安定性，セキュリティ）
（４）運営組織力（人的パワー，受講者への対応，事務処理）
（５）ＬＭＳの機能

教育の質は上記要因の複合で決まるが，ＬＭＳにおいて教育の質の劣化を防ぐ機能を研究することは重要である．

５．ＶＯＤによる非同期双方向授業の利点と課題

e-Learning（遠隔授業）にはライブ形式の同期双方向とＶＯＤによる非同期双方向があるが，下記の利点からＶＯＤ方式が主流である．加計サイバーキャンパスにおいてもＶＯＤが主力である．

ＶＯＤの利点として

・いつでも，どこでも受講ができる
・何度でも復習が可能
・講師は授業内容（ＶＯＤコンテンツの内容）を十分に検討できる
・講師や受講生の都合による影響を受けない

があるが，一方次の問題がある．

・通信環境の影響を受ける
・受講生の学習意欲に大きく影響される

これらの問題の解決が課題となっている．

６．ＶＯＤにおける教育の質劣化の要因

ＶＯＤによる授業は下記の要因で，教育の質の劣化をまねくことが懸念される．

劣化要因１：ＶＯＤでの学習は，集中力が持続しにくい
劣化要因２：計画的な学習をせず，学期末に集中する傾向がある
劣化要因３：受講生の本人確認が困難
劣化要因４：教員は受講生の学習状況を把握しにくい
劣化要因５：受講生と教員とのコミュニケーションが不足する
劣化要因６；通信環境やアクセス集中により画像・音声が劣化する
劣化要因７：セキュリティ攻撃による個人情報漏洩やサーバダウンによる受講不能

これらの要因による教育の質劣化を防ぐことも，ＬＭＳの重要な役割であると考える．

７．ＬＭＳ「MOMOTARO」の教育の質保証に関連する機能

教育の質保証に関して，MOMOTARO に次の機能を実装した．

・ＶＯＤ教材の分割配信機能
・ＶＯＤ教材の学習状況確認機能
・出欠管理機能（授業アンケートと小テスト機能）
・コミュニケーション機能
・受講生の学習行動パターン把握機能（アクセスログの収集）
・情報セキュリティの強化

（１）ＶＯＤ教材の分割配信機能

１コマ９０分のＶＯＤを集中して連続受講することは極めて困難である．２０分～３０分程度に分割配信することで，劣化要因１に対処している．また，連続的に読み込むＶＯＤの容量が減少し，ネットワーク及びサーバ負荷が軽減されることになり，劣化要因６に対処している．

図１はＶＯＤ受講のメイン画面である．１コマを３分割している．

図１．ＶＯＤ受講のメイン画面

（２）ＶＯＤ教材の学習状況確認機能

ＶＯＤへのアクセス時間をアクセスログから計算し，ＶＯＤの受講済み割合を％で図１の画面にリアルタイム表示する．受講者にＶＯＤの学習済み時間を認識させ学習を促すことにより，劣化要因２と４に対処している．

（３）出欠管理機能（授業アンケートと小テスト問題）

出席カードを MOMOTARO 上から送信することにより出席確認をしている．出席カードを送信すると自動的にデータベースに登録され出席表にマーキングされる．出席カードには次の機能がある．

・ＶＯＤをまじめに学習しているかを確認するためにアンケートや授業に関する小テスト問題を付加することができ，小テストの自動採点とアンケートの自動集計が可能である．
・ＶＯＤへのアクセス時間（学習時間）がＶＯＤの時間の９０％を越えなければ，出席カードを送信できない．
・出席カードの送信期限が設定できる．

これらの機能により，劣化要因１，２，４に対処している．また，授業毎のアンケートで授業の改善にすばやく対応できる．図２は講師用の出席確認画面である．図３は出席カードの例である．

図２．講師用出席確認画面

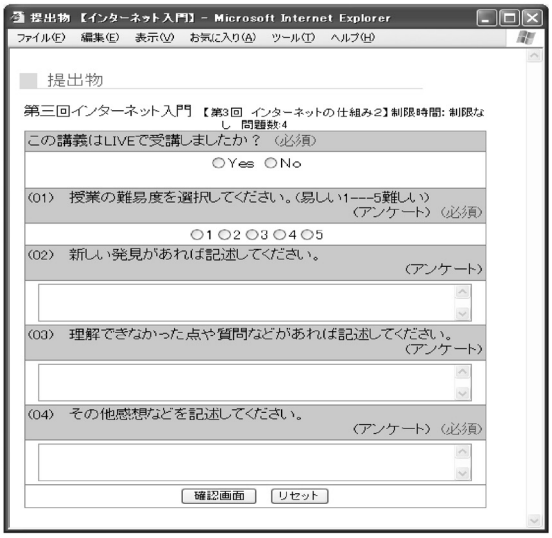

図３．出席カード

（４）コミュニケーション機能

全受講生に対して，管理者からの「お知らせ」と管理者への「お問い合わせ」機能で受講生と管理者間とのコミュニケーションがとれる．各科目ごとの「科目のお知らせ」機能，「講義内容に関する質問／回答」機能で講師と受講生間のコミュニケーション，「チャット」機能で受講生間及び受講生と講師間のコミュニケーション，「メール一斉送信」機能で講師から指定した受講生への連絡がとれる．これらの機能により劣化要因５に対処している．

図４．は「講義内容に関する質問／回答」の画面，図５．は「メール一斉送信」の画面である．

（５）受講生の学習行動パターン把握機能

受講生の学習行動を把握するために，各種アクセスログをとっている．アクセスログは，システムログ（MOMOTARO へのアクセス），科目ログ（科目へのアクセス），教材ログ（科目内の教材へのアクセス）があり，ログ内容は「氏名」，「学生番号」，「ログイン日時」，「ログアウト日時」，「クライアントのＩＰアドレスもしくはドメイン名」である．これらの情報を分析することによって，時間帯別学習度数，曜日別学習度数，連続学習時間数度数，学習傾向，Blended Learning にお

ける対面授業とＶＯＤ授業の比較，などが把握できる．講師が受講生の学習行動パターンを知ることにより，生活面も含めた学習指導が可能になる．この機能により，劣化要因２，４に対処している．

図4.「講義内容に関する質問／回答」の画面

図5.「メール一斉送信」の画面

（６）情報セキュリティの強化

ＬＭＳ「MOMOTARO」は Web アプリケーションの一種である．Web アプリケーションとしてのセキュリティ対策をしなければならない．セキュリティ対策としては，ＳＱＬインジェクション，クロスサイトスクリプティング，ＯＳコマンドインジェクションの攻撃に耐えられるよう設計している．また，ＣＶＳＳ（ Common Vulnerability Scoring System ）[19][20]で MOMOTARO の脆弱性を評価している．これらにより，劣化要因７に対処している．

８．今後の課題と取り組み

本論文であげた教育の質の劣化要因７項目のうち劣化要因３：受講生の本人確認が困難　いわゆる「なりすまし受講」について，MOMOTARO はいろいろな機能の複合で対策しているが，十分な対策ができているとは言えない．個人認証技術は生体認証など進歩はしているが，e-Learning の特徴から「なりすまし」の完璧な防止は極めて困難である．費用対効果を考えると，単位認定時に面接試験を実施し，厳密な成績評価をするのが得策と考える．インターネットの世界は好ましいことではないが、性悪説で考えざるを得ない状況である。ＬＭＳも各種攻撃に備える十分なセキュリティ対策が要求される。「なりすまし」防止対策とセキュリティ対策は今後の大きな課題である．

ＬＭＳ「MOMOTARO」の今後の主な取り組みを次にまとめる．

（１）「なりすまし」防止機能
（２）アクセスログの自動分析機能
（３）成績評価及び管理機能
（４）個別データバックアップ機能及びビューアーの開発
（５）携帯電話との連携機能
（６）セキュリティ強化

謝辞

本取り組みの推進にあたり，連携高校関係者，岡山理科大関係者，NTT 西日本・岡山支店関係者のご協力を得ましたことに感謝いたします．また，岡山県情報ハイウェイ，岡山市地域情報水道を利用しており，これらの設置者である岡山県，岡山市に感謝いたします．

参考文献

［1］　大西荘一，山本英二，市田義明，惣臺聖治
「２．４ＧＨｚ帯無線 LAN によるキャンパス内　どこでも学習　基盤の構築
岡山理科大学情報処理センター研究報告第 22 号，pp. 29-35，2001 年 3 月

［2］大西荘一，榊原道夫，市田義明，堂田周治郎，山本英二，惣臺聖治
「インターネット利用遠隔授業による高大連携教育」
岡山理科大学情報処理センター研究報告第 23 号，pp. 15-20，2002 年 3 月

［3］大西荘一，榊原道夫，橋井幸子，鶴将幸，村山真一，市田義明，堂田周治郎，惣臺聖治
「インターネット利用遠隔授業による 7 高校との高大連携教育」
岡山理科大学情報処理センター研究報告第 24 号，pp11-17，2003 年 3 月

［4］　大西荘一，榊原道夫，秋山雄亮，青嶋　智，田坂仁昭
「インターネットを利用した広域高大連携教育」
岡山理科大学情報処理センター研究報告第 26 号，pp. 31-37，2005 年 3 月

［5］　大西荘一，北川文夫，榊原道夫，河野敏行、青嶋　智，山本敏弘、西崎書彦、田坂仁昭
「加計グループ・サイバーヤンパスを支えるＬＭＳ「MOMOTARO」」
岡山理科大学情報処理センター研究報告第 27 号，pp. 37-43，2006 年 3 月

［6］　大西荘一　「インターネット遠隔授業による高大連携教育」
私学経営 Vol. 346，　pp. 23-30，　2003 年

［7］　大西荘一　「インターネット遠隔授業による高大連携の広域化」
（独）日本学生支援機構　大学と学生　Vol. 25，pp21-27，　2006 年　３月

［8］　北川文夫，大西荘一
「対面講義と e-learning(LMS+VOD) とを併用した講義形式の実践と分析」
日本教育情報学会論文誌　22 巻 3 号，　pp. 57-66　2007 年 1 月

［9］橋井幸子，鶴将幸，村山真一，榊原道夫，大西荘一，市田義明，堂田周治郎，惣臺聖治
「インターネット利用遠隔授業による高大連携教育」
日本教育工学会第 19 回全国大会論文集，pp. 911-912，2003 年 10 月

［10］　青嶋智，秋山雄亮，大西荘一，榊原道夫
「インターネット利用による高大連携の仕組みと評価」
日本教育情報学会第 21 回年会論文集 pp. 142-145，2005 年 8 月

［11］　村山真一，持田龍也，橋井幸子，鶴将幸，秋山雄亮，藤本貴壽，大西荘一，榊原道夫
「インターネット利用遠隔授業における Web システムの開発」
情報処理学会第 66 回全国大会論文集 pp. 4-395-396，2004 年 3 月

［12］　秋山雄亮，鶴将幸，村山真一，青嶋智，大西荘一，榊原道夫
「独立した複数組織の連携教育用 Web システム」
情報処理学会第 67 回全国大会論文集 pp. 4-447-448，2005 年 3 月

［13］　秋山雄亮，青嶋　智，大西荘一，榊原道夫
「広域遠隔授業における Web システムの開発」
日本教育情報学会第 21 回年会論文集 pp. 236-239，2005 年 8 月

［14］　青嶋　智，大西荘一，榊原道夫，河野敏行，山本敏弘
「加計グループの e-Learning への取組」
日本教育情報学会第 22 回年会論文集，pp. 148-149，2006 年 8 月

［15］　山本敏弘，大西荘一，青嶋智，榊原道夫，河野敏行
「広域連携授業に対応した LMS MOMOTARO V4 の開発」
日本教育情報学会第 22 回年会論文集，pp. 130-131，2006 年 8 月

［16］　山本敏弘,大西荘一,荒川智昭,榊原道夫,　北川文夫,河野敏行
「LMS「MOMOTARO」における管理の効率化」
日本教育情報学会年会論文集,　pp.170-171,　2007 年 8 月
［17］　荒川智昭,大西荘一,山本敏弘,榊原道夫,　北川文夫,河野敏行
「LMS「MOMOTARO」における受講者の受講状況の把握と学習促進機能～e-learning における教育の質保障のために～」　日本教育情報学会年会論文集 pp.172-173,　2007 年 8 月
［18］　特集「e-Learning における高等教育機関の質保証への取り組み」
（独）メディア教育開発センター　メディア教育研究　Vol.3,　No.2,　2007
［19］　共通脆弱性評価システム CVSS v2 概説
http://www.ipa.go.jp/security/vuln/SeverityCVSS2.html
［20］　CVSS でぜい弱性を評価してみよう
http://itpro.nikkeibp.co.jp/article/COLUMN/20070808/279327/