

最小二乗法標準プログラムSALSの紹介

中川, 徹
東京大学理学部科学教室

<https://doi.org/10.15017/1474249>

出版情報：九州大学大型計算機センター広報. 12 (3), pp.210-219, 1979-09-20. 九州大学大型計算機センター
バージョン：
権利関係：



最小二乗法標準プログラム SALS の紹介

中 川 徹*

最小二乗法解析のための使いやすい標準プログラムが、最近東京大学大型計算機センター（以下、東大センターと略記）で開発された。統計学的にも数値解析的にもよく吟味されたシステムであり、自然科学をはじめ種々の分野のデータ解析に便利である。九州大学大型計算機センター（以下、九大センターと略記）にも移植され、公開使用を開始した。

1. SALS の開発の目的

最小二乗法は、自然科学などの測定データの解析に広く使われている基本的な手法である。我々が SALS の開発を始めるより以前にも、多くのユーザーが自作の最小二乗法プログラムを使用しており、東大センターにもいくつかライブラリーとして登録されていた。しかし、そのほとんどが統計学的にも数値解析的にも旧式で、性能や使い勝手に多くの問題があった。特に自然科学においては、大規模で複雑な理論式を用いることが多いので、社会科学などを対象として作られている“統計プログラムパッケージ”などはほとんど使えない。パラメーターについて非線形の理論式（モデル）で速く安定に収束させ、同時に高精度の解が要求される。それには、複雑で緻密なプログラミングが必要で、ユーザー一人一人が今まで大変な苦労を強いられてきた。解を求めるのが精一杯であったから、統計的な情報を詳しく出力することなどは手が回らないのが実情であった。

しかし、考えてみると、最小二乗法や統計的手法はいろいろな分野に共通に使えるはずである。そこで、共通の困難に直面している多くのユーザーが、物理や化学などの個々の分野にとらわれずに、データ解析一般の問題としてとり組めば、共同で解決できるはずである。ただし、それには、ユーザーだけでは不可能で、統計や数値解析の専門家と協力することが必要なことは明らかであった。

そこで、我々は 1975 年 4 月に呼びかけをして、統計・数値解析・計算機の専門家と物理や化学のユーザーとで、十数名の共同のグループを作った。そして、「自然科学における測定データの最小二乗法解析のための汎用標準プログラム」を目標として、開発を始めたのである。統計学的にも数値解法の面でも、近年の新しい研究成果を取り入れて、使いやすく信頼できるシステムを作ることが基本的な方針であった。その後、東大センターおよび科研費特定研究の援助を得て、4 年がかりで完成したのが、今回公開する SALS（第 2 版）である。

SALS の説明と使い方は、東大センター発行**の「SALS 利用の手引き」に詳しく記してある。

* 東京大学理学部化学教室

** 東大センター発行の「SALS 利用の手引き」のジョブ制御文の部分を九大用に変更した「SALS 利用の手引き（九大センター版）」を 10 月中旬に九大センターから発行する予定である。

このうち、「第1部, 基礎篇」は初めて使う人が必ず読むべきもので, これだけでSALSが使える。「第2部, 制御・解法篇」はSALSを使いこなすための, 一般ユーザーの手引きである。実際に使う時にはこれらを読んでいただくことにして, 本稿ではSALSの概要だけを述べる。

2. SALS の機能

SALSとは, **S**tatistical **A**nalysis with **L**east-Squares **F**itting の略であり, サルスと呼ぶ。

SALSは, 最小二乗法を中心とするデータ解析のための汎用のプログラム・パッケージである。その適用分野には一切制限がない。もともとは, 物理や化学を中心とする分野の, 精密な実験データの解析を主眼としたものであるが, 薬学・医学・農学さらには社会・人文科学などの諸分野のデータ解析にも使える。実際, 東大センターでのSALSのユーザーの分布は, 後者のような周辺分野のユーザーの方が多いぐらいである。

SALSの機能は, (広義の) 最小二乗法解析, いいかえるとあてはめ(Fitting)である。すなわち, 測定データに対して, それを説明できると思われる理論式を用いて, その理論式中の未知パラメータを調節し, 測定データを最もよく再現出来るパラメータの値を決定する。また, 測定値の誤差に由来するパラメータ値の不確定性など, 統計情報を出力する。

あてはめるべき理論式は, それぞれの問題によって異なるから, 各ユーザーが自分でサブルーチンの形にプログラミングしなければならない。この理論式は, 多項式のように未知パラメータに対して線形のものでよいし, フーリエ級数や指数関数・ガウス関数などのようにパラメータに対して非線形であってもよい。たいていのちよとした問題なら, 後述の指数減衰関数の重ね合わせの例のように, フォートランカード20枚程度でこのサブルーチンが作れる。その一方で, この理論式のサブルーチンがどんなに長大で複雑なものになってもよい。例えば, 分子分光学的問題では, 数千のデータがあるが, 個々の測定値に対応した理論式は, 数十次元のエネルギー行列の固有値の差として, 数値的にしか表現できない。また別の例では, 多次元連立微分方程式の答えが個々の測定値と対応する。このような場合に, SALSに接続すべき理論式のサブルーチン(SALSではモデル関数サブルーチンと呼んでいる)が, それ自体で構造を持ったサブルーチンの体系になっていてもよいのである。

非線形のモデルで最小二乗法を行なえることがSALSの特長の1つである。非線形問題では, 解が発散したり振動したりして収束しにくいことが多いから, これを安定に速く収束させるようにアルゴリズムを工夫している。現在は, ダンピング付きのGauss-Newton法とFletcherの修正Marquardt法の2種を組みこんである。これらは, 単純なGauss-Newton法では発散してしまうような場合にも, 残差二乗和が小さくなる方向を見出すための工夫を含んでいる。

また, 線形最小二乗解法についても, 速く・小容量で・高精度の解法を採用している。正規方程式を解くには掃出し法ではなくCholesky法を用いる。さらに, 正規方程式を立てずにヤコビアン行列を直接解く方法が, 計算時間は同等で, より精度が高いことが知られており, この新しい解法がSALSの標準オプションである。これには, 修正Gram-Schmidt法, Householder法, お

よび特異値分解法がある。これらの解法はすべて、計算機の精度を基準にしてランク落ち（パラメーター間の相関が強いため独立には決められない場合）を判定しており、その場合には低ランクでの解を求める。特異値分解法（および正規方程式を扱う場合の固有値分解法）は、このようなランク落ちの場合にパラメーターを決定する最も合理的な解法である。このように、線形解法が高精度で、ランク落ちの判定を行なっていることは、非線形最小二乗法のアルゴリズムを安定にするためにも良い効果をもっている。

次に統計的な機能を挙げる。

単純な重みなしの最小二乗法のほか、個々の測定値の測定精度を指定して重みつき最小二乗解法を行なうことができる。

あてはめ結果の良否（適否）を診断するためには、測定値・計算値・残差の一覧表の出力の他に、残差のヒストグラムや残差の正規確率プロット（正規分布からのずれを見る）を出力する。さらに、データ番号を横軸にした残差のグラフが出力されるので、系統的な残差が残っているかどうかを一目で見ることができる。残差二乗和から標準偏差を計算して出力しており、 χ^2 検定に用いることもできる。

誤差伝播の法則を用いて、決定したパラメーターの誤差（不確定さ）とその誤差行列（分散共分散行列）および相関係数行列を計算する。また、最終パラメーター値を用いて計算した理論値（測定値に対応する）の推定誤差をも計算して出力する。

もし、モデルのパラメーターに関する微分係数（ヤコビアン行列）が解析的に計算できるならば、ユーザーのモデル関数サブルーチン中でヤコビアン行列を計算することが、あてはめの計算速度と精度の上から、望ましい。しかし、数値微分（差分近似）でしか計算できない場合には、ヤコビアン行列を計算しないオプションを使ってもよい。この時にはSALS内で数値微分によりヤコビアン行列を計算する。また、ユーザーが解析的に計算したヤコビアン行列と、SALSによる数値微分のものと比較して、ユーザーが作成したプログラム中に誤り（矛盾）がないかどうかチェックする機能がある。あてはめを実行する前に必ずこのチェックを試みるとよい。

モデル関数を実験式であったり、近似式であったりするときには、どのようなパラメーターの項を入れるかを判断することが大事である。この目的のためには、モデル関数中のパラメーターを独立変数として動かすか、あるいは初期値に固定するかが、選択できるようにしてある。さらに、他の実験によるパラメーターの値（例えば文献値）を先験情報として入力するためには、パラメーターの初期値と共に初期値の推定誤差を入力して、初期値のまわりにゆるく束縛する機能がある。このゆるい束縛の機能は、高次のパラメーターを安定に導入・決定するために有効である。

数種のモデル関数のうち最適のものを選び出すための基準として、赤池の情報量基準A I C

$$\begin{aligned} A I C &= -2 \log_e (\text{最大尤度}) + 2 (\text{パラメーター数}) \\ &= (\text{データ数}) \log_e (\text{残差二乗和}) + 2 (\text{パラメーター数}) \end{aligned}$$

が用いられる。SALSが出力するA I Cの値を、数種のモデル関数について比較し、A I Cが最小になるモデルを選べばよい。

また統計学の新しい考え方として、ロバスト推定法というのがあり、SALSにも積極的に組み

入れられている。これは、あてはめべき測定値の中に、一つでもおかしい（かけはなれた値をもつ）ものがあると、最小二乗法がうまく動かなくなることの反省から生まれた考え方である。実際の測定データは、最小二乗法の理論的な前提をいつも満たしているとは限らない。データ解析の初期段階では、測定値にもモデルにも不完全なところが残っていて、最小二乗法の前提を満たしていない場合の方が多いくらいである。このとき、大部分の測定値を信頼し、少数の異常な測定値には目をつぶってあてはめを行なうのが、ロバスト推定法の考え方である。実際には、残差が大きな測定値に対する重みを自動的に小さくして、あてはめをくりかえす。このロバスト推定法は、あらゆる測定値に対して有効であり、異常があると思われる測定値を見つけ出して診断するのによい。測定値やモデルが修正されて、最小二乗法の基本前提が満たされるようになると、ロバスト推定法は実質的に最小二乗法と同等になる。

以上のように、**SALS**は数値解析的にも統計学的にも種々の新しい手法の機能を備えている。より詳しい説明は、「利用の手引き、第2部、制御・解法篇」を参照されたい。

3. **SALS** システムの構成

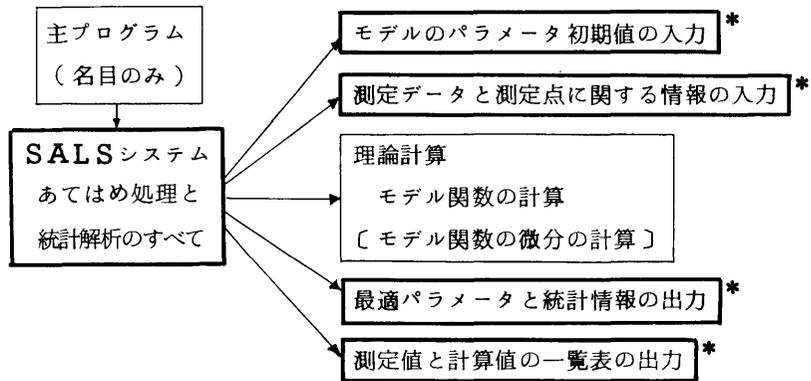
ユーザーが外側から見た時の**SALS**の構成は、図1(a)のようである。すなわち、**SALS**はあてはめ処理と統計解析の一切を行なう1つのサブルーチン(系)であり、ユーザー作成の簡単な(カード4枚の)主プログラムで呼び出す。**SALS**は、パラメーターと測定値に関する入出力サブルーチン群を持っており、それと共に、ユーザー作成のモデル関数の理論計算のサブルーチンを呼んでいる。この図1(a)と、その下の理論計算(シミュレーション)の図(図1(b))とを比べていただきたい。処理内容・構成がほとんど1対1に対応していることがわかる。複雑な非線形モデルを扱うような場合のデータ解析の最大のコツは、理論計算をいろいろ行なって、パラメーターやモデルのふるまいをつかんでから、測定データのあてはめを行なうことである。**SALS**を用いれば、シミュレーションのプログラムを、極めて容易にあてはめのプログラムに変換できるのである。この変換の際に、モデル関数の微分係数(ヤコビアン行列)を計算するプログラムを作ることが望ましいが、前述のように、これを省略して**SALS**内の数値微分で代用することもできる。入出力ルーチン群は、**SALS**中に標準的なきちんとしたものが組み込んであるから、通常はそれを使えばよい。大量の測定データを扱ったり、予め作成したデータファイルを使ったり、複雑な前処理をしたりする必要があれば、ユーザーが独自に入出力ルーチンを作成し、**SALS**のもの置きかえることができる。

SALSシステムの内部構成を、図2に示す。**SALS**は、フォートランカード約14,000枚、サブルーチン約130からでき上がっている。図中のそれぞれのブロックが、実際には数個~20個のサブルーチンから構成されているのである。全体は階層的な構造になっており、特に解法に関するものは、**SALS**管理ルーチン以下、線形解法群まで6層の構造をなしている。非線形解法のうちのどれを使い、線形解法のどれを使うかはユーザーが自由に制御できる。

ユーザーが扱おうとする問題のパラメーターの数や測定値の数は、小さいものもあれば大きいものもある。 **SALS**では、これらの大きさに関する配列はすべて動的に作業領域のわりつけを

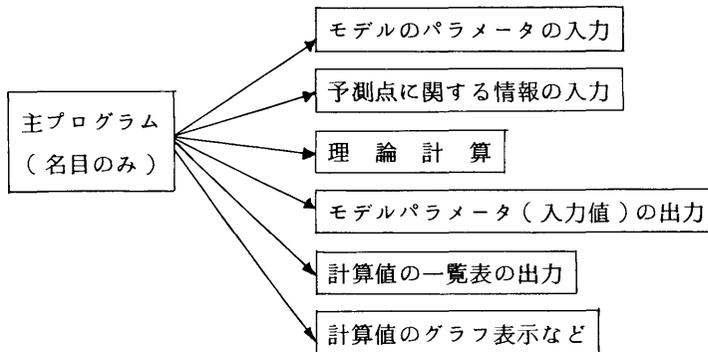
行なう。この結果、パラメーター数個、測定値数十個の小さな問題から、パラメーター数十個、測定値数千個の大きな問題まで柔軟に対応できる。SALSの内部では磁気ディスクなどファイルの操作は一切行っていないので、ユーザーに許されている主記憶（または仮想記憶）の容量が受け入れ可能な問題の大きさを決めることになる。（もちろん、この他に計算精度と計算速度から来る限界がある。）

SALSの実行時には、オーバーレイ構造を用いてリンクしたオブジェクトモジュールを用いることができ、いくつもの解法の選択枝があってもオブジェクトプログラムのサイズを小さく保つようになっている。



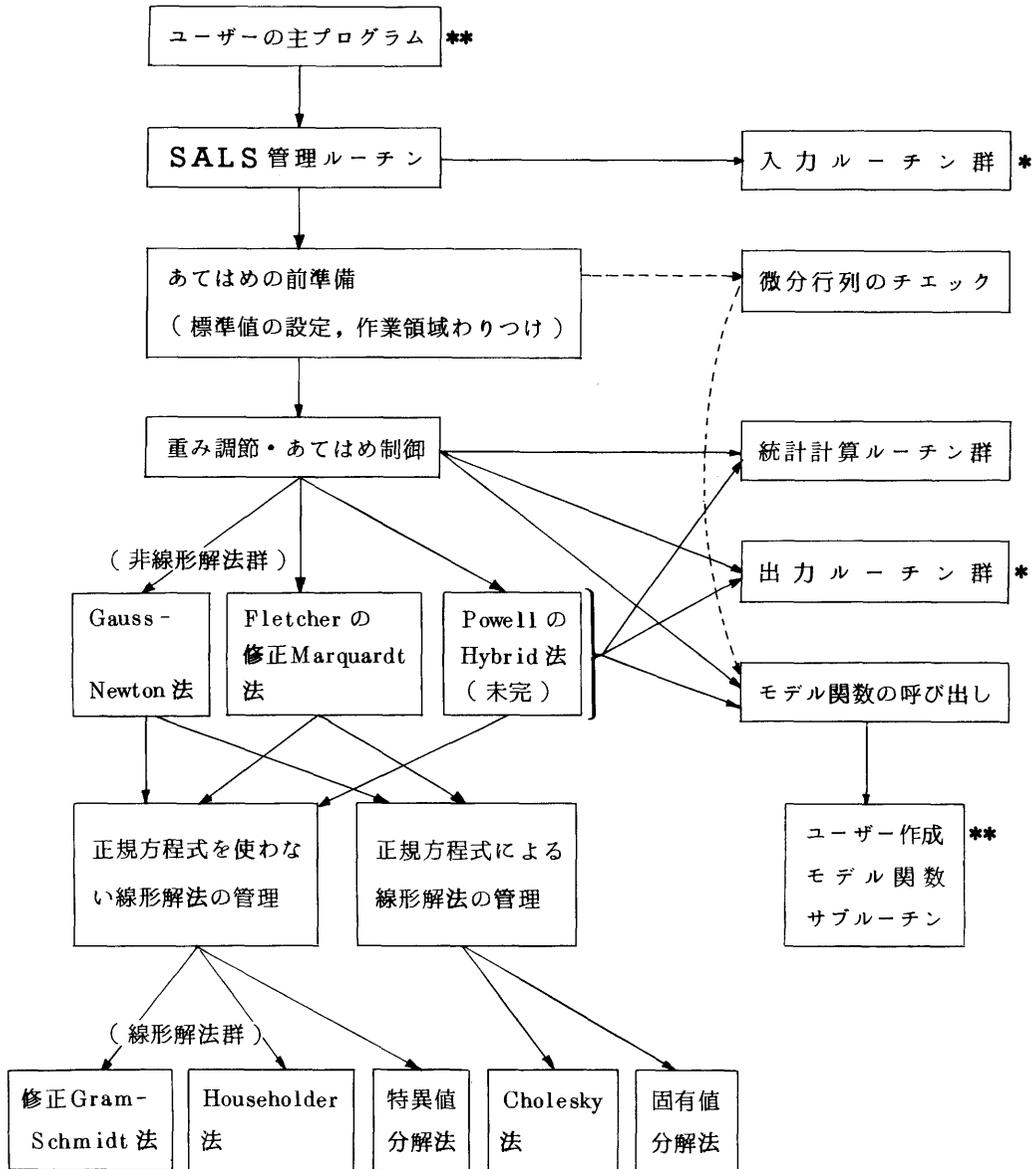
- (注) SALS標準
 * SALS標準またはユーザー作成
 ユーザー作成

図1(a) SALSを用いたあてはめのプログラムの構成



- (注) すべてユーザー作成

図1(b) 理論計算(シミュレーション)のプログラムの構成



(注) *SALS 標準またはユーザー作成

**ユーザー作成

図2. SALS システムの構成

4. SALS の使い方

図3に、指数減衰関数の重ね合わせを使ってあてはめを行なう場合の、SALSの入力ジョブの例を示す。理論式（モデル関数）は

$$f_i = \sum_k a_k \exp(-b_k q_i)$$

である。このような問題は、放射性壊変による放射能の減衰、薬物投与後の新陳代謝による薬物濃度の変化、気相衝突反応による消光過程など、さまざまなところで見られる。

入力デッキは、3枚のジョブコントロールカードに続いて、ユーザーの主プログラム（カード4枚）、ユーザー作成のモデル関数サブルーチン、そして入力データだけからでき上っている。これらは必要不可欠なものばかりであり、最少限の労力でSALSを使えることが了解できよう。

入力データの最初には、PROBLEMコマンド（実際はフォートランの入力データ）を置き、SALSを使う解法や出力を制御する。この例では、(1)モデル関数サブルーチン中でヤコビアン行列を計算していること、(2)Gauss-Newton法で非線形性に対する安定化を強めた解法を使うこと、(3) Gauss-Newton法の反復改良は3回で打ち切りとすること、(4)ロバスト推定法を用いること、(5)出力レベルを第2段階とし、ロバスト推定法によるあてはめ経過と、最終結果とだけを出力すること、(6)統計情報は、最終結果についてだけ詳しく出力すること、を指定している。実はSALSには、解法や出力を制御するコマンドがこの他に20ばかり組みこまれているが、ユーザーが特別な制御をしようとしないう限り、全部省略することが許されている。省略された場合には、PROBLEMコマンドのデータを頂点とする階層構造によって、SALSの標準値がセットされるのである。

次のTITLEコマンドは標題カード2枚を入力する指定であるが、標題入力を省略してもよい。

次にPARAMETERコマンドにより、パラメーターの数を指定し、その数だけのパラメーターの初期値を入力する。線形最小二乗法の場合はパラメーター初期値はすべて0でよいが、非線形の問題では、できるだけよい初期推定値を入力することが大事である。

DATAコマンドは、測定値入力を先導するものであり、測定データの個数（の上限値）、測定点の“横座標”の次元数、重みまたは測定精度の入力のしかた（この例では均一重み）、そして標準誤差（重み1とするものに対する測定精度）などを指定する。ついで、測定データについて、データの名前（番号）、横座標の値、測定値をカードで1つつつ入力させる。

最後のENDSALSがSALSへの入力の終りを示す。

この例では現われないが、SALSでは、解法の組、パラメーターの組、測定値データの組を複数同時に入力することが出来、この場合にはこれらのすべての組み合わせについてあてはめを実行する。このやり方を使うと、パラメーターの初期値をかえたり、パラメーター数を増減したりして、あてはめを試みる事が極めて容易にできる。

SALSは、入力データおよび実行時のエラーチェックをていねいに行なっており、エラーが検出されれば、メッセージを出した上で、適当な処理をして止まるか先に進む。このようなデフォルト機能・エラーチェック機能は、システムとしてのSALSの信頼性をできる限り保証しようとするものである。

```

//A000100 JOB F0001,PASSWORD,REGION=512K.....ジョブカード
// EXEC PORTCG
//FORT.SYSIN DD *
C*** ANALYSIS OF DRUG METABOLIBH USING SALS.
CALL SALS(4000)
STOP
END
C*** MULTIPLE EXPONENTIAL DECAY MODEL.
SUBROUTINE MODEL(LC, NP, ND, NC, X, Q, FOBS,
1 F, RES, DF, LFIX, MDF)
DIMENSION X(NP), Q(ND), FOBS(ND), F(ND), RES(ND),
1 DF(MDF, NP), LFIX(NP)
DO 20 I=1, ND
F(I)=0.0
DO 10 J=1, NP, 2
E=EXP(-X(J+1)*Q(J))
T=X(J)*E
F(I)=F(I)+T
C* CALCULATE THE JACOBIAN COEFFICIENTS, IF LC=1.
IF(LC .LE. 0) GO TO 10
DF(I, J)=E
DF(I, J+1)=-T*Q(I)
10 CONTINUE
20 RES(I)=FOBS(I)-F(I)
RETURN
END
//GO.SYSIN DD * .....以下入力データ
PROBLEM 2. 5. 2. 4. 2. 3. ....解法・出力の制御
TITLE 2.
ANALYSIS OF DRUG METABOLISM.
TORU NAKAGAWA MARCH 27,1979.
PARAMETER 4.
A1 5.0
B1 3.0
A2 1.5
B2 0.1
DATA 12. 1. 1. 0.05
1 0.1 9.17
2 0.2 7.66
12 10.0 0.14
ENDSALS
//

```

ユーザの主プログラム

ユーザ作成モデル関数サブルーチン

$$f_i = \sum_k a_k \exp(-b_k q_i)$$

微分行列の計算

残差の計算

課題の入力

パラメータの初期値の入力

測定値の入力

SALSの入力の終り

図3 SALSの使用例

5. SALS 開発の経過

はじめに述べたように、SALSのグループが発足したのが1975年4月である。その後、7月には東大センターから応用プログラムライブラリーの開発の一環として認められ、計算機使用料や作業室などの援助を受けた。また1976年度から1978年度まで、科研費特定研究「統計プログラムパッケージの研究」(代表者：丘本正阪大教授)にサブグループとして参加し、活動の基盤を得た。

グループ発足後、活発な討論を行ない1975年10月には基本仕様を決定した。その後、設計とコーディングに悪戦苦闘した。1976年12月にSALS-MINI版が始動し、1977年4月SALS-MINI-B版をコーディング、そして1978年3月SALS基本システム第1版を完成し、東大センターで公開使用を開始した。さらに改良を続け、1978年4月にSALS基本システム第2版を完成し、5月より東大センターの正式ライブラリプログラムとして公開使用している。この第2版は、SALSシステムの当初の目標をほぼ完全に達成し、一部には当初目標よりはるかに進んだものもある。

なお、第2版以後、開発を続けているものには、PowellのHybrid法(ヤコビアン行列不要で、速くて安定な非線形解法)、倍精度版SALS、TSS版SALSなどがある。また、第2版として公開したものの性能テストを体系的に行ない、改良する努力を続けている。

この4年余の間のSALS開発グループの主要メンバーは表1のようである。ただし、この他にも多数の人々が開発に寄与されたことを付記し、改めて感謝する。SALSシステムの開発の手法・組織作り・苦労話は東大センタニュース5月号に載せたので、参照されたい。

表1. SALS開発グループ主要構成メンバー

中 川 徹 (東大理)	竹 中 章 郎 (東工大理)
小 柳 義 夫 (高エ研 → 筑大情報)	町 田 勝之輔 (京大薬)
田 辺 國 士 (統数研)	渋谷 政 昭 (日本IBM)
力 久 正 憲 (早大理工)	山 本 毅 雄 (東大センター)
戸 川 隼 人 (日大理工)	鈴 木 功 (東大教計セ → 筑大情報)
伊 藤 徹 三 (理 研)	平 島 守 (日大理工)
上 田 澄 江 (統数研)	松 田 久 夫 (日大理工)
大 岩 元 (東大理 → 豊橋技科大)	

6. SALS の公開と移植

SALS第2版は、5月より東大センターの応用プログラムライブラリーの一つとして公開されている。東大センターのユーザーは誰でも使うことができ、ソースプログラムも公開されている。ユーザーが負うべき義務は、(i)SALSを利用した結果を公表する時には、プログラム名(SALS

第2版)と作成代表者名(中川・小柳)を明記すること、(ii)プログラムをコピーしたり他者に譲渡したりしないことである。

東大センターでは、「SALS利用の手引き*」を発行し、ユーザーのためのSALS講習会を開催、センタニュースに解説記事を連載している。このような努力が実ったためか、SALSの利用回数は急増しつつある。5月は21日以後で116回、6月が508回、7月が687回、8月が20日までで308回であった。10回以上利用したユーザーは20人を越え、そのうち千葉大学薬学部の人、女子栄養大学の化学の人、名古屋大学の地球化学の人などがビッグユーザーなものには、開発者としてもいささか驚いている。

この度、九大センターよりSALSの移植の申し入れを受け、東大センターの公的手続きを経て移植が実現したことは、喜ばしいことである。移植作業そのものは、九大センターのスタッフの人達がすべてやって下さった。同様の移植は大阪大学大型計算機センター、名古屋大学大型計算機センターなどでも進行中である。良いプログラムを作って、みんなが使えるようにしたいというのがSALSを開発した趣旨であり、我々開発グループとしても九大センターのユーザーの皆さんに安心して使ってもらえるように最善の努力をしたいと考えている。11月12日(月)、13日(火)に九大センターでSALSの利用説明会が計画されており、関心のあるユーザーは参加いただきたい。SALSは、ユーザーの発意によって開発されたシステムである。今後もユーザーの皆さんが育ててさらに良くして下さることを願っている。

7. 参 考 文 献

1. あてはめ法 — その理想と現実(I), (II)
中川 徹・朽津耕三：分光研究, **24**(2), 109;(3)165(1975).
2. あてはめ法の新しいアプローチ
中川 徹：ぶんせき, **1979**(4), 84(1979).
3. 最小二乗法標準プログラムSALS(第2版)利用の手引き*, 第1部 基礎篇。(改訂版)
中川 徹・小柳義夫：東大大型計算センター(1979).
4. 同上*, 第2部 制御・解法篇。(改訂版)
中川 徹・小柳義夫・戸川隼人：東大大型計算センター(1979).
5. 最小二乗法標準プログラムの開発(5) SALSシステムの開発をかえりみて
中川 徹・小柳義夫：東大大型計算センターニュース, **11**(5), 63(1979).
6. SALSの使い方
小柳義夫・中川 徹：東大大型計算センターニュース **11**(6), 63, (7)32(1979).

* 東大センター発行の「SALS利用の手引き」のジョブ制御文の部分を九大用に変更した「SALS利用の手引き(九大センター版)」を10月中旬に九大センターから発行する予定である。