

原子準位の計算における non-linear parameter の最適化例

北大、理、佐々木不可止 堀野寧世

§ 1. まえおき

関数 $F(x_1, \dots, x_n)$ の最小値の計算には、大別して二つの型があるように思われる。それらは

i) 最小値を与える x_i の値そのものが計算の主要な目的である場合、

ii) x_i の値そのものは補助的な役割しかない場合、

である。ii) に属するものとして、汎関数

$F(\psi_1, \dots, \psi_n)$ の各 ψ_i を

$$\psi_i(\xi) = \sum_{ij} C_{ij} \varphi_j, x_j(\xi)$$

という x_j というパラメータを持つ関数で展開し、 x_j に関して F を最小化する種の計算がある。この場合には、実際に意味のあるものは ψ_i そのものであって、 x_j の値そのものにはあまり意味はない。

多くの場合、 C_{ij} については $\frac{\partial F}{\partial C_{ij}}$ は比較的容易に解析

的に求めることが出来るので、それらを使って C_{ij} の最適値を、相当精度良く求めることが出来る。原子分子の計算で、Roothaan SCF 方程式といわれるのは、その一例である。

普通、 x_i について $\frac{\partial F}{\partial x_i}$ の表現を求めることは、相当に面倒であって、従って $\frac{\partial F}{\partial x_i}$ を使わず F の値だけから x の最適値を求める必要がある。このような場合、non-linear parameter の最適化という呼び方をしている。以下は Roothaan SCF の例である。この場合には、

$$\psi_i(r) = \sum_j C_{ij} r^{n_j} e^{-\delta_j r}$$

と展開した。三つの n_j と δ_j の組とそれに対する F の値と。以下の表に示す。

n_j	δ_j	δ_j	δ_j
0	3.044	3.044	3.289
0	6.229	6.102	5.979
1	1.2	1.113	1.083
1	3.23	3.124	3.063
1	0.82	0.980	0.977
1	0.3	0.430	0.348
1	0.7	0.459	0.463
1	1.5	1.712	1.740
1	5.0	5.975	6.228

$F = 14.39150 \quad -14.39470 \quad -14.39473$

この例で見られるように、 δ_j の値が相当大きく変っても、 F の値は殆んど変化しない。数値的に二次形式を求めるためには、 δ のステップは

i) F の値が計算精度に比べて充分大きく変化すること

ii) その範囲で二次形式からのずれが充分小さいこと

の二条件が満足されるようにとられなければならない。

この計算誤差の問題について考えてみる。

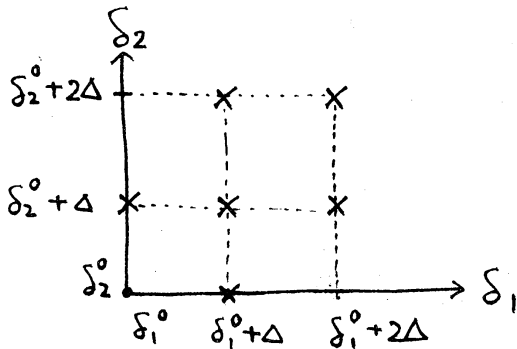
§2. 計算誤差の問題

二変数関数を、下図の×印を付した6点を通る二次形式

$$(1) \quad F(\delta_1, \delta_2) = \sum_{i,j} a_{ij} \delta_i \delta_j + \sum_i b_i \delta_i + C$$

で近似したとき、 Δ によって、 a および 極小値を与える

δ が、どのように変わるかの例を図1に示した。



(δ_1^0, δ_2^0) は最適値を選んだ。

F の値に誤差がなければ

δ, a, b, C は Δ で

展開出来て、 a, b, C

は Δ に比例する項を持ち

δ は Δ について、一次

の項はない。実際に、数

値的に求めれば、 Δ が

ある値よりも大きいときには、大体期待出来るような振舞いを示すが、 Δ が小さいとき、この例では $\Delta \leq 0.03$ 、不規

則になる。

一般に、ある関数 $F(x_1, \dots, x_n)$ を展開すると

$$(2) \quad F(x_1, \dots, x_n) = a + \sum_i^n a_i x_i + \frac{1}{2!} \sum_{i,j}^n a_{ij} x_i x_j + \frac{1}{3!} \sum_{i,j,k}^n a_{ijk} x_i x_j x_k + \dots$$

と書ける。 F の計算誤差を δF とし、各 x について、 Δ 程度の大きさを持つ有限のステップで F の二次形式を求めると、理想的には三次以上の項のない場合の F の値に対して実際に求まる値は、 A_3 を a_{ijk} の程度の量として、

$$(3) \quad \delta F + A_3 \Delta^3$$

の違いが生ずる。従って

$$(4) \quad a \text{ の誤差は } \delta F + A_3 \Delta^3$$

$$(5) \quad a_i \text{ の誤差は } (\delta F + A_3 \Delta^3) / \Delta$$

$$(6) \quad a_{ij} \text{ の誤差は } (\delta F + A_3 \Delta^3) / \Delta^2$$

で見積られる。この例では、 $A_3 \sim 1$, $\delta F \sim 15 \times 10^{-6}$ で見積り、誤差を最小にするステップ、 Δ は

$$\frac{\partial}{\partial \Delta} \left\{ \frac{\delta F + A_3 \Delta^3}{\Delta^2} \right\} \sim -2 \frac{\delta F}{\Delta^3} + A_3 = 0$$

より、 $30 \times 10^{-6} \sim \Delta^3$, $\Delta \sim 3 \times 10^{-2} = 0.03$ となる。実際の計算例とよく一致する。

a_{ij} の入きさを A_2 程度とすれば、

$$\delta F \sim A_2 (\delta x)^2$$

以下の精度で x の値を求めることは無意味であり、従って
この場合では $A_2 \sim 1$ だから.

$$(7) \quad \delta x \sim \sqrt{15 \times 10^{-6}} \sim 4 \times 10^{-3}$$

程度までしか求め得ない.

一方、二次形式として見積った x の値のもつ誤差は、 x の
値を Δ の order とすれば、各 a_i は ΔA_2 の order と
なるから.

$$(8) \quad -a_i = \sum_j a_{ij} x_j$$

とぶくと.

$$(9) \quad \frac{A_1}{A_2} \left\{ \frac{(\delta F + A_3 \Delta^3)/\Delta}{A_1} + \frac{(\delta F + A_3 \Delta^3)/\Delta^2}{A_2} \right\} \\ \sim \frac{\Delta}{A_2} \{ \delta F + A_3 \Delta^3 \} / \Delta^2 \sim \frac{\delta F}{\Delta A_2} + \frac{A_3 \Delta^2}{A_2}$$

(但し、 $A_2 \Delta^2 > \delta F + A_3 \Delta^3$ の場合)

となり、 $\frac{\delta F}{\Delta A_2}$ の項がなければ(計算精度による誤差がなければ)
) Δ^2 に比例することが確かめられる.

このような関数の極値問題を単純に一変数ずつ変化させて
求めると二次までの項であっても有限の回数で収束しない.

二次の関数

$$(10) \quad F = \sum_{ij}^n x_i a_{ij} x_j$$

を考えてみる。 $x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}$ から出発して、順次 x_i に

ついでに $i=1, \dots, n$ の極小値を求め、その値を $x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$ とし、同様に $x^{(k)}$ から $x^{(k+1)}$ を求めると

$$(11) \quad x_i^{(k+1)} = \sum_j^n W_{ij} x_j^{(k)}$$

と書かれる。ここで W は

$$(12) \quad W = \frac{(-)^n}{\prod_i a_{ii}} \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & 0 & \ddots & 1 \\ a_{n,1} & a_{n,2} & \dots & a_{n,n-1} & 0 \end{pmatrix} \begin{pmatrix} 1 & \dots & 1 & 0 \\ & 0 & & 1 \\ a_{n-1,1} & \dots & a_{n-1,n-2} & 0 & a_{n-1,n} \\ 0 & & 0 & 0 & 1 \end{pmatrix} \\ \dots \begin{pmatrix} 0 & a_{1,2} & \dots & a_{1,n} \\ & 1 & & 0 \\ & 0 & \ddots & 1 \end{pmatrix}$$

と書かれる。 $x^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$ を W の固有ベクトル $\mathbb{C}_l$ で展開すれば、固有値と λ_l とすると

$$(13) \quad x^{(k)} = \sum_l (\lambda_l)^k \mathbb{C}_l$$

となり、各 $\mathbb{C}_l$ の展開係数で見れば等比的に変化する。特に二次の場合は

$$(14) \quad \lambda_1 = 0, \quad \lambda_2 = \frac{a_{12}^2}{a_{11} a_{22}}$$

となる。最終的に求める x の精度を $\sqrt{\delta F/A_2}$, 始めに与えられた x の値の精度を Δ とすると、このように一つずつ順次に minimum を求める方法で、minimum に到達するためには、

k は

$$(15) \quad |\lambda_{\max}|^k \Delta \sim \sqrt{\delta F/A_2}$$

$$(16) \quad R \sim \frac{\log(\Delta^2 A_2 / \delta F)}{2|\log \lambda_{\max}|}$$

となる。前の例と同様に $\Delta \sim 0.3$, $A_2 \sim 1$, $\delta F \sim 1.5 \times 10^{-5}$ とおけば $\lambda_{\max}$ に対して

$$R \sim 4 \quad |\lambda_{\max}| \sim 1/3$$

$$(17) \quad R \sim 6 \quad |\lambda_{\max}| \sim 1/2$$

$$R \sim 11 \quad |\lambda_{\max}| \sim 2/3$$

となる。我々の計算例では minimum に達するのに、大体 R が 3~6 cycle となっている。変数の数に対する大体の傾向を以下の表に示す。

変数の数	= 二次形式を求める のに必要な点の数	一つずつ順次、二次形式で 近似する方法で求めた点の数
4	15	26 ~ 31
5	21	49 ~ 60
6	28	38 ~ 49
7	36	49 ~
8	45	55 ~ 71
9	55	22 ~ 64
10	66	89 ~ 133

前述の二次形式で求める方法では、1 cycle で x の誤差は (9) より、 $A_3 \Delta^2 / A_2$ となり、 $\Delta \sim 0.3$ で出発すると、 $A_3 \sim A_2 \sim 1$ とし、 $\Delta^{(1)} \sim 0.1$, $\Delta^{(2)} \sim 0.01$, $\Delta^{(3)} \sim 0.0001$ となる。

最終的な α の精度を(7)により、0.004程度と見積ると、3 cycleを要する。

一次元で順次に探す方法と、二次形式を求める方法と比較すれば、後者は、誤差が一回毎にその2乗に比例して減少し前者は、等比的である。具体例では、前者は、3~6 cycleを要し、後者は3 cycleを要する。従って、後者の2乗に比例して減少するという利点は、それ程明らかでない。1 cycleに要する計算量は、二次形式を求めるためには、 $(n+1)(n+2)/2$ 点必要であるが、順次求めるためには、大略 $2n$ 点のevaluationで済む。従って点の数で比較すれば、 $8n$ と $(n+1)(n+2)$ で $n \geq 5$ の場合には、順次に探す方が有利である。但し、これは rough estimationで得られた結果である。

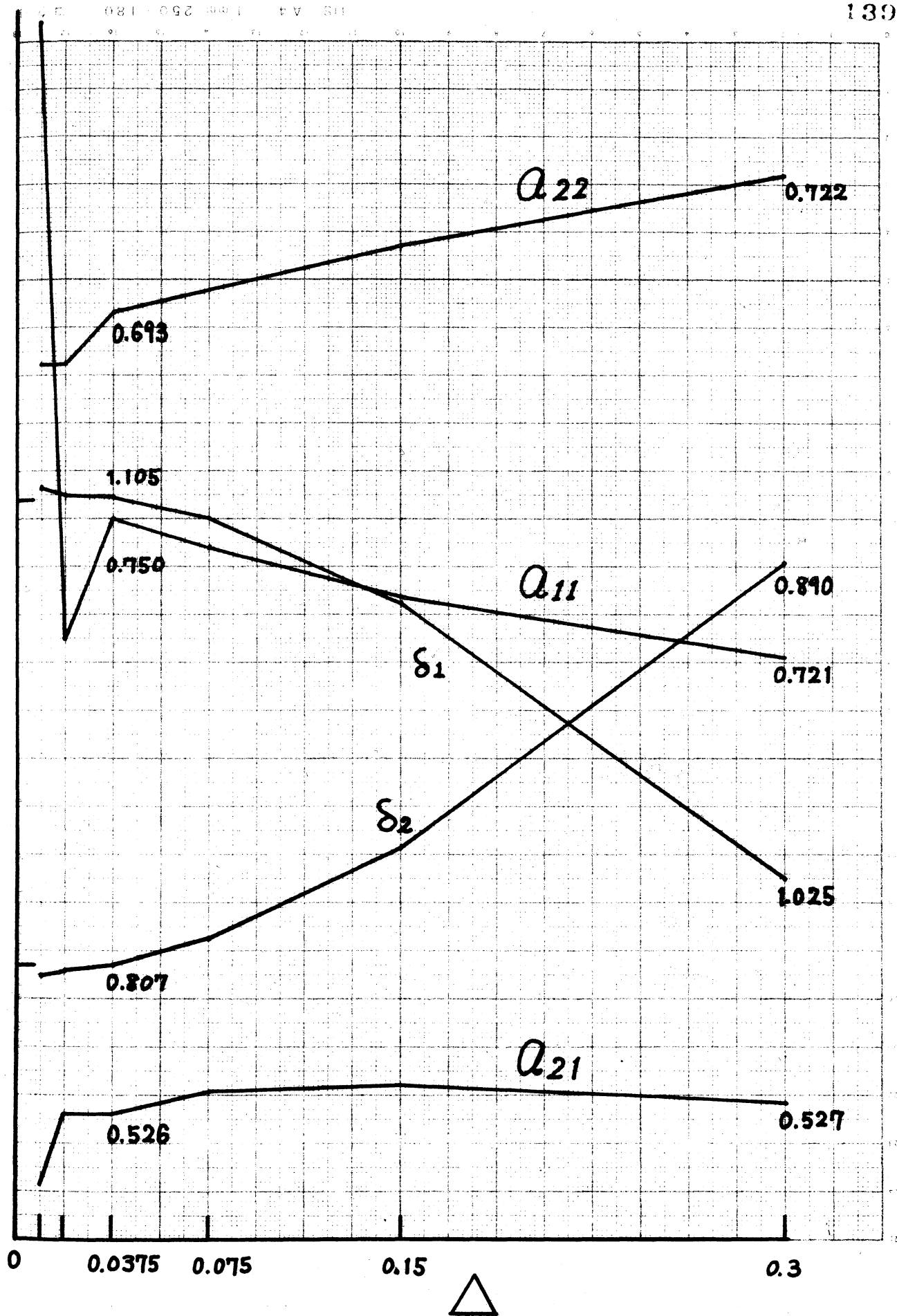


图 1.