計算機を利用した関数の極小化

これは、３年生に対して行った物理実験学の一コマの読みきり講義の講義ノートである。主題は、計算機を用いた関数の極小化を取り上げている。極小化は最小２乗法のみならず、変分法を通じて非常に大きな分野の問題を包含しうるものである。

ここでは、計算機を利用した最小２乗法の数値計算に利用される手法を３種類提示する。最初の物はガウスが提示したとされる観測方程式を解く手法で有り、線形近似である。ゆとりが有れば最小ノルムの最小２乗法についても触れる。後の二つは非線形の極小化の問題を解く手法であり、微分を使用せずにアメーバーの様な振舞いをしながら極小値を探すもので、シンプレックス法と呼ばれる。最後の手法は、非線形極小化を微分を用いて実現するものであり、パラメータ空間の曲率を推定しながら計算を行う。共役勾配法とか可変計量法いう名前で呼ばれる。最後に、時間があれば決定したパラメータの誤差に付いて付け加える。

最後に、対称とする関数を別の関数で adaptive に展開するとアイデアの基本部分を追加した。元もと、この部分は model independent な展開手法として、電子散乱を用いた原子核電荷分布の抽出に使用されたものであるが、これを少し一般化したものである。直交化の手法を、Gram-Schmidt の方法に置き換えた版である。

線形の最小２乗法

独立変数を x、従属変数を y として $y = y(x, a)$ という関数関係を仮定する。

ここに a は関数に含まれるパラメータであり、(x, y) の対データを複数個用意してこのデータを用いてパラメータ（複数も可）を出来るだけ合理的に決めよという問題である。与えられた点の数をｎ、パラメータの数をｍとしておく。

話しを具体的にするために、多項式近似をとる。但し、数値計算上は、多項式近似は誤差が入り込み易いから勧められない。かわりに、直交多項式による展開を勧める。直交多項式には基本漸化式と呼ばれる関係式が有ることをコメントしておく。線形近似では、観測方程式は任意のｉ (i $\leq$ n) に対して以下の様に書ける。

$$y_i = \sum_{j=1}^{j=m} x_i^j \, a_j$$

説明を分かり易くするために、誤差は一定であり、従ってパラメータの決定には関与しないと仮定する。もしも各点に誤差 (ϵ_i) があるならば、次の置き換えをしておけば良い。

$$y_i \to y_i/\epsilon_i, \qquad x_i^j \to x_i^j/\epsilon_i$$

ベクトルを用いて書くと、

$$\vec{y} =_1 X \, \vec{a}$$

ここで、ベクトルと行列の成分は以下の通りである。

$$(\vec{y})_i = y_i, \quad (\vec{a})_i = a_i, \quad X_{ij} = x_i^j$$

最小２乗法は、以下のノルムを最小にすることにより、パラメータａを決める。

$$S \equiv (\vec{y} - X\,\vec{a},\, \vec{y} - X\,\vec{a})$$

ここで、括弧でくくられているのは、ベクトルのスカラー積を表す。Ｓの最小化の為に対称な直交行列Ｐを導入する。

$$S = (\vec{y} - X\vec{a}, P^2(\vec{y} - X\vec{a})) = (P\vec{y} - PX\vec{a}, P\vec{y} - PX\vec{a})$$

行列 PX はｎ行m列であるが、m行m列の正方行列Ｕと（ｎ－m）行m列の行列Ｒを積み重ねた物だと見る。そして、行列Ｐを旨く選ぶとＵは上三角行列になっていて、Ｒは成分が全て０である０行列になっていると仮定する。後で、このようなＰの作り方を示す。$P\vec{y}, P\vec{a}$ も上からm個の成分で作られるベクトル $\vec{y}_1, \vec{a}_1$ と下の（ｎ－m）個の成分の作るベクトル $\vec{y}_2, \vec{a}_2$ に分割する。このとき、Ｓは以下の様に書かれる。

$$S = (\vec{y}_1 - U\vec{a}_1, \vec{y}_1 - U\vec{a}_1) + (\vec{y}_2, \vec{y}_2) \geq (\vec{y}_2, \vec{y}_2)$$

これにより、Ｓの最小値が決定でき、等号が成立する時のパラメータベクトルａが最小２乗法の解を与える。

$$U\vec{a} = \vec{y}_1$$

を解けば良いわけだが、これはＵが上三角行列であるから簡単に逆代入方で解ける。先ず、

$$a_m = (\vec{y}_1)_m / U_{mm}$$

次に、

$$a_{m-1} = \{(\vec{y}_1)_{m-1} - U_{m-1,m}\, a_m\} / U_{m-1,m-1}$$

一般のｉに対しては

$$a_i = \{(y_1)_i - \sum_{j=i+1}^{m} U_{i,j}\, a_j\} / U_{ii}$$

このように、足の大きな方から順番に解が書き下せる。

残差は当然 $(\vec{y}_2, \vec{y}_2)$ である。

残った宿題は、行列Ｘを与えたとき如何にして行列Ｐを決定するかという問題だけである。この問題は段階を追って少しずつ処理する。行列Ｘをmこの列ベクトル $\vec{x}_i, (i = 1, \cdots m)$ を並べて書いたものだと見る。更に次の事実に注目する。任意の単位ベクトル $\vec{u}$ が与えられたとして、次の行列Ｐは対称な直交行列である。

$$P = 1 - 2\vec{u}\,\vec{u}^T$$

ここでベクトルについたTは転置を表す。この行列はelementary orthogonal matrixという。覚えておくと非常に便利である。

PXの第1列のみを取り上げ、この変換されたベクトルは第1成分（λ_1）のみが残り、他の成分は全て0になるようにベクトル$\vec{u}$を決める。

$$P\vec{x}_1 = \vec{x}_1 - 2\vec{u}(\vec{u}, \vec{x}_1) = \lambda_1 \vec{e}_1$$

が成立する。ここで$\vec{e}_1$は第1成分のみが1、他は0という基本単位ベクトルである。両辺のノルムを作ると、$\lambda_1^2 = (\vec{x}_1, \vec{x}_1)$だから、$\lambda_1$の大きさが決定された。符号は後の計算で誤差が小さくなるように決定される。

次に未知数$(\vec{u}, \vec{x}_1)$を決定する作業が残っている。上の式は

$$\vec{u} = \{\vec{x}_1 - \lambda_1 \vec{e}_1\}/2(\vec{u}, \vec{x}_1)$$

を意味するが、ベクトル$\vec{u}$は単位ベクトルでなければならないという条件をここで使用する。その結果、

$$2(\vec{u}, \vec{x}_1)^2 = \lambda_1^2 - \lambda_1 \times (\vec{x}_1)_1$$

右辺の最後の記号はベクトル$\vec{x}_1$の第1成分を指している。ベクトル$\vec{u}$は必ず2次形式で登場するからこの平方根を計算する必要は無いので、符号の選択と言う問題は存在しない。こうして決った$\vec{u}$から作られた行列PをP_1とする。P_1により変換された行列XをX_1と書くとX_1の第1列は第1成分のみが0でない。次にX_1から第1列と第1行を除いて作られる行列を新しいXだと考えて上と同じ操作を行う。

このような操作をm回繰り返して行うと、X_mは上三角行列に変換され、行列の積$P_m P_{m-1} \cdots P_1$は所定の行列Pになっている。行列Xを変換するとき、ベクトル$\vec{y}$も同時に変換しておかねばならないし、行列P_iの左上の(i－1)行(i－1)列の部分は単位行列である。

少しプログラミングの経験を積めば、この理論を用いたプログラムを書いて実用に供することは出来るだろう。メモリーを共有する事や、余分な計算をしないこと、行列Xが特異な時にはどうするかといった点に注意を向けるのが必要である。行列Xの成分をどんな順番で並べるのが精度を保つために必要か？とう事に注意が向けられる様になると、人様にプログラムの話しをしてもいいだろう。

最小2乗法でパラメータを精度良く決定するためには、測定点の精度と点数が多い方が良いのは分かるが、無暗に測定点を増やしても独立性の高いデータでなければパラメータの決定精度は上がらない。測定点を如何に選択すべきかというのは非常に面白い問題であり、有る種の条件付きでは答も提示できるが、最小2乗法とは必ずしも関係無いので次の話題に移ろう。

実験データの点数が沢山が与えられると、フィットを改善しようとして、展開の次数を上げるという欲望が湧いてくる。展開の次数を上げると、展開係数がどんどん大きくなって近似式は(大きな数)ー(大きな数)＝(小さな数)という操作をして実験値を再現する様になる。展開係数のノルムをも考慮するのが良いと思えるだろう。ノルムを最小にすると

いう条件を付けた最小２乗法という概念が出てくる。このときには、一般化（擬）逆行列という概念（道具）を用いる。

参考文献:R. Penrose, Cambridge Phylosophical Society Proceedings, vol 51(1955) 406.

以後の二つの節では独立変数をベクトル $\vec{x}$ としその次数はmとする。この $\vec{x}$ が作る空間を対象とし、値がいつも正の関数 $f(\vec{x})$ を対象とし、この関数値が極小となる $\vec{x}$ を探す問題を扱う。

シンプレックス法

m次元空間に独立な（m＋１）個の点をとりこの点で作られる空間をシンプレックスと呼ぶ。先ず、概念的な説明をしよう。この（m＋１）個の頂点の内で、最大値と最小値を与える点を取り出して最大値の点から最小値の点へむ向けて転がして行こう。うまくいくと、最大値はこれまでの値よりも小さくなるだろう。この操作を繰り返すとそのうちに、極小値はこのシンプレックスの内部に取り込まれるだろう。そうしたら、シンプレックスを小さくする。転がす事と小さくする事を繰り返すと、その内にシンプレックスは極小値のまわりに収縮するだろう。手法を改良するには、転がす時にシンプレックスを大きくしたり小さくしたり歪ませたりすると効率が改善されるだろう。

手法をコワーリック著山本と小山訳の非線形最適化問題（培風館）から再現しておこう。

シンプレックスの頂点 $\vec{x}_i$ での関数値を f_i とし、次の記号を用いる。

最大値を $f_h = f(\vec{x}_h)$

第２番目の最大値 (2nd maximum) を $f_s = f(\vec{x}_s)$

最小値を $f_l = f(\vec{x}_l)$

シンプレックスの図心 $\vec{x}_0 = \sum_i^{n+1}(\vec{x}_i - \vec{x}_h)/n$

次の３個の独立な操作を組み合せ $\vec{x}_h$ を動かし、失敗したら第４の操作をする。

鏡像。最大点 $\vec{x}_h$ から図心 $\vec{x}_0$ に向けたベクトル $\vec{x}_0 - \vec{x}_h$ を作り、この延長上に $\vec{x}_h$ を移す。

$$\vec{x}_r = \vec{x}_0 + \alpha(\vec{x}_0 - \vec{x}_h)$$

拡張。鏡像操作により f_r が小さくなったときには、もっと大きく同じ方向にシンプレックスを動かす事を考える。次の式で与える $\vec{x}_e$ を考える。

$$\vec{x}_e = \vec{x}_0 + \gamma(\vec{x}_r - \vec{x}_0)$$

収縮。鏡像操作が失敗したときには、次の点 $\vec{x}_c$ の値を評価する。

$$\vec{x}_c = \vec{x}_0 + \beta(\vec{x}_h - \vec{x}_0)$$

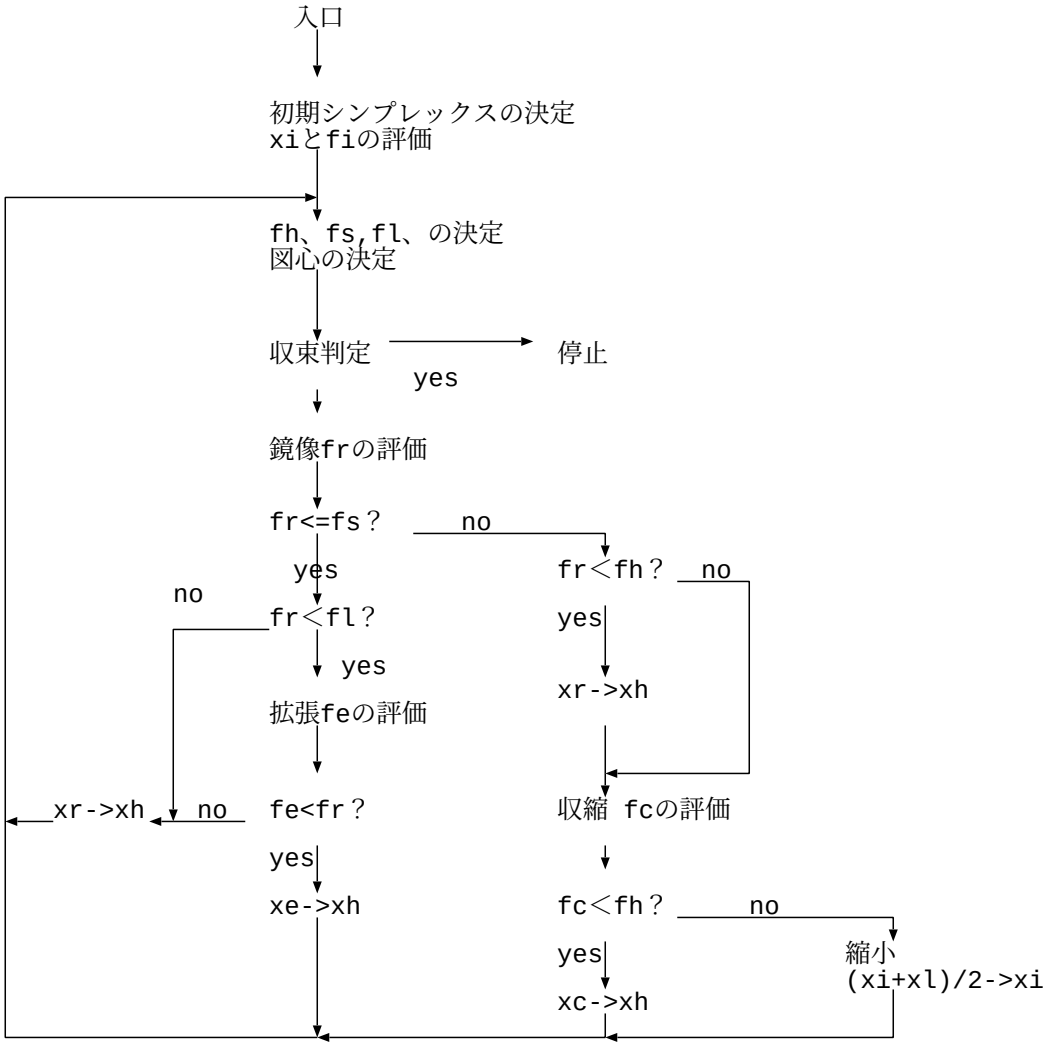
縮小。シンプレックスを最小点方向に向けて半分する。全ての i について、

$$新\vec{x}_i = (旧\vec{x}_i + \vec{x}_l)/2$$

代表的な係数の値は $\alpha = 1, \beta = 1/2, \gamma = 2$ であろう。

以下に参考書のフロウチャートをそのまま引用しておく。

シンプレックス法は効率は悪いが非常に安定であり、プログラムを書くのも易しいので簡単な問題にはお勧めの手法である。微分の計算も必要無い。シンプレックス法はシンプレックスを大きくするのはどちらかと言うと不得手であるが、小さくする方向に働くときには威力を発揮する。最初はシンプレックスを大き目に取っておくと失敗が少ない。



シンプレックス法の
流れ図

共役勾配法

微分 $\nabla f(\vec{x})$ が利用できるならば、極小値は微分が示す逆方向に探して行って、微係数を

0となるようにすれば良いのだから、こんな易しい事は無いと思った人もいるだろう。しかし、実際にやってみると問題は極端に難しく、鳥篭に捕まる（CAGE IN ）とか、悪名高き魔の谷の問題（Notorious deep valley problem ）に出くわしてとまどってしまうのがおちである。

多次元空間での極小化は、出発点を与えたとき、探すべき方向の決定とその方向での極小値の探索という二つの問題に分けることが出来る。後半の問題は直線探索と呼ばれる。他に利用できる情報が無いならば、探索方向の単位ベクトルを $\vec{s}$ とすると、関数値 $f(\vec{x})$ と方向微分 $g = \vec{s} \cdot \nabla f(\vec{x})$ が利用できる訳だから、探査方向に一歩進んでみて出発点での値 f_0, g_0 と幾らか進んだ点での値 f_1, g_1 を３次近似してその極小点を調べるしかないだろう。経験的には、この時深追いをするなという事である。出発値よりも小さな関数値が得られたらさっさとその方向の直線探索は切り上げよという事である。この事は肝に銘じておかないと、猟師森を見ずという事態が生ずる。

方向探索では、変化の急な谷底を目指す気持したいは分かるのだが、深い谷を目がけてスキーで滑べり降りると、対岸の高い岡まで登ってしまい、対岸でもう一度深い谷を目がけて滑べり降りると結局出発点へ戻ってしまうという現象を経験することになる。これが深い谷の問題とが鳥篭に閉じ込められるという表現の直感的な説明である。

直線探索が終了した後の探索方向の決定に大切なのは、現在又は過去に調べた方向とはある意味で直交した方向を探索方向とするという配慮である。直感的に言って、m次元パラメータ空間の直交したm個の方向に直線探索を１回ずつ完成させれば、極小値は決定できる。但し、直線探索を一つの方向だけに拘っていると、崖に沿って直線探索している可能性もあるので、先に肝に銘じたようにさっさと方向は変えるのである。その結果、方向探索はm回では済まなくなるが、最終的な効率は上がる事がしばしば経験されるという訳である。

共役勾配法は Hestenes と Steifel のいわゆる世紀の大発見を Davidon が極小化に応用したものであるが、Davidon の論文は難しいので Fletcher の解説を皆が引用する様である。

参考文献: R. Fletcher, The Computer Journal 13(1970) 317

簡単な解説

最小化すべき関数 $f(\vec{x})$ が、$\vec{x}$ の２次形式で書かれているという単純な場合を先ず想定する。

$$f(\vec{x}) = (\vec{x}, G\,\vec{x})/2 + (\vec{g}, \vec{x}) + c$$

この場合、最小値を与えるのは $\vec{x}_m = G^{-1}\vec{g}$ と与えられる事は自明である。即ち、目的関数 $f(\vec{x})$ の勾配 $\vec{g} = \nabla f(\vec{x})$ 及び、ヘッシアンと呼ばれる行列 G ($G_{ij} \equiv \dfrac{\partial^2 f}{\partial x_i\, \partial x_j}$ の逆行列 H が分かれば極小値を簡単に計算出来ると言う事が理解できよう。

行列 G 及び二つのベクトル $\vec{a}, \vec{b}$ が与えられた時、以下の関係が成立するならばベクトル $\vec{a}, \vec{b}$ は共役であるという。

$$(\vec{a}, G\vec{b}) = 0$$

ベクトル $\vec{a}, \vec{b}$ は直交はしていないが、有る意味で直交、従って独立である。このお互いに

共役なm個の方向を極小値を探す方向として利用する事ができる。

共役勾配法は、非常に完成された方法であって、現在及び過去の情報を利用してGの逆行列を推定し、新しく共役な方向をつくり出す。更に、有る程度逆行列が推定された後では、直線探索を一回の計算で済ませてしまう。

最初はHの近似値は単位行列としておく。出発点$\vec{x}$と直線探索の結果の到達点$\vec{x}^*$での勾配を$\vec{g}, \vec{g}^*$とし、次のベクトルを定義する。

$$\vec{delta} = \vec{x}^* - \vec{x}, \qquad \vec{\gamma} = \vec{g}^* - \vec{g}$$

行列Hの改良公式は以下の通りである。

$$H^* = H + \frac{\vec{\delta}\vec{\delta}^T}{(\vec{\delta}, \vec{\gamma})} - \frac{H\vec{\gamma}\vec{\gamma}^T H}{(\vec{\gamma}, H\vec{\gamma})} + \phi\vec{\nu}\vec{\nu}^T$$

ここで、ベクトル$\vec{\nu}$は次式で定義される。

$$\vec{\nu} = \sqrt{(\vec{\gamma}, H\vec{\gamma})}\{\frac{\vec{\delta}}{(\vec{\delta}, \vec{\gamma})} - \frac{H\vec{\gamma}}{(\vec{\gamma}, H\vec{\gamma})}\}$$

最後にϕを以下に選ぶかという基準を定義する。ベクトル$\vec{\nu}$の分母に登場した二つの数字をａ，ｂとする。

$$a = (\vec{\delta}, \vec{\gamma}), \quad b = (\vec{\gamma}, H\vec{\gamma})$$

もしも$a > b$ならば$\phi = 1$、逆ならば$\phi = 0$とする。始めの方で述べたが、探索の初期には直線探索は必要であるが、その内に直線探索は必要無くなる。その内とは、直線探索がすべての共役方向に対して一巡したころである。

誤差について

誤差という言葉を使う時にはかなり慎重にならなばならない。直接測定の実験誤差という概念は安心して、信用出来る。しかし、間接測定の誤差には直接の測定量と間接に導かれる物理量を繋ぐ関係式が介在する。この関係式定義式ならば信用するしかないが、そうでない場合にはその関係式自体が含んでいる誤差をどのように評価するかという問題を十分に検討するべきである。

ここでは、上に述べた基本的な問題を無視して、最小２乗法そのものがどの程度信頼できるパラメータを与えるかという小さな問題に限定して考えてみよう。もう一度言っておくと、最小２乗法で決められた、最小化すべき目的関数の大小は、定性的傾向としては小さい方がより現実的らしいけれども、この目的関数の大小と、現実が起こり得る確率とがどのような関数関係にあるかという事は、かなり大胆な仮定を設けない限り分からないので、不問に付す。

問題は以下の様に設定する。何らかの方法で決定されたパラメータは、お互いに相関があるだろうが、どの程度変化させると目的関数の値が、極小値の２倍(５０％増し)になるか？もっと抽象的には、パラメータ空間の極小点での曲率はいくらか？

こう言った問いにならば自信を持って答える事が出来る。パラメータを変化させた時に、目的関数がどの程度変化するかを調べれば良い。それには、目的関数を極小点でテイラー展開すれば良い。極小点では、勾配が０であるから１次の展開係数は０であるので、

$$f(\vec{x}_m + \delta\vec{x}) = f(\vec{x}_m) + (\delta\vec{x}, G\delta\vec{x})/2 + cdots$$

ここで、

$$(G)_{ij} = \frac{\partial^2}{\partial x_i \partial x_j} f(\vec{x}_m)$$

これにより、誤差を論ずる事はこの行列の性質を論ずる事である事が分かった。共役勾配法では先に説明したように、副産物として G の逆行列が計算されている。以下の議論では G の議論をするが、逆行列の固有値は元の行列の固有値の逆数であり、固有ベクトルは同じとする事が出来るので、逆行列の計算をする必要は無い。他の方法で極小値を評価したときには、適当な方法でこの行列 G を計算しなければならない。

先ず注意すべきは、たとえ直交関数系を用いて展開したとしても、パラメータには相関があるという事である。形式的な言葉で言えば、G は非対角要素が０では無い。即ち、同時の複数のパラメータを変化させて、しかも目的関数の値は変化しないという状況を作り得る。パラメータの独立性と直交性とをはっきり区別せねばならない。直交したパラメータならば一方の変化に他方の変化が影響を与える事は無い。これは、行列 G を対角化する事により実現される。形式的には、直交行列 V を持って来て行列 G を対角化する。

$$G\vec{v}_i = \lambda_i \vec{v}_i, (i = 1, \cdots m)$$

ここで、行列 V と Λ は固有ベクトルと固有値を対角型に並べて出来た行列だとすると、以下の様な関係式を書く事が出来る。

$$f(\vec{x}_m + \delta\vec{x}) - f(\vec{x}_m) = (\delta\vec{x}, G\delta\vec{x})/2 + \cdots = (\delta\vec{y}, \Lambda\delta\vec{y}) + \cdots = \sum_i \lambda_i \delta y_i^2 + \cdots$$

最後の式を見ると目的関数の増分は、交差項が無いから、完全に独立な項の和で書けている。お互いに影響しあう事は無い。δy_i は企画化された固有ベクトル $\vec{v}_i$ 方向にパラメータを動かす時の増分である。

この次の議論が難しい。誤差は、上の意味で独立な方向にパラメータを動かした時、目的関数が $f(\vec{x}_m)$ がどれくらい増加するかを言えば良いのだが、この意味での増加量を習慣的にいくらととっているか知らない。ガウス分布を仮定するならば、簡単に決定できるだろう。例えば、目的関数が $f(\vec{x}_m)/m$ 増える部分を独立なパラメータの組で分担すると仮定するならば、$\delta y_i = \sqrt{2f(\vec{x}_m)/(m\lambda_i)}$ と決まる。元の x というパラメータに戻すと、

$$\delta x_i = \delta y_i \vec{v}_i$$

増分をいくらに取るかと言う瑣末な議論は物理的にはあまり重要ではないから、固有ベクトルの方向と、固有値 (従ってパラメータ空間の曲率) をそのまま誤差だと表現しておくほうがよほどすっきりする。

adaptive な手法での関数の展開

ある種の極小化問題では、展開すべき関数 O(x) があり、これを別の関数系 $b_i(x),\quad (i = 1, \cdots N)$ で展開する。一般的に言えば、$b_i(x)$ は直交関数系であることが望ましいが、最初から直交関数を与える事は面倒であるので、計算機に直交化の手続きを教えて、事前に直交化してから処理をする。

しかし直交関数系は無限にあるという事実もある。即ち、無限の可能性の中からどの直交関数系を採用するかという問題は、考えてみる価値がありそうに思える。この問題は、直交関数系という言葉を、基底として採用する関数系 $b_i(x)$ と置き換えても、ほとんど同様に意味を持つ問であろう。

一方、対象とする関数 O(x) に関する情報は、一般的に言って充分に与えられている訳ではないから、$b_i(x)$ としては、最初は常識的なものを使い、問題を解いて行く段階で少しずつ改良して行くのが妥当な手段だと思われる。以下の記録するのは、この改良をある程度自動的に行う方法に対する一つの提案である。これまでの常識では、展開の精度を上げるには展開の項数を増やせという事であったが、項数を増やすだけが能ではないというのが、この文の趣旨である。

$b_i(x)$ の直交化を如何に行うかで、アルゴリズムはある程度異なるが、ここでは、Gram-Schmidt の手法を使う。最初に Gram-Schmidt の手法を、記号を定義する目的を込めて、復習する。

与件： 非直交関数系 $b_i(x), (i = 1 \cdots N)$ が与えられたとする。ここで、正の荷重関数 w(x) を用いた以下の積分が計算可能であるとする。

$$(b_i, b_j) \equiv \int dx\, b_i(x) b_j(x) w(x)$$

但し、積分範囲は問題に依存し確定しているものとする。

この関数系の線形結合として、以下の条件を満足する直交関数系を作れ。

$$c_i(x) = \sum_{j=1}^{i} B^j b_j(x); \quad (c_i, c_j) \equiv \int dx\, c_i(x) c_j(x) w(x) = \delta_{i,j}$$

Gram-Schmidt の手法は、以下の通りである。

0 先ず、i=1 として計算を開始する。

1 $\hat{b}_i(x) = b_i(x) - \sum_{j=1}^{j=i-1} c_j(x)(c_j, b_i)$ を計算する。但し、右辺の計算が j=0 のみならば、この内積の計算は行わない。

2 $c_i(x) = N_i \hat{b}_i(x), \quad N_i = (b_i, b_i)^{-1/2}$ により、$c_i(x)$ とする。

3 i を一つ増やし、$i \leq N$ ならば、[1] から作業を再開し、そうでなければ、作業を停止する。

パラメータを含んだ基底 $b_i(x,p)$ に、問題を拡張することを考える。パラメータは複数個あっても良いが、簡単の為に一つであるとする。

問題を以下の様に定める。直交関数系 $c_i(x)$ と同時に、$\dfrac{\partial c_i}{\partial p}$ も同時に計算するように、アルゴリズムを拡張せよ。当然の事ながら、$\dfrac{\partial b_i(x)}{\partial p}$ や、これが関係する内積は計算可能であるとする。

以下の様にするのが、一つの解法である。

0 i=0 として、作業を開始する。

1 $\hat{b}_i(x) = b_i(x) - \sum_{j=1}^{j=i-1} c_j(x)(c_j, b_i)$ を計算する。右辺の和に関しては、j=0 だけならば、この項は無視する。

2 内積 $(\hat{b}_i, \hat{b}_i)$ 及び、$\left(\dfrac{\partial \hat{b}_i}{\partial p}, \hat{b}_i\right)$ を計算する。

$$\frac{\partial \hat{b}_i}{\partial p} = \frac{\partial b_i}{\partial p} - \sum_j \frac{\partial c_j}{\partial p}(c_j, b_i) - \sum_j c_j\{(\frac{\partial c_j}{\partial p}, b_i) + (c_j, \frac{\partial b_i}{\partial p})\}$$

3 $N_i = (\hat{b}_i, \hat{b}_i)^{-1/2}$ とおくと、

$$c_i(x) = N_i\,\hat{b}_i(x), \qquad \frac{\partial c_i(x)}{\partial p} = -N_i^3\left(\frac{\partial \hat{b}_i}{\partial p}, \hat{b}_i\right)\hat{b}_i(x) + N_i\frac{\partial \hat{b}_i}{\partial p}$$

4 i を一つ増やし、N 以上ならば作業停止、さもなくば [1] へ戻る。

この知識があると、パラメータ p を一つに固定して 展開係数 B^i を決定した後で、例えば χ^2 の様な展開係数の決定指標 (これを罰金関数と呼ぶ) をパラメータ p の関数と考えて微分する可能性がある。即ち、もう一段進んだ展開が、展開の基底を変更する事により可能である事が分かる。

勿論この知識は、展開係数を決定する途中で使用する事も可能である。罰金関数の展開係数による微分以外に、パラメータによる微分も並列的に計算するならば、更に効率の良いアルゴリズムが可能であることは、自明であろう。

この手法は、罰金関数を極小化する手法として、共役勾配法と組み合わせると、威力を更に発揮できる。ヘッシアンを計算しているはずだから、誤差行列を引き出せるから、展開係数の誤差や相関関係を簡単に評価する事が可能になる。

上に書いた文だけでは、抽象的すぎてピンと来ない人もいるだろうから、少し解説を兼ねた蛇足を加えておこう。

例を長い橋を作るときの橋脚にとろう。橋脚の上には、橋桁その他の構造物が載り、更に上を交通手段としての列車や車が通過する。橋に屋根を架けたり、櫓を組む人達もい

る。動荷重の問題は、複雑にするだけで、ここでの例としての適切性を欠くだろうから無視しよう。

一つの橋脚の強度は一定だと仮定したとき、幾らの橋脚を必要とするか、橋脚のスパンはどのように分布するのが良いかという事は、設計上当然問題とされるべきだろう。橋をある関数 O(x) に対応させ、橋脚の数を展開する次数 N に対応させたとすると、橋脚の位置は、直交系のゼロ点の分布に対応させて考えるのが良いだろう。荷重分布に付いて何も考えないならば、両端で考えるべき事が少しあるが、橋脚は均等に分布させるのが妥当だろう。これは、上の例では橋の両端を変数 x の定義域と考えて、$b_i(x)$ としてこの区間での Legendre 関数を使用する事に対応する。しかし、橋の上の荷重に付いての知識があれば、等間隔というのは最良の判断では無いことは、自明である。あらかじめ橋の全体像が掴めていれば、これを設計に際して考慮すれば良い。一方、実際の問題を解くときには全体像は不明である場合が多い。不明だから問題を解く必要がある！解の性質がはっきりとは分からない時に、問題を解くと言うのが数値計算の困難な点の一つである。そこで、対策として解きながら考えるという事がある。問題をある程度の精度で一度解いてみて、その結果を見て次の手を打つ。これを自動的にする一つのやりかたは、先ず適当と予想した関数系で展開を試みる。その過程で、またはその結果、展開の関数系を改良するならば、最初固定した関数空間から、次数を大きく増やさずに、飛び出す事が出来る。どちらかと言うと、アメーバが自分の行きたいところを探すような、adaptive な方法を与える事になる手法の原型をここで与えた事になる。