

# 非線形計画法 (3)

## —無制約最適化問題—

八巻 直一, 矢部 博

### 1. はじめに

今月は、いよいよ非線形計画法のアルゴリズムのお話に入りましょう。ここで、もう一度非線形計画問題を確認しておきます。

非線形計画法とは、以下のような問題の構成と、構成された問題を数値的に解く手法のことを指します。

#### 非線形計画問題

$n$  変数の非線形関数  $f(x)$  (目的関数) を、等号制約条件

$$h_i(x) = 0, \quad i = 1, \dots, m$$

と不等号制約条件

$$g_j(x) \leq 0, \quad j = 1, \dots, l$$

のもとで、最小化せよ。

上の問題で、制約条件がない場合は、無制約最適化問題と、また制約条件のある場合を、制約条件付き最適化問題といいます。無制約最適化問題と、制約条件付き最適化問題とは、最適解の持つべき条件 (最適性条件といいます) や数値解法に、本質的な違いがありますので、本連載ではそれらを区別して、今回と次回に分けて解説します。

### 2. 無制約最適化法

本章では、制約条件の無い問題 (無制約最適化問題) の数値解法を取り扱います。最適化問題には、

目的関数を最小化するタイプと、目的関数を最大化するタイプがありますが、これらは本質的には同じといえます。何故ならば、最大化問題で目的関数の符号を変えれば、問題の本質を変える事なく、最小化問題となるからです。したがって、最適化問題は、最小化問題として定義することができます。

以上より、ここで取り扱う問題は、次のように表すことができます。

#### 無制約最小化問題

$n$  変数の目的関数  $f(x)$  を最小にする、 $x_* \in R^n$  を見つけよ。

以下、問題といえば、上の無制約最小化問題を意味することとします。

さて、問題の最適解  $x_*$  の満たすべき条件は、明かに、

$$f(x_*) \leq f(x) \quad (1)$$

です。ただし  $x$  は任意の点です。このような  $x_*$  を大域的最小点といいます。また、 $x_*$  の適当な近傍内の任意の点で (1) が満たされるとき、局所的最小点といいます。

最適解の満たすべき条件を、最適性条件といいます。上の最適性条件は、このままでは、実用的ではありませんが、 $f$  が微分可能であるか、形状がサラダボウルのように丸く下に凸であるなどの、滑らかさや形状の情報によって、上の条件はいろいろに書き換えることができます。

もし、目的関数になんの情報もないときは、連続性も凸性も仮定できませんので、なんらかのアルゴリズムによって得られる数値解が、(1) を満たす点に近いかどうかを推定するしかありません。現実には遭遇する問題では、目的関数に凸性や微分可能性を仮定できな

やまき なおかず システム計画研究所

〒 150 渋谷区桜丘町 2-9 カサヤビル

やべ ひろし 東京理科大学工学部

〒 162 新宿区神楽坂 1-3

い場合がよくあります。そのような問題では、本当の最適解を求めることは、大変に難しいといえます。そんな場合では、できるだけ最適解に近い点を、いかに効率よく求めるかという戦略が採られることになります。最近話題の、遺伝的アルゴリズムや、シミュレーテッドアニーリング法などは、このような戦略に基づくものといえましょう。

### 3. 微分を用いない方法

微分を用いない手法は、直接探索法と呼ばれています。直接探索法では、目的関数の値を計算することだけが許されるので、微分の計算は必要ありませんが、収束性にやや難点があるといわれています。しかし、現実問題では、目的関数の微分が容易には得られないことは珍しくありません。したがって、直接探索法は重要な手法であるといえましょう。

ここでは、ランダム法とシプレックス法をご紹介します。

#### 3.1 ランダム法

ランダム法は、非線形最適化法の歴史では新しい方法ではありませんが、最近話題のシミュレーテッドアニーリング法や、遺伝的アルゴリズムなどとも大いに関係があると考えられます。

ランダム法は、定められた  $n$  次元超立方体の内部で、目的関数  $f(x)$  の最小値を探索する問題に対して定義されます。

ランダム法では、超立方体の中にランダムに多くの点を取り、その各点での目的関数の値を比較することによって、最小点を探索します。もちろん、単純にデータな点をとるのでは効率的ではありませんので、さまざまな工夫を加えます。ここでは詳細を述べる余裕はありませんが、確率的な要素をアルゴリズムの中に持つ方法は、極小点がいくつか存在する問題や、目的関数の滑らかさが保証されない問題では、おおいに力を発揮することが知られています。

#### 3.2 シプレックス法

シプレックスとは、 $n$  次元空間における  $n+1$  個の点の集合のことです。3次元空間ならば、四面体の頂点が形成されることとなります。シプレックス法は、シプレックスの各端点での目的関数値を比較して、反転、縮小、拡大の三つの操作を繰り返しながら最小点を探索する手法です。正多面体が坂道を転げ落

ちる様子を想像していただければ、その移動の原理が直観的に理解できるでしょう。

シプレックス法は、適用範囲が広く、アルゴリズムが比較的簡単なので、目的関数値のみを用いる手法としては、大変実用的なものとして使われています。

### 4. 微分を用いる方法

本章では、関数の微分情報を用いるアルゴリズムについて述べます。

以下では、目的関数  $f(x)$  は十分に滑らかであるとし、必要な回数だけ微分できるものと仮定します。目的関数が微分可能であれば、 $x_*$  が局所的な最小点であるとき、

$$\nabla f(x_*) = 0 \quad (2)$$

が満足されます。ここで、 $\nabla f(x)$  は、 $f(x)$  を  $x$  の各成分で偏微分した値で作られるベクトルです。

さらに、 $f(x)$  が凸ですと、 $x_*$  は大域的な最小点となります。したがって、(2) を満たす点を発見することを目指して、アルゴリズムが構成されます。しかし一般的には、大域的な最適解を見つけることはなかなか難しく、現実的には局所的な最適解で妥協することがむしろ普通です。

ここで扱う解法は、反復法に基づくもので、次のような手順で記述されます。

#### 反復法のプロトタイプ

**Step 0.** 初期点  $x_0$  と行列  $B_0$  を与える。  $k=0$  とおく。  
( $B_k$  については、後で説明します。)

**Step 1.** 停止条件が満たされていれば、 $x_k$  を解の近似点として停止する。さもなければ、Step 2 へ行く。

**Step 2.** 探索方向  $d_k$  を次の方程式の解として決定する。

$$B_k d = -\nabla f(x_k)$$

**Step 3.**  $d_k$  方向での刻み幅  $\alpha_k$  を求める (直線探索)。

**Step 4.**  $x_{k+1} = x_k + \alpha_k d_k$  とおく。

**Step 5.**  $B_{k+1}$  を生成し、 $k+1$  を  $k$  とおいて Step 1 へ行く。

よく用いられる停止条件は、勾配ベクトルの大きさ  $\|\nabla f(x_{k+1})\|$  や点列  $\{x_k\}$  の変動  $\|x_{k+1} - x_k\|$  を拠り所としています。すなわち、上のどちらかの値があらかじめ定めた許容値より小さくなったら、解に収束したとみなすことにします。ただし、ベクトル  $a = (a_1, \dots, a_n)^T$  について、 $\|a\|$  は  $l_2$  ノルムといい

$$\|a\| = \sqrt{\sum_{i=1}^n a_i^2}$$

で計算されます。

アルゴリズム中の  $B_k$  は、探索方向ベクトルを求めるための方程式の係数行列で、アルゴリズムの性格を本質的に決定します。

反復法のプロトタイプは、目的関数が下がる探索方向（降下方向）をうまく見つけて、その方向で刻み幅を決定するもので、直線探索 (line search)、もしくは1次元探索 (linear search) と呼ばれる手順を利用する方法です。他方、近年、信頼領域法 (trust region method) と呼ばれる、別のカテゴリの手法が提案されて注目されています。信頼領域法は、現在の解の近似点の周りに適当な領域を作って、その領域の中で目的関数がある程度下げる点を次の点とする方法で、この”適当な領域”が信頼領域と呼ばれます。

#### 4.1 最急降下法

アルゴリズムの性質は、探索方向の決定に用いる  $B_k$  によって、特徴づけられます。目的関数の勾配ベクトルは関数を局所的に最大にする方向を示すので、この方向の反対に進めば、最大の傾斜に沿って  $f(x)$  の値を下げることができるでしょう。これが最急降下法 (steepest descent method) といわれる方法です。このとき、 $B_k = I$  すなわち単位行列となり、 $k$  回目の探索ベクトル  $d_k$  は次式で与えられます。

$$d_k = -\nabla f(x_k) \quad (3)$$

しかしながら、この探索方向は大域的な解を探索するときの最適な探索戦略とは限らず、しばしば途中で停滞してしまうことが知られています。最急降下法は、直線探索を上手にやれば、初期点の選び方によらず  $x_*$  に収束することが保証されますが、反面、ほとんど実用に耐えられないほどに収束が遅い場合があるという欠点をもっています。

#### 4.2 ニュートン法

よく知られているニュートン法は、目的関数を点  $x_k$  で2次近似したモデル関数

$$f(x_k + d) \approx f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla^2 f(x_k) d$$

の最小点に向かう方向を、探索ベクトル  $d_k$  に選びます。ただし、 $\nabla^2 f(x)$  は  $f(x)$  のヘッセ行列を意味します。上の近似式から、ニュートン法で選ばれる探索方向は、連立1次方程式

$$\nabla^2 f(x_k) d = -\nabla f(x_k) \quad (4)$$

を解くことによって求まることがわかります。すなわち、 $B_k = \nabla^2 f(x_k)$  となります。

ニュートン法は収束が速いという長所をもつ一方で、関数のヘッセ行列が必要であり、利用者の大きな労力が要求されるという欠点をもっています。また、初期点をうまく選ばないと、収束が保証できないのも短所といえます。

#### 4.3 準ニュートン法

無制約最小化のアルゴリズムの例として、最急降下法とニュートン法を紹介しました。その各々の長所、欠点をあげました。ここでは、最急降下法とニュートン法の長所を継承し、欠点を修正した方法として、準ニュートン法を紹介しましょう。

準ニュートン法 (quasi-Newton method) は、Davidon(1959), Fletcher and Powell(1963) らが、また非線形方程式に対するニュートン法と関連して Broyden(1965) が、それぞれ開発したアルゴリズムで、セカント法 (secant method) とか可変計量法 (variable metric method) とも呼ばれています。

準ニュートン法は、ニュートン法の探索ベクトル(4)におけるヘッセ行列（あるいはその逆行列）の計算の手間を緩和した解法であり、無制約最小化アルゴリズム中で、ヘッセ行列  $\nabla^2 f(x_k)$  を行列  $B_k$  で近似する方法と、ヘッセ行列の逆行列  $\nabla^2 f(x_k)^{-1}$  を行列  $H_k$  で近似する方法とに分けられます。

その際に、 $B_{k+1}(H_{k+1})$  に対して課される条件として、セカント条件（あるいは、準ニュートン条件とも呼ばれる）があります。これは、 $\nabla f(x)$  の点  $x_k$  のまわりの1次の Taylor 展開

$$\nabla f(x_{k+1}) - \nabla f(x_k) \approx \nabla^2 f(x_{k+1})(x_{k+1} - x_k)$$

に基づくもので、 $B_{k+1}(H_{k+1})$  について、

$$B_{k+1}s_k = y_k \quad (H_{k+1}y_k = s_k)$$

が課されます。ただし、

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

とします。

一般に、上の条件式を満たす  $B_{k+1}(H_{k+1})$  は一意に決まらないので、さまざまな更新公式が提案されています。

以下では、 $B_k$  に関する更新公式を  $B$  公式、 $H_k$  に関する更新公式を  $H$  公式と呼びます。

$H$  公式を最初に研究したのは、Davidon, Fletcher と Powell です。この公式は 3 人の頭文字をとって DFP 公式と呼ばれ、

$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{s_k^T y_k} \quad (5)$$

で与えられます。

準ニュートン法の創世期における DFP 公式の登場は画期的であり、当時は、Fletcher-Powell法とか DFP 法とか呼ばれたりもしました。その後、DFP 公式と同じ様な性質をもつ公式が研究され始め、なかでも、Broyden, Fletcher, Goldfarb, Shanno らがそれぞれ別々に提案した公式が有名で、4 人の頭文字をとって BFGS 公式と呼ばれます。BFGS 公式は、次のようなものです。

$$H_{k+1} = H_k + \left(1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} \quad (6)$$

BFGS 公式と DFP 公式との間には、非常に興味深い関係があります。DFP 公式で  $B = H$  とおき、 $s$  と  $y$  を交換すれば、以下に示す  $B$  公式の BFGS 公式が得られます。

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}$$

逆に、BFGS 公式で、 $B = H$  とおき、 $s$  と  $y$  を交換すれば、以下に示す  $B$  公式の DFP 公式が得られます。

$$B_{k+1} = B_k + \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right) \frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k y_k^T + y_k s_k^T B_k}{s_k^T y_k}$$

このことから、BFGS 公式と DFP 公式は互いに双対であるといわれています。

## 4.4 直線探索

直線探索の目的は、大域的収束性を実現することにあります。大域的収束性とは、初期点の選び方によらず、点列  $\{x_k\}$  が  $x_*$  に収束するという性質です。刻み幅  $\alpha_k$  に関する適当な条件が満たされれば、大域的収束性が得られることが知られています。大域的収束性を保証するための直線探索の基準はいくつか提案されていますが、代表的な基準として、Curry の基準、Goldstein の基準、Wolfe の基準、Armijo の基準などがあります。

とくに、 $d_k$  方向で刻み幅  $\alpha$  に関して  $f(x)$  を最小にする  $\alpha_k$  に選ぶこと、すなわち、

$$f(x_k + \alpha_k d_k) = \min_{\alpha} f(x_k + \alpha d_k)$$

となる  $\alpha_k$  を求めることを、「正確な直線探索」といい、Curry の基準はその一つです。しかしながら、 $d_k$  方向で目的関数を最小化することは、数値計算上、実現不可能です。したがって、実際の計算では、Wolfe の基準や Armijo の基準などの緩和された基準が用いられます。

## 5. 非線形最小二乗問題

たとえば、 $p$  組のデータ  $(t_j, b_j)$  ( $j = 1, \dots, p$ ) を、モデル関数  $F(x, t)$  に当てはめて、係数ベクトル  $x$  を決定する場合の一つの手段として、

$$\sum_{j=1}^p (F(x, t_j) - b_j)^2$$

を未知係数ベクトル  $x$  について最小にすることが考えられます。関数  $F$  が  $x$  について非線形であるとき、この問題は非線形最小二乗問題 (nonlinear least squares problem) と呼ばれ、実験式の当てはめや経済モデルの当てはめなど、現実問題でしばしば発生します。非線形最小二乗問題の数値解法は、問題の構造を活かしたものが工夫されています。そこで、本章では、非線形最小二乗問題に対する数値解法を紹介します。

無制約非線形最小二乗問題は、次のように定式化されます。

$r_j: \mathbb{R}^n \rightarrow \mathbb{R}$  ( $j = 1, \dots, p$ ) ( $p \gg n$ ) が与えられたとき、目的関数

$$\begin{aligned} f(x) &= \frac{1}{2} r(x)^T r(x) \\ &= \frac{1}{2} \|r(x)\|^2 \end{aligned} \quad (7)$$

$$= \frac{1}{2} \sum_{j=1}^p r_j(x)^2$$

を  $x \in R^n$  について最小にせよ。ただし、

$$r(x) = (r_1(x), \dots, r_p(x))^T \quad (8)$$

である。

$p$  はデータの個数 (サンプル数) に相当するので、通常は  $n$  に比べて非常に大きくなります。以下、非線形最小二乗問題を単に問題とといいます。問題に対する数値解法として、目的関数が2回微分可能ならば、ニュートン法が有力な手法として考えられます。この場合、勾配ベクトルおよびヘッセ行列は次のように表わされます。

$$\begin{aligned} \nabla f(x) &= \sum_{j=1}^p r_j(x) \nabla r_j(x) \\ &= J(x)^T r(x), \\ \nabla^2 f(x) &= J(x)^T J(x) + \sum_{j=1}^p r_j(x) \nabla^2 r_j(x), \end{aligned}$$

ただし、 $J(x)$  はベクトル値関数  $r(x)$  のヤコビ行列です。

ニュートン法の探索方向  $d_k$  は、

$$\Omega d_k = -J(x_k)^T r(x_k),$$

ただし、

$$\Omega = \left( J(x_k)^T J(x_k) + \sum_{j=1}^p r_j(x_k) \nabla^2 r_j(x_k) \right)$$

から求められます。ニュートン法は局所的に収束速度が速いという長所がありますが、ヘッセ行列の第2項を毎回計算するのは大変な手間がかかりそうです。しかしながら、勾配ベクトルの計算からヤコビ行列は求まっていますから、 $J(x_k)^T J(x_k)$  の部分は利用できます。したがって、このような特殊性を利用したニュートン法の変種が考えられ、かなりの成功をおさめています。ここでは、Gauss-Newton法、Levenberg-Marquardt法、および準ニュートン法を紹介します。

### 5.1 Gauss-Newton 法

Gauss-Newton法の反復公式は、

$$x_{k+1} = x_k + d_k, \quad (9)$$

ただし  $J(x_k)^T J(x_k) d_k = -J(x_k)^T r(x_k)$  で与えられます。ここで、 $\text{rank} J(x_k) = n$  ならば、探索方向  $d_k$  は一意に決定され、かつ、目的関数の降下方向になります。

Gauss-Newton法について、次の2つの解釈ができます。

1. ニュートン法の反復公式で、ヘッセ行列の第2項を無視して、 $\nabla^2 f(x_k)$  を  $J(x_k)^T J(x_k)$  で近似している。
2. 現在の点  $x_k$  で  $r(x)$  を1次近似して得られる線形最小二乗問題

$$\min_d \frac{1}{2} \|J(x_k) d + r(x_k)\|^2$$

の解を探索方向  $d_k$  に選ぶ。

Gauss-Newton法の収束性は、ヘッセ行列の第1項の大きさと第2項の大きさに関係しています。すなわち、最適解  $x_*$  において第1項  $J(x_*)^T J(x_*)$  に比べて第2項が無視できるならば収束し、残差  $r(x_*)$  が零の場合には本来のニュートン法と同様の挙動を示します。逆に、第2項の影響が大きい場合には収束速度が遅くなり、場合によっては収束しないことも起こり得ます。ここで、ヘッセ行列の第2項が第1項に比べて無視できる場合とは、最適解  $x_*$  での残差  $\|r(x_*)\|$  が非常に小さい場合、および各  $r_j(x)$  がほとんど線形に近い場合をいいます。

### 5.2 Levenberg-Marquardt 法

Levenberg(1944)とMarquardt(1963)によって、それぞれ提案されたアルゴリズムで、反復公式は、

$$x_{k+1} = x_k + d_k, \quad (10)$$

ただし、

$$(J(x_k)^T J(x_k) + \mu_k I) d_k = -J(x_k)^T r(x_k) \quad (11)$$

で与えられます。ここで、 $\mu_k$  が正であれば、 $d_k$  は目的関数の降下方向となります。パラメータ  $\mu_k$  については、いくつかの意味付けがされています。ひとつは、Gauss-Newton法で  $J(x_k)^T J(x_k)$  の正定性が崩れた場合でも、 $\mu_k$  に正の値を与えることで、降下方向  $d_k$  が得られるという効果です。もうひとつとして、 $x_k$  のまわりでの  $f(x)$  を線形近似したモデルに対する、 $x_k$  の近傍での線形最小二乗解が、ちょうど(11)の解として与えられるという解釈です。この解釈は、さきに述べた信頼領域法と深い係りを示すものとされています。

### 5.3 準ニュートン更新公式を用いた方法

Gauss-Newton 法は、残差  $\|r(x_*)\|$  が小さい場合、および  $r(x)$  が線形に近い場合には効率よく働きますが、そうでない場合には必ずしも効率が良いとは限りません。これは、ヘッセ行列の第2項を無視したことによりです。そこで、一般の無制約最小化問題に対して非常に有効な、準ニュートン法を導入することが考えられます。非線形最小二乗問題の場合にはヘッセ行列が特別な構造をしているので、一般の場合のようにヘッセ行列全体を近似するのではなく、第2項を近似行列  $A_k$  で置き換えて、

$$(J(x_k)^T J(x_k) + A_k) d_k = -J(x_k)^T r(x_k) \quad (12)$$

から探索方向  $d_k$  を決定するアルゴリズムが考えられます。ここで、行列  $A_k$  は準ニュートン更新公式によって生成されます。

本節では、Dennis-Gay-Welsch 法（以下、略して DGW 法）を紹介します。

準ニュートン法を非線形最小二乗法に適用するとき、問題の特殊構造を活かしたセカント条件を選びます。ヘッセ行列の第1項はヤコビ行列から形成されますが、第2項は  $r_j(x)$  のヘッセ行列が必要で複雑です。そこで、第2項の近似行列として  $A_k$  を導入することが考えられます。このとき、 $A_{k+1}$  に関するセカント条件を

$$(J(x_{k+1})^T J(x_{k+1}) + A_{k+1}) s_k = z_k \quad (13)$$

ただし、

$$z_k = J(x_{k+1})^T J(x_{k+1}) s_k + (J(x_{k+1}) - J(x_k))^T r(x_{k+1})$$

とすることが可能です。

行列  $A_k$  を導入した理由は、ヘッセ行列の第2項が無視できない場合には、Gauss-Newton 法の効率が悪くなるからでした。しかし逆に、残差が非常に小さな問題に対しては、Gauss-Newton 法に準ずることが望ましいでしょう。すなわち、行列  $A_k$  が零行列に近づくことが望ましいこととなります。そこで、Gauss-Newton 法に準ずるようになるために、伸縮パラメータ  $\beta_k$  を導入して、 $\beta_k A_k$  を更新するような工夫が考えられました。パラメータ  $\beta_k$  の選び方としては、以下の様なものがあります。

$$\beta_k = \min\left(\left|\frac{s_k^T q_k}{s_k^T A_k s_k}\right|, 1\right),$$

ただし、

$$q_k = (J(x_{k+1}) - J(x_k))^T r(x_{k+1}).$$

以上の考えに基づく DGW 法の更新公式は、次式で与えられます。

$$A_{k+1} = \beta_k A_k - \frac{s_k^T u_k}{(s_k^T z_k)^2} z_k z_k^T + \frac{u_k z_k^T + z_k u_k^T}{s_k^T z_k}. \quad (14)$$

ただし、

$$u_k = q_k - \beta_k A_k s_k.$$

もし、 $r(x_{k+1}) = 0$  ならば、 $q_k = 0$ 、 $\beta_k = 0$  なので  $A_{k+1} = 0$  となり、Gauss-Newton 法に帰着します。他方、残差が大きい問題では  $\beta_k = 1$  が選ばれるでしょうから、ヘッセ行列の第2項が行列  $A_k$  で近似されます。以上のことから、伸縮を行った更新公式(14)を用いたアルゴリズムは、最適解  $x_*$  での残差が大きい問題にも、小さい問題にも効率が良いことが期待されます。また初期行列として、初回が Gauss-Newton 法の探索方向になるように、普通は  $A_0 = 0$  が選ばれます。

## 6. おわりに

上に紹介した数値解法は、いずれも第一回に紹介した文献群に、基本的な手法として詳しく解説されています。とくに、教科書に相当として紹介したテキストには、分かりやすく丁寧な解説がありますので、もし、詳しいところまで調べたいと考えておられる方は、それらのテキストをご覧ください。

さて、いよいよ次は最終回ですが、最終回は制約条件のある非線形最適化問題について概説します。