第10章　非線形方程式の解法

非線形方程式は様々な応用分野を持ち、数多くの解法が提案されている。数理計画問題のその最適性条件を満たす点をもとめる問題は、非線形方程式に定式化できる。そのため、数理計画問題に対する解法の多くは、本質的には非線形方程式の解法に基づいている。そのため、非線形方程式の解法を勉強することは、数理計画法においても重要である。この章では，微分可能な関数 $F : R^n \to R^n$ が与えられたとき，非線形方程式

$$F(x) = 0$$

を解く解法を紹介する．

10.1　数理計画問題との関係

この節では、まず、非線形方程式と数理計画問題の関係を見る。

非線形方程式

$$F(x) = 0 \tag{10.1}$$

は、次の制約なし最小化問題に定式化できる。

$$\min\ \frac{1}{2}\|F(x)\|^2 \tag{10.2}$$

この目的関数はすべての $x \in R^n$ に対して $\frac{1}{2}\|F(x)\| \geq 0$ であり、点 $\bar{x}$ において最小値 $\frac{1}{2}\|F(\bar{x})\| = 0$ をとることと、$\bar{x}$ が非線形方程式 $F(x) = 0$ の解であることは等価である。そのため、等価な最小化問題 (10.2) に対して、制約なし最小化問題に対するニュートン法などの手法を使うことによって、非線形方程式 (10.1) を解くことができる。

一方、数理計画問題

$$\begin{array}{ll} \min & f(x) \\ \text{subjectto} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

の最適性の必要条件 (Karush-Kuhn-Tucker 条件) を求める問題は、以下のように、非線形方程式に変換することができる。まず、Karush-Kuhn-Tucker 条件は、ラグランジュ関数を用いることによって,

$$\begin{gathered} \nabla_x L(x, \lambda, \mu) = 0 \\ h(x) = 0 \\ g(x) \leq 0,\ \mu \geq 0,\ g(x)^T \mu = 0 \end{gathered}$$

表すことができたことを思い出そう。このシステムは次の方程式系と等価である．

$$F(x,\lambda,\mu) \equiv \begin{pmatrix} \nabla_x L(x,\lambda,\mu) \\ h(x) \\ \phi(g_1(x),\mu_1) \\ \vdots \\ \phi(g_r(x),\mu_r) \end{pmatrix} = 0 \tag{10.3}$$

ここで，$\phi : R^2 \to R$ は

$$\phi(a,b) = \sqrt{a^2+b^2} + a - b$$

で定義される関数で Fischer-Burmeister 関数と呼ばれている．この ϕ は

$$\phi(a,b) = 0 \ \Leftrightarrow \ a \le 0, b \ge 0, ab = 0$$

という性質がある．このため，方程式 (10.3) は最適性の必要条件と等価になることがわかる[1]．このように、Karush-Kuhn-Tucker 条件は非線形方程式に定式化できる。このことは、凸計画問題は非線形方程式に定式化して解くことができることを意味している。また、F を

$$F(x,\lambda,\mu) \equiv \begin{pmatrix} \nabla_x L(x,\lambda,\mu) \\ h(x) \\ g_1(x)\mu_1 \\ \vdots \\ g_r(x)\mu_r \end{pmatrix} \tag{10.4}$$

と定義すれば、$g(x) \le 0$, $\mu \ge 0$ 以上という制約上での非線形方程式

$$F(x,\lambda,\mu) = 0$$

は、Karush-Kuhn-Tucker 条件と等価である。この制約つきの非線形方程式を解くという解法が、後の章で紹介する内点法である。

　このように数理計画法と非線形方程式は密接な関係がある。そのため、非線形方程式の解法を理解することは、数理計画法の解法を理解する上でも重要である。

10.2　ニュートン法

　非線形方程式の解法としてニュートン法を紹介する。このニュートン法は、制約なし最小化問題のニュートン法と本質的に一緒であることを説明する．

　非線形方程式の解法として、古典的ではあるが、今なお有効な手法がニュートン法である。ニュートン法では、非線形方程式を表す関数 F を、反復点 x^k において 1 次近似する。

$$F^k(x) := F(x^k) + \nabla F(x^k)^T (x - x^k)$$

この一次近似した関数に対する線形方程式の解は、非線形方程式の近似解とみなすことができる。この近似解が満足できる解でなければ、より満足できる解を求めるために、その近似解を x^{k+1} として、よりよい近似解を求める手法がニュートン法である。ここで、$F^k(x) = 0$ の解は、

$$x^{k+1} := x^k - (\nabla F(x^k)^T)^{-1} F(x^k)$$

図 10.1: ニュートン法

で与えられる。なお 1 次元の方程式に対するニュートン法は、図 10.2 のように表される。

ニュートン法が点列を生成する上で重要なことは、$\nabla F(x^k)$ が正則となることである。もし正則とならなければ、次の反復点を求めることすらできない。そのような欠点に対する対処方法として、次の節で紹介するガウスニュートン法や修正ガウスニュートン法 (Levenberg-Marquardt 法) がある。

古典的なニュートン法が使われる理由は、ニュートン法の収束が速い、つまり 2 次収束することである。ここで、$\nabla F(x^k)$ は正則であり、次の不等式を満たす正の定数 c が存在するとしよう。

$$\|nablaF(x^k)^{-1}\| \leq c \tag{10.5}$$

さらに、F が 2 回微分可能だとしよう。そのとき、2 回微分可能であることから、

$$\|F(x) - F(y) - \nabla F(x)^T(x-y)\| = O(\|x-y\|^2) \tag{10.6}$$

が成り立つ。ニュートン法が高速な収束をするのは、この 2 つの性質が成り立つからである。この性質を用いると、x^* をこの非線形方程式の解として、

$$\begin{aligned}
\|x^{k+1} - x^* \quad | &\leq \|x^k - (\nabla F(x^k)^T)^{-1}F(x^k) - x^*\| \\
&= \|(\nabla F(x)^T)^{-1}(\nabla F(x^k)(x^k - x^*) - (F(x^k) - F(x^*)))\| \\
&\leq \|(\nabla F(x)^T)^{-1}\|\|(\nabla F(x^k)(x^k - x^*) - (F(x^k) - F(x^*)))\| \\
&= O(\|x^k - x^*\|^2)
\end{aligned}$$

となる。このため、2 次収束していることがわかる。

なお、制約なし最小化問題のニュートン法は、最適性の 1 次の必要条件

$$\nabla f(x) = 0$$

を非線形方程式と考え、その非線形方程式に対してニュートン法を行なったことと同じである。

10.3 ガウスニュートン法

ニュートン法にはいくつかの欠点があった。ここでは、その欠点を克服する手法として、ガウスニュートン法と、その修正版である修正ガウスニュートン法の説明をする。

[1] ϕ は $(0,0)$ で微分不可能な関数のため，この方程式系も微分不可能となるが、狭義の相補性 $-g(x) + \lambda > 0$ が成り立つ点において微分可能なため、たいていの場合は微分可能であると想定してよい。

ニュートン法では、$\nabla F(x^k)$ が正則でなければ点列を生成できないという欠点があった。それは、ニュートン法では、

$$F^k(x) = 0$$

とする解 x^k を見つけることができなかったことを意味している。そこで、この線形方程式を解く代わりに、その残差を最小にする次の最小化問題を解くことを考える。

$$\min \quad \frac{1}{2}\|F^k(x)\|^2$$

このとき、この最小化問題は F^k が線形関数であることから、制約なしの凸 2 次計画問題となる。そこで、その最適性の条件を満たす点が、この問題の解となる。つまり、

$$0 = \nabla(\frac{1}{2}\|F^k(x)\|^2) = \nabla F(x^k)\nabla F(x^k)^T(x - x^k) + \nabla F(x^k)F(x^k) \tag{10.7}$$

を満たす点である。この線形方程式の解 x^{k+1} は必ず存在する。また、$\nabla F(x^k)$ が正則であれば、その解 x^{k+1} はニュートン法で生成された反復点と一致することに注意しよう。このように点列 $\{x^k\}$ を生成する手法をガウスニュートン法と呼ぶ。ガウスニュートン法も、ニュートン法と同様にして、解の近傍において $\nabla F(x)$ が正則であり、F が 2 回微分可能であれば、解に 2 次収束する。一方、もし $\nabla F(x^k)$ が正則でなければ、線形方程式 (10.7) は性質の良くない問題となり、またその解は無数に存在する。そこで、そのような欠点を補うため次の修正ガウスニュートン法が考えられている。この手法は Levenberg-Marquardt 法と呼ばれることもある。

ここで、線形方程式 10.7 において、$d = x - x^k$ として、探索方向 d を求める問題とすることにしよう。

$$\nabla F(x^k)\nabla F(x^k)^T d = -\nabla F(x^k)F(x^k)$$

この線形方程式の性質がよくないときは、$\nabla F(x^k)\nabla F(x^k)^T$ が正則でない場合である。ところで、この行列の構造を見てみると、この行列は半正定値行列であることがわかるそのため、この行列に適当な $\mu I, \mu > 0$ を足してやることによって、正定値行列にすることができる。実際、$y \neq 0$ である任意のベクトル $y \in R^n$ に対して、

$$y^T(\nabla F(x^k)\nabla F(x^k)^T + \mu I)y = y^T(\nabla F(x^k)\nabla F(x^k)^T y + \mu\|y\|^2 > 0$$

となることより、そのような行列が正定値行列だることがわかる。そこで、探索方向 d を次の線形方程式の解とする。

$$(\nabla F(x^k)\nabla F(x^k)^T + \mu I)d = -\nabla F(x^k)F(x^k)$$

こうすることによって、d は一意に求めることができる。そして、反復

$$x^{k+1} := x^k + d$$

を繰り返すのが修正ガウスニュートン法である。修正ガウスニュートン法は、k が大きくなるにつれて $\mu \to 0$ になるよう μ を変更していけば、ガウスニュートン法とほとんど同じになることがわかる。そのため、修正ガウスニュートン法も、2 次収束するアルゴリズムである。

ところで、これまで紹介したニュートン法、ガウスニュートン法、修正ガウスニュートン法とも、局所的に速い収束は保証されていたが、大域的に収束することは保証されていなかった。修正ガウスニュートン法のよい点のひとつに、大域的収束を保証するアルゴリズムを構成しやすい

ことがある。この章の始めに述べたように、非線形方程式 $F(x)=0$ は、$\theta(x) := \frac{1}{2}\|F(x)\|^2$ を目的関数とする制約なし最小化問題

$$\min \quad \theta(x)$$

と等価である。そこで、この最小化問題を解くために修正ガウスニュートン法を組み合せてやることを考える。実は、修正ガウスニュートン法で求められる探索方向 d は θ の降下方向になっているのである。

$$d^T\nabla\theta(x^k) = d^T(\nabla F(x)^k F(x^k)) = -d^T(\nabla F(x^k)\nabla F(x^k)^T + \mu I)d \le -\mu\|d\|^2$$

このことを用いれば、適当なステップサイズ t を用いて、反復を繰り返せば等価な最小化問題の解を求めることができる。また、解の周辺でステップサイズが 1 となれば、修正ガウスニュートン法の効果により、2 次収束することが期待できる。

この観点から、次の修正ガウスニュートン法を構成することができる。

修正ガウスニュートン法

Step 0. (初期化)： 適当にパラメータ $p>2,\ \beta\in(0,\frac{1}{2})$ と $\mu^0>0$, 初期点 $x^0\in R^n$ を選ぶ. $k:=0$ とする.

Step 1. (終了のチェック)： もし終了条件を満たしていれば終了.

Step 2. (探索方向の計算)： 次の方程式を満たす探索方向 d^k を求める.

$$(\nabla F(x^k)\nabla F(x^k)^T + \mu^k I)d = -\nabla F(x^k)F(x^k) \tag{10.8}$$

Step 3. (直線探査)： 次の不等式を満たす最小の非負の整数 l を求める.

$$\theta(x^k + 2^{-l}d^k) \le \theta(x^k) + \beta 2^{-l}\nabla\theta(x)^T d^k.$$

$x^{k+1} := x^k + 2^{-l}d^k$ とする．$k:=k+1$ として Step1 へ.

このアルゴリズムに対して，次の定理が成り立つ.

定理 1 生成される点列の集積点は θ の停留点である．さらに，集積点の一つ x^* において、$\nabla F(x^*)$ が正則であり、F は x^* の近傍で *2* 回微分可能であるとする。このとき、$\mu^k = O(\|F(x^k)\|)$ であれば、生成される点列は x^* に *2* 次収束する.

この定理より、このアルゴリズムが解の周辺では非常に速い収束をすることがわかる。しかしながら、解から遠いところから、解の周辺に近づくためにかなり反復回数を必要とすることもあり、いかに速く解の周辺に近づけるようにするかの工夫が必要となる。そのような工夫として、次の節で紹介するホモトピー法がある。

10.4 ホモトピー法

ニュートン型手法は解の周辺では高速に収束するという長所を持つが、一方、解から遠いところからではかなりの反復回数を必要とする場合がある。そのような欠点を補うために、解くべき方程式を変形した方程式を各反復で作り、その方程式に対してニュートン型手法を用いる手法が開

発されている。この変形した方程式の解が現在の反復点の近くになるように構成できれば、ニュートン型手法の長所により、その解を高速に求めることができる。そして、変形した方程式を徐々に元の方程式に近づけてやれば、トータルでも高速に解を持つことが期待できる。このような手法がホモトピー法である。

10.4.1　ホモトピーパス

ここで次のような関数 $G : R^{n+1} \to R$ を考える。

- $G(x, 0) = F(x)$
- ある $\bar{\mu}$ に対して、$G(x, \bar{\mu}) = 0$ の解は簡単に求まる。
- G は微分可能
- $\nabla G(x, \mu)$ は正則

例えば、

$$G(x, \mu) = \mu x + (1 - \mu)F(x)$$

とすると、$\bar{\mu} = 1$ のとき、$G(x, \bar{\mu}) = 0$ の解は簡単に求まる。

また、$\nabla G(x, \mu)$ が正則であれば、陰関数定理を用いれば、

$$G(x, \mu) = 0$$

を満たす点 x は唯一に求まることがわかる。そのため、その点を μ に対する関数として $x(\mu)$ とすることができる。このとき、この点の集合 $\{x(\mu) \| \mu \in [0, \bar{\mu}]\}$ は、なめらかな曲線となり、$x(0)$ は元の非線形方程式の解となる。この曲線をホモトピーパスと呼ぶ。このホモトピーパスを逐次的に追跡する手法がホモトピー法である。

具体的には、ホモトピー法は以下のように与えられる。

ホモトピー法

Step 0. (初期化)： $\mu_0 = \bar{\mu}$, 初期点 $x^0 \in R^n$ を選ぶ. $k := 0$ とする.

Step 1. (終了のチェック)： もし終了条件を満たしていれば終了.

Step 2. (反復点の計算)： 次の方程式の近似解 x^{k+1} を求める.

$$G(x, \mu_k) = 0$$

Step 3. (μ_k の更新)： $0 \leq \mu_{k+1} < \mu_k$ となる μ_{k+1} を求める。$k := k + 1$ として Step1 へ.

ホモトピー法で、ステップ 2 において、前の点 x^k を初期点として、ニュートン型手法を使えば、x^{k+1} を求めることができる。特に、μ_k と μ_{k+1} が近ければ、高速に次の点が求まる。しかし、μ_k と μ_{k+1} が近ければ、アルゴリズムの目的は、$\mu_k \to 0$ とすることであるので、全体の収束が遅くなる。そのため、μ_k の更新の決め方は非常に重要である。また、アルゴリズム中では ”近似解 ” としか記述されていなかったが、どの程度の近似解を求めるかも、アルゴリズムの速度に影響を与える。さらに重要なことは、いかに性質のよい G を見つけるかということもある。凸計画問題の解法である内点法は、(i) よい μ_k の更新規則を与え、(ii) 近似解のルールを明確に与え、(iii) 性質のよい G を用いている。このため、この手法により、凸計画問題は高速に解くことができる。次の章ではこの内点法を紹介する。