第 1 章　はじめに

非線形最適化モデルは, 典型的に 2 つの問題に定式化される.

制約なし最適化問題

n 実変数 $x_1, x_2, \cdots, x_n$ の非線形実数値関数 $f(\boldsymbol{x})$ の最小値を与える $\boldsymbol{x}$ を求める問題. ただし $\boldsymbol{x}$ は第 i 要素を x_i とする n 項列ベクトルである.

制約つき最適化問題

方程式 $g_1(\boldsymbol{x}) = 0, \cdots, g_r(\boldsymbol{x}) = 0$ および不等式 $g_{r+1} \geq 0, \cdots, g_m \geq 0$ で実現される制約条件下で, 非線形目的関数 $f(\boldsymbol{x})$ の最小値を与える n 次元列ベクトル $\boldsymbol{n}$ を求める問題.

非線形最適化問題を有限回の四則演算で解くことは, 一般的にはできない. 最適化法のアルゴリズムは逐次近似 (反復) 解法の形を取る. すなわち, 適当な初期近似解 $\boldsymbol{x}_0$ から出発して, 反復公式,

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \tag{1.1}$$

によって最適解 $\boldsymbol{x}^*$ に収束する点列 $\boldsymbol{x}_k$ を生成する. ここで α_k は「ステップ幅」, $\boldsymbol{x}_k$ は「探索方向」と呼ばれる. この両者をどう選ぶかによってアルゴリズムが決まる. とりわけ探索方向 $\boldsymbol{d}_k$ の選び方がその性能を左右する. 最適解において極小値を取り, 探索方向がその降下方向 (すなわち $\boldsymbol{d}_k^T M(\boldsymbol{x}) < 0$) となるような評価関数 $M(\boldsymbol{x})$ を構成し, 反復の段階ではこの関数を減少させるようにする. すなわち,

$$M(\boldsymbol{x}_{k+1}) < M(\boldsymbol{x}_k) \tag{1.2}$$

となるようにする. 制約なし最適化問題では評価関数として目的関数自体が, 制約つき最適化問題では評価関数として罰金関数や罰金付き Lagrange 関数がよく用いられる.

この文章では非線形の最適化問題の数学的構造に関してまとめる. 最適化問題の解法アルゴリズムについては W. C. Davidson (1959) の可変計量法以来, 数多くのアルゴリズムが提案されてきた. その中でもギブス自由エネルギーを最小化する平衡熱力学計算に良く利用される Newton 法をここでは取り上げる. 制約なし最適化問題として Newton 法と準 Newton 法を, 制約つき最適化問題のとして Newton 法を紹介する.

第2章　制約なし最適化問題

滑らかな実数値関数 $f(\boldsymbol{x})$ の最小値を与える $\boldsymbol{x}$ を求める問題を考える．この問題を解くための反復解法 (1.1) 式において, 探索ベクトル $\boldsymbol{d}_k$ は,

$$\boldsymbol{d}_k \equiv -H_k^{-1}\nabla f(\boldsymbol{x}) \tag{2.1}$$

という形式を取ることが多い．ここで $\boldsymbol{H}$ はヘッセ (Hesse) 行列と呼ばれ, 例えば最急降下法では $\boldsymbol{H} \equiv \boldsymbol{I}$ (単位行列) であり, Newton 法では $\boldsymbol{H} \equiv \nabla^2 f(\boldsymbol{x})$ である．

本章では制約なし最適化問題として, ギブス自由エネルギーを最小化するという平衡熱力学計算に良く用いられる Newton 法, 準 Newton 法を解説する．

2.1　Newton 法

制約なしの最適化解は, 停留点を定義する連立非線形方程式

$$\nabla f(\boldsymbol{x}) = 0 \tag{2.2}$$

を満たす．この関係式を用いることで Newton 法の探索方向を決める．

目的関数 $f(\boldsymbol{x})$ を近似解 $\boldsymbol{x}_k$ のまわりでテーラー展開し, 2 次の微小量まで考慮した 2 次近似関数を Q とすると, 近似関数 Q は,

$$Q = f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}\nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k)^2 \tag{2.3}$$

と書ける．ただし, $\nabla = \partial/\partial \boldsymbol{x}_k$ である．(2.3) 式の $\boldsymbol{x}$ 微分をとると ($\boldsymbol{x}_k$ 微分でないことに注意),

$$\begin{aligned} \frac{dQ}{d\boldsymbol{x}} &= \frac{d}{d\boldsymbol{x}}f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)\cdot\frac{d}{d\boldsymbol{x}}(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}\nabla^2 f(\boldsymbol{x}_k)\cdot\frac{d}{d\boldsymbol{x}}(\boldsymbol{x} - \boldsymbol{x}_k)^2 \\ &= \nabla f(\boldsymbol{x}_k) + \nabla^2 f(\boldsymbol{x}_k)\cdot(\boldsymbol{x} - \boldsymbol{x}_k) \end{aligned} \tag{2.4}$$

となる．(2.4) 式を停留点の条件 (2.2) 式に代入することで,

$$\nabla f(\boldsymbol{x}_k) + \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d}_k = 0 \tag{2.5}$$

が得られる．ただし (2.4) において $\boldsymbol{x} - \boldsymbol{x}_k$ を探索方向 $\boldsymbol{d}_k$ とみなした．

Newton 法では, (2.5) 式より探索方向を決め, ステップ幅を常に $\alpha_k \equiv 1$ にすることによって, 最適解 $\boldsymbol{x}^*$ に収束させる．(2.5) の定式化からわかるように, この探索方向は目的関数を変数 $\boldsymbol{x}$ に関する 2 階の方程式に近似し, その停留点により目的関数の最小値を求める方法といえる．

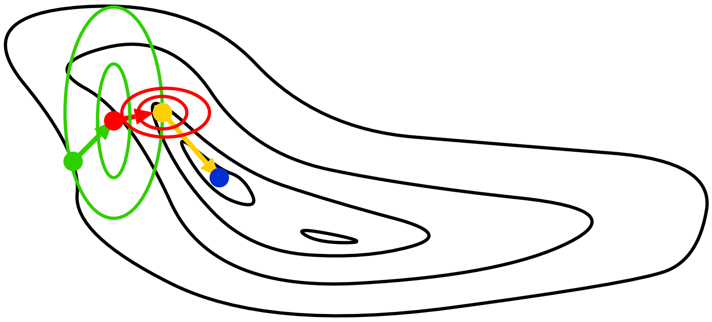

図 2.1: Newton 法が生成する点列の意味

2.2 準 Newton 法

(2.5) 式の $\nabla^2 f(\boldsymbol{x}_k)$ の直接計算は手間がかかるため, 1959 年に W. C. Davidon は $\nabla^2 f(\boldsymbol{x}_k)$ を直接計算せずに済む「可変計量法」を開発した. 現在ではその変種が数多くしられており, それらは総称して「準 Newton 法」と呼ばれている. この方法群においては, $\{\boldsymbol{x}_k\}, \{\nabla f(\boldsymbol{x}_k)\}$ の情報を用いて $\nabla^2 f(\boldsymbol{x}_k)$ の近似行列の系列 $\{H_k\}$ を構成して $\nabla^2 f(\boldsymbol{x}_k)$ に代用する.

2 つのベクトルを $\boldsymbol{s}_k \equiv \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$, $\boldsymbol{y}_k \equiv \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$ と定義するとき, 近似的に $\nabla^2 f(\boldsymbol{x})\boldsymbol{s}_k \approx \boldsymbol{y}_k$ が成り立つ. そこで $\boldsymbol{s}_k$ と $\boldsymbol{y}_k$ の対が得られるたびに, セカント条件,

$$H_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k \tag{2.6}$$

を満たすように, 近似行列を $H_{k+1} = H_k + \Delta H_k$ の形式によって逐次近似していくと, 系列 $\{H_k\}$ はある意味で $\nabla^2 f(\boldsymbol{x}_k)$ の近似になることが期待できる. そこで $\boldsymbol{x}_0, H_0$ を決め, (2.1) 式によって探索方向 $\boldsymbol{d}_0$ を決め, $\boldsymbol{x}_1$ を生成し, (2.6) 式に基づいて H_1 を更新する. これを反復して点列 $\{\boldsymbol{x}_k\}$ を生成する. ステップ幅の決定は Wolfe の条件,

$$f(\boldsymbol{x} + \alpha\boldsymbol{d}_k) - f(\boldsymbol{x}_k) \le \sigma_1\alpha(\nabla f(\boldsymbol{x}_k))^T\boldsymbol{d}_k \tag{2.7}$$

$$(\nabla f(\boldsymbol{x} + \alpha\boldsymbol{d}_k))^T\boldsymbol{d}_k \le \sigma_2(\nabla f(\boldsymbol{x}_k))^T\boldsymbol{d}_k \tag{2.8}$$

によって行う. ただし, 準 Newton 法においては, この条件が極端に緩くなるようにこの定数 σ_1, σ_2 を選択すると, 総合性能が良くなることが経験的に知られている. 典型的には $\sigma_1 \equiv 10^{-4}, \sigma_2 \equiv 0.9$, さらに H_0 として単位行列が選ばれることが多い.

今日, 経験的にも理論的にも最も優れているとされている準 Newton 法は, C. G. Broyden, R. Fletcher, D. Goldfalb, D. F. Shanno による BFGS 公式,

$$H_{k+1} = H_k - \frac{H_k\boldsymbol{s}_k\boldsymbol{s}_k^T H_k}{\boldsymbol{s}_k^T H_k\boldsymbol{s}_k} + \frac{\boldsymbol{y}_k\boldsymbol{y}_k^T}{\boldsymbol{s}_k^T\boldsymbol{y}_k} \tag{2.9}$$

に基づくものである.

準 Newton 法は制約のない最適化問題の中で最もよく使われる方法であるが, 行列 H_k のために大きな記憶領域を必要とする点に適用上の限界がある. 90 年代以降の最適化法による平衡熱力学計算においては, 準 Newton 法が最も良く利用されている.

第3章 制約つき最適化問題

本章では,

$$g_1(\boldsymbol{x}) = 0, \cdots, g_r(\boldsymbol{x}) = 0 \tag{3.1}$$

$$g_{r+1}(\boldsymbol{x}) \geq 0, \cdots, g_m(\boldsymbol{x}) \geq 0 \tag{3.2}$$

という条件下で, 目的関数 $f(\boldsymbol{x})$ を最小化する問題を考える.

3.1 Newton 法

$\boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}) = 0, \cdots, g_r(\boldsymbol{x}))^T = 0$ という等式制約のみの場合を考える. このとき, 制約想定下で最適解 $\boldsymbol{x}^*$ に対して $\boldsymbol{\lambda}^*$ が存在し, これらは Karush-Kuhn-Tucker 条件, すなわち変数 $\boldsymbol{x}, \boldsymbol{\lambda}$ に連立非線形方程式,

$$\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla_x f(\boldsymbol{x}) - (J(\boldsymbol{x}))^T \boldsymbol{\lambda} = 0 \tag{3.3}$$

$$\nabla_\lambda L(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{g}(\boldsymbol{x}) = 0 \tag{3.4}$$

を満たす. ただし λ は未定乗数, $L(\boldsymbol{x}, \boldsymbol{\lambda}) \equiv f(\boldsymbol{x}) - \sum \lambda_j g_j(\boldsymbol{x})$ とし, Jacobi 行列は $J(\boldsymbol{x}) \equiv \nabla \boldsymbol{g}(\boldsymbol{x})$ とする. この Karush-Kuhn-Tucker 条件はラグランジュの未定乗数法の条件と同じものである. 制約つき最適化法は究極的にはこの方程式を解かねばならないと言える.

Newton 法を (3.3), (3.4) 式から成る連立方程式に適用すること考える. (3.3), (3.4) 式より L を目的関数とみなせばよいことがわかる. ただし制約なしの最適化問題では目的関数を $\boldsymbol{x}$ に関する 2 次方程式に近似するだけでよかったが, 制約つき最適化問題では $\boldsymbol{x}$ と $\boldsymbol{\lambda}$ について停留点を求めるので, 目的関数を $\boldsymbol{x}$ と $\boldsymbol{\lambda}$ で展開せねばならない. まず目的関数 L を $\boldsymbol{x}_k$ の近傍でテーラー展開し, 2 次の微小量まで考慮する. そのようにして求めた $L(\boldsymbol{x}, \boldsymbol{\lambda})$ の近似関数を Q とすると,

$$\begin{aligned} Q &= f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k) \cdot (\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2} \nabla^2 f(\boldsymbol{x}_k) \cdot (\boldsymbol{x} - \boldsymbol{x}_k)^2 \\ &\quad -\boldsymbol{\lambda} \boldsymbol{g}(\boldsymbol{x}_k) - \boldsymbol{\lambda}_j \nabla \boldsymbol{g}(\boldsymbol{x}_k) \cdot (\boldsymbol{x} - \boldsymbol{x}_k) - \frac{1}{2} \boldsymbol{\lambda}_j \nabla^2 \boldsymbol{g}(\boldsymbol{x}_k) \cdot (\boldsymbol{x} - \boldsymbol{x}_k)^2 \end{aligned} \tag{3.5}$$

となる. さらに Q を $\boldsymbol{\lambda}_k$ の近傍でテーラー展開すれば,

$$\begin{aligned} Q &= f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k) \cdot (\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2} \nabla^2 f(\boldsymbol{x}_k) \cdot (\boldsymbol{x} - \boldsymbol{x}_k)^2 \\ &\quad -\boldsymbol{\lambda}_k \boldsymbol{g}(\boldsymbol{x}_k) - \boldsymbol{g}(\boldsymbol{x}_k) \cdot (\boldsymbol{\lambda} - \boldsymbol{\lambda}_k) \\ &\quad -\boldsymbol{\lambda}_k \nabla \boldsymbol{g}(\boldsymbol{x}_k) \cdot (\boldsymbol{x} - \boldsymbol{x}_k) - \nabla \boldsymbol{g}(\boldsymbol{x}_k) \cdot (\boldsymbol{x} - \boldsymbol{x}_k) \cdot (\boldsymbol{\lambda} - \boldsymbol{\lambda}_k) \\ &\quad -\frac{1}{2} \boldsymbol{\lambda}_k \nabla^2 \boldsymbol{g}(\boldsymbol{x}_k) \cdot (\boldsymbol{x} - \boldsymbol{x}_k)^2 - \frac{1}{2} \nabla^2 \boldsymbol{g}(\boldsymbol{x}_k) \cdot (\boldsymbol{x} - \boldsymbol{x}_k)^2 \cdot (\boldsymbol{\lambda} - \boldsymbol{\lambda}_k) \end{aligned} \tag{3.6}$$

となる．この近似関数 Q を $\boldsymbol{x}$ について微分すれば,

$$\begin{aligned}\nabla_x Q &= \nabla f(\boldsymbol{x}_k) + \nabla^2 f(\boldsymbol{x}_k)\cdot(\boldsymbol{x}-\boldsymbol{x}_k) - \boldsymbol{\lambda}\nabla \boldsymbol{g}(\boldsymbol{x}_k) - \boldsymbol{\lambda}\nabla^2 \boldsymbol{g}(\boldsymbol{x}_k)\cdot(\boldsymbol{x}-\boldsymbol{x}_k)\\ &= \left\{\nabla^2 f(\boldsymbol{x}_k) - \boldsymbol{\lambda}\nabla^2 \boldsymbol{g}(\boldsymbol{x}_k)\right\}(\boldsymbol{x}-\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k) - \boldsymbol{\lambda}\nabla \boldsymbol{g}(\boldsymbol{x}_k) \end{aligned} \tag{3.7}$$

となるので, 停留点の条件を用いると,

$$\begin{aligned}&\left\{\nabla^2 f(\boldsymbol{x}_k) - \boldsymbol{\lambda}\nabla^2 \boldsymbol{g}(\boldsymbol{x}_k)\right\}(\boldsymbol{x}-\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k) - \boldsymbol{\lambda}\nabla \boldsymbol{g}(\boldsymbol{x}_k) = 0\\ &H(\boldsymbol{x}_k,\boldsymbol{\lambda})\boldsymbol{d}_k - J(\boldsymbol{x}_k)\boldsymbol{\lambda} = -\nabla f(\boldsymbol{x}_k)\end{aligned} \tag{3.8}$$

である．ただし,

$$H(\boldsymbol{x},\boldsymbol{\lambda}) = \nabla^2 L(\boldsymbol{x},\boldsymbol{\lambda}) = \nabla^2 f(\boldsymbol{x}_k) - \boldsymbol{\lambda}\nabla^2 \boldsymbol{g}(\boldsymbol{x}_k) \tag{3.9}$$

という置き換えをおこなった．同様に近似関数 Q を $\boldsymbol{\lambda}$ 微分すると,

$$\nabla_\lambda Q = -\boldsymbol{g}(\boldsymbol{x}_k) - \nabla \boldsymbol{g}(\boldsymbol{x}_k)\cdot(\boldsymbol{x}-\boldsymbol{x}_k) - \frac{1}{2}\nabla^2 \boldsymbol{g}(\boldsymbol{x}_k)\cdot(\boldsymbol{x}-\boldsymbol{x}_k)^2 \tag{3.10}$$

となり, 停留点の条件を用いると,

$$\begin{aligned}&-\boldsymbol{g}(\boldsymbol{x}_k) - \nabla \boldsymbol{g}(\boldsymbol{x}_k)\cdot(\boldsymbol{x}-\boldsymbol{x}_k) = 0\\ &J(\boldsymbol{x}_k)\boldsymbol{d}_k = -\boldsymbol{g}(\boldsymbol{x}_k)\end{aligned} \tag{3.11}$$

が得られる．ここで $\boldsymbol{d}_k^2$ の項を微小量とみなして無視した.

(3.8) と (3.11) 式より, Newton 法を (3.3), (3.4) 式から成る連立方程式に適用することで次のアルゴリズムが得られる．探索方向 $\boldsymbol{d}_k$ を連立一次方程式,

$$\begin{bmatrix} H(\boldsymbol{x}_k,\boldsymbol{\lambda}_k) & -J^T(\boldsymbol{x}_k)\\ J(\boldsymbol{x}_k) & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{d}_k\\ \boldsymbol{\lambda}_{k+1}\end{bmatrix} = -\begin{bmatrix}\nabla_x f(\boldsymbol{x}_k)\\ \boldsymbol{g}(\boldsymbol{x}_k)\end{bmatrix}$$

によって定め, (1.1) 式においてステップ幅を常に $\alpha = 1$ にとって点列 $\{\boldsymbol{x}_k\}, \{\boldsymbol{\lambda}_k\}$ を生成する．$\boldsymbol{\lambda}_{k+1}$ は連立方程式の解として直接得られることに注意を払う必要がある[1].

この方法は, 一般に収束域が狭く, 初期値 $\boldsymbol{x}_0$ の選び方によっては発散したり, 解とは異なる Karush-Kuhn-Tucker 点に収束する可能性がある.

[1] $H = H(\boldsymbol{x}_k, \boldsymbol{\lambda}_{k+1})$ とならずに $H = H(\boldsymbol{x}_k, \boldsymbol{\lambda}_k)$ となる理由は?

第 4 章　参考文献

藤田宏, 今野浩, 田辺國士, 著:
「最適化法」, 岩波講座応用数学 方法 7, 岩波書店, 1994.

田中正次, 二宮市三, 鳥居達生, 長谷川武光, 秦野和郎, 秦野, 吉田年雄, 刀根薫, 古林隆, 細野俊夫, 著:
「SSL II 使用手引書」, 富士通株式会社, 1987.