第6章　微分を使わない最適化法

今回の授業では，関数の微分などを使わないで，関数値だけをもちいる最適化のアルゴリズムを紹介する．そのようなアルゴリズムは，多くの問題に対して適用可能であり、計算機への実装が容易である．一方、微分値などの情報を用いていないため、大域的収束性や速い収束性などは保証されていない。

ここでは、まず 1 次元の最小化問題を解くアルゴリズムである黄金分割法を紹介する。1 次元の最小化問題は反復法でステップサイズを用いるときなどに現れる問題である。次に n 次元の制約なし最小化問題を解く直接探索法の紹介する。

6.1　黄金分割法

ここでは、目的関数 $f : R^n \to R$ を区間 $[l, u]$ の中で最小化する次の問題を考える (図 6.1)。

$$\begin{array}{ll} \min & f(x) \\ \text{subjectto} & l \le x \le u \end{array} \tag{6.1}$$

このような問題として、反復法においてステップサイズを求める問題がある。今、$\hat{f} : R^n \to R$ を最小化することを考えており、k 回目の反復で x^k と探索方向 d^k が与えられているとする。このとき、$\hat{f}(x^k + td^k) < \hat{f}(x^k)$ となるようなステップサイズ t を求める問題は、$f(t) = \hat{f}(x^k + td^k)$、$l = 0, u = 1$ として問題 (6.1) となる。(ここで簡単のため $l = 0, u = 1$ としたが他の値にしても構わない。)

問題 (6.1) を解く手法として、黄金分割法がある。ここでは、まず

仮定： f は凸関数である。

として、このアルゴリズムを説明しよう。黄金分割法では、まず、始めに $l < a < b < u$ となるよう a, b を用意する。そして、初期点として、$x^{0,0} := l, x^{0,1} := a, x^{0,2} := b, x^{0,3} := u$ とする (図 6.1)。その後、各反復では 4 つの点の組 $(x^{k,0}, x^{k,1}, x^{k,2}, x^{k,3})$ を更新する。ここで、$x^{k,j}$ の上についた添え字は、k が反復回数、j が 4 つの点につけた番号である。今、初期点においては、

$$x^{0,0} < x^{0,1} < x^{0,2} < x^{0,3}$$

となっていることに注意しよう。このとき、2 番目と 3 番目に大きい $x^{0,1}, x^{0,2}$ に着目する。今、$f(x^{0,1}) < f(x^{0,2})$ であったとしよう。このとき、区間 $[x^{0,2}, x^{0,3}]$ の中には、この問題の最小解は存在しない (図 6.2 左)。

これは次のように示すことができる。もし、その区間中に最小解 x^* があったとすると (背理法)、$x^{0,2}$ は $x^{0,1}$ と x^* の間にあるので、$x^{0,2} = \alpha x^{0,1} + (1 - \alpha)x^*$ となる $\alpha \in [0, 1]$ が存在する。今、f は凸関数であると仮定しているので、

$$f(x^{0,2}) = f(\alpha x^{0,1} + (1 - \alpha)x^*) \le \alpha f(x^{0,1}) + (1 - \alpha)f(x^*)$$

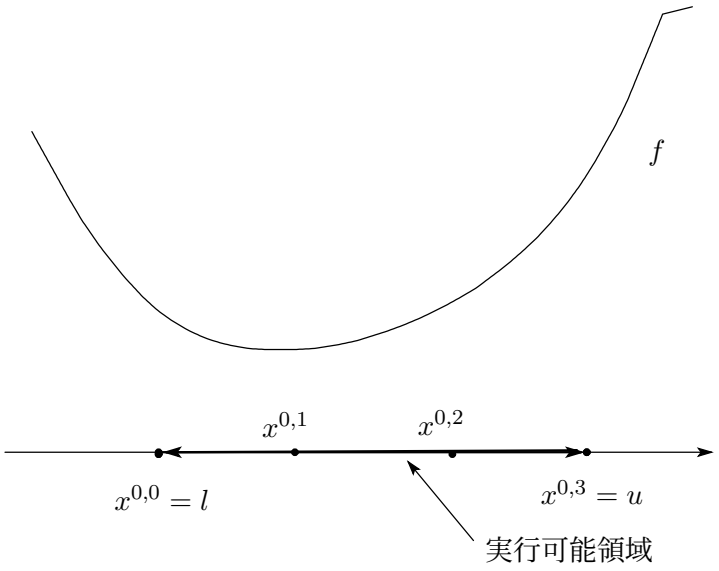


図 6.1: 1 次元の最小化問題と黄金分割法の初期点

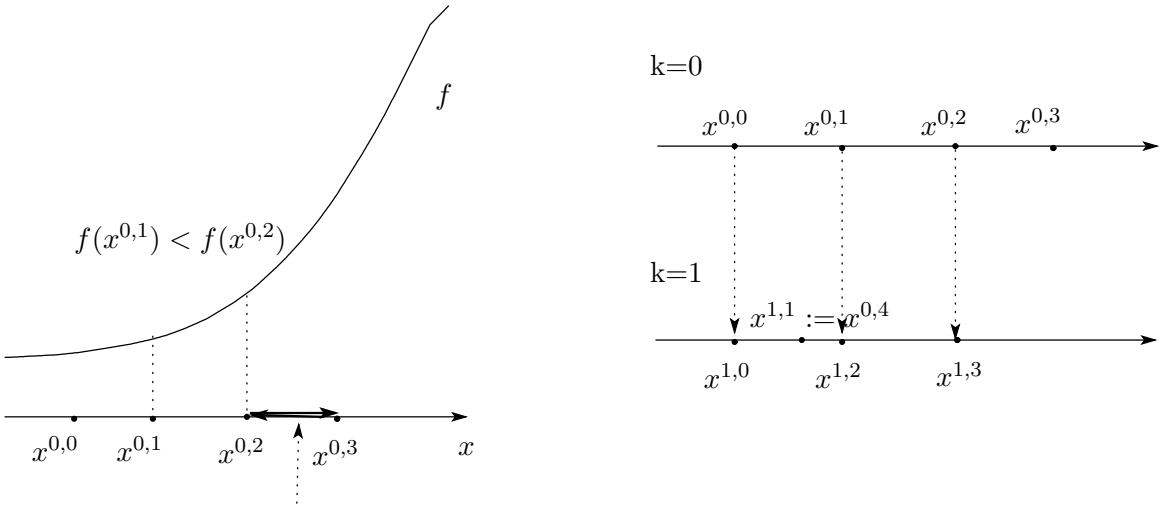


図 6.2: 最小解の場所と反復点

となる。一方、x^* は最小解であるので、$f(x^*) \leq f(x^{0,1})$ である。よって、

$$f(x^{0,2}) \leq \alpha f(x^{0,1}) + (1-\alpha) f(x^*) \leq f(x^{0,1})$$

となる。これは、$f(x^{0,2}) > f(x^{0,1})$ であったことに矛盾する。このため、$f(x^{0,1}) < f(x^{0,2})$ であれば区間 $[x^{0,2}, x^{0,3}]$ の中には最小解は存在しない。このことより、$f(x^{0,2}) > f(x^{0,1})$ のときには、区間 $[x^{0,0}, x^{0,2}]$ の中に最小解があることになる。そこで、必要ない $x^{0,3}$ の代わりに、区間 $(x^{0,0}, x^{0,1})$ の中から、$x^{0,4}$ をひとつ選び、

$$x^{1,0} := x^{0,0},\ \ x^{1,1} := x^{0,4},\ \ x^{1,2} := x^{0,1},\ \ x^{1,3} := x^{0,2}$$

とする (図 6.2 右)。このとき、

$$x^{1,0} < x^{1,1} < x^{1,2} < x^{1,3}$$

になっており、区間 $[x^{1,0}, x^{1,3}]$ の長さは、始めの区間 $[x^{0,0}, x^{0,3}]$ の長さよりも短くなっている。

一方、$f(x^{0,2}) < f(x^{0,1})$ のときには、同様に議論することによって、区間 $[x^{0,1}, x^{0,3}]$ の中に最小解が存在することがわかる。そこで、このときは、$x^{0,4}$ を区間 $(x^{0,2}, x^{0,3})$ の間に取る。そして、次の反復点を、

$$x^{1,0} := x^{0,1},\ \ x^{1,1} := x^{0,2},\ \ x^{1,2} := x^{0,4},\ \ x^{1,3} := x^{0,3}$$

とする。

これらのことを繰り返していく手法が黄金分割法である。このとき、反復 k において、最小解 x^* が存在する区間は $[x^{k,0}, x^{k,3}]$ となり、この区間の幅は $k \to \infty$ としたときに 0 に収束する。そして、その収束した先が最小解となる。

ところで、このアルゴリズムの各反復 k において、$x^{k,4}$ をどの点に求めるかは説明していなかった。特にこうすればアルゴリズムが速くなるという指針はないが、4 つの点 $x^{k,0}, x^{k,1}, x^{k,2}, x^{k,3}$ が、毎回等比率で並んでいる, つまり、

$$x^{k,1} - x^{k,0} : x^{k,2} - x^{k,1} : x^{k,3} - x^{k,2} = 1 : \alpha : 1$$

となるような α が存在すれば美しい (図 6.3 左)。

ここで、まず、$f(x^{k,1}) > f(x^{k,2})$ のときの $x^{k,4}$ を求める方法を説明する。このとき、区間 $[x^{k,0}, x^{k,3}]$ の長さを β とし、区間 $[x^{k,0}, x^{k,2}]$(および区間 $[x^{k,1}, x^{k,3}]$) の長さを $\gamma := \frac{1+\alpha}{2+\alpha}\beta$ とし, $\tau := \beta - \gamma$ とする (図 6.3 右)。そして、

$$x^{k,4} := x^{k,1} + \tau$$

とする。(α の決め方によっては、$x^{k,4} \notin [x^{k,2}, x^{k,3}]$ にならないこともある。) 今、$k+1$ の反復点でも比率が保たれるには、

$$\begin{aligned} \beta : \gamma &= x^{k+1,3} - x^{k+1,0} : x^{k+1,2} - x^{k+1,0} \\ &= x^{k,3} - x^{k,1} : x^{k,4} - x^{k,1} \\ &= \gamma : \tau \end{aligned}$$

が成り立たなければならない。ここで、$\tau = \beta - \gamma$ を代入すると、

$$\frac{\beta}{\gamma} = \frac{1+\sqrt{5}}{2}$$

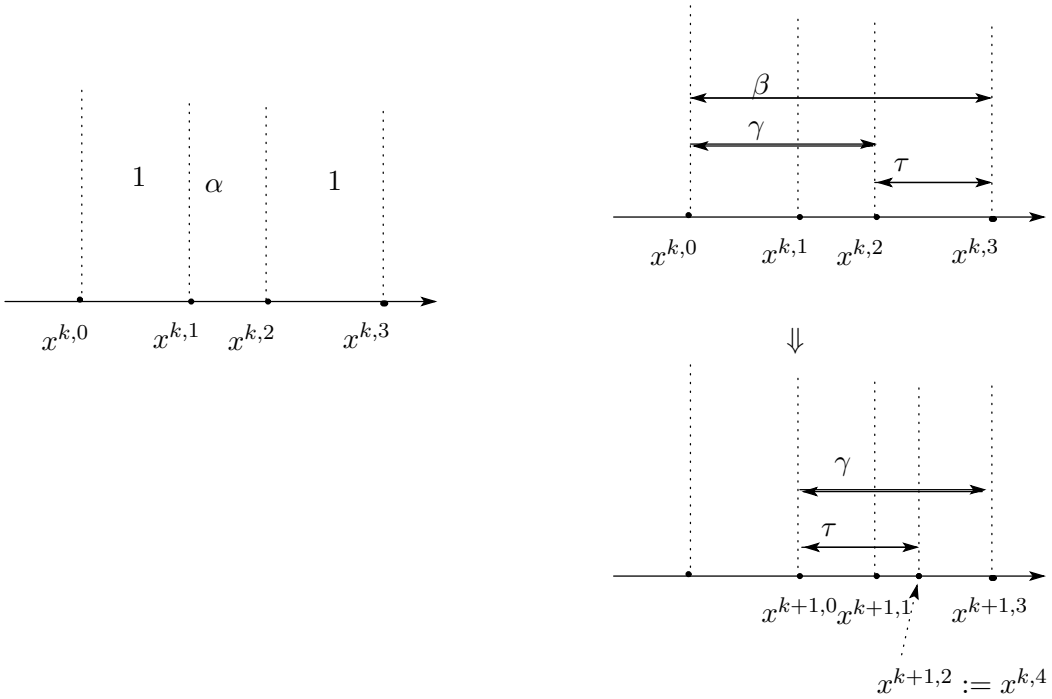


図 6.3: 黄金分割比

を得る。この比率を黄金分割比と呼ぶ。なお、このとき、$\alpha = \frac{3-\sqrt{5}}{\sqrt{5}-1}$ となり、$x^{k,4} \in [x^{k,2}, x^{k,3}]$ となる。

$f(x^{k,1}) < f(x^{k,2})$ のときも, $x^{k,4}$ は黄金分割比を用いて $x^{k,4} := x^{k,3} - \tau$ とすることによって求められる。

以上のことをまとめると、黄金分割法は以下のようになる。

Algorithm 6.1.1 ステップ 0：初期化　$\eta := (\sqrt{5}-1)/(\sqrt{5}+1)$, $\beta^0 := u - l$, $\tau^0 := \eta\beta^0$ とする。$x^{0,0} := l, x^{0,1} = l + \tau^0, x^{0,2} := u - \tau^0, x^{0,3} := u$ とする。$k := 0$ とする。

ステップ 1：終了条件　もし β^k が十分近ければ終了する。

ステップ 2：反復点の生成 $f(x^{k,1}) < f(x^{k,2})$ であれば、

$$x^{k+1,0} := x^{k,0},\ \ x^{k+1,1} := x^{k,2} - \tau^k,\ \ x^{k+1,2} := x^{k,1},\ \ x^{k+1,3} := x^{k,2}$$

とし、そうでなければ、

$$x^{k+1,0} := x^{k,1},\ \ x^{k+1,1} := x^{k,2},\ \ x^{k+1,2} := x^{k,1} + \tau^k,\ \ x^{k+1,3} := x^{k,3}$$

とする。$\beta^{k+1} := \tau^k$, $\tau^{k+1} := \eta\beta^{k+1}$ とする。$k := k+1$ としてステップ *1* へ。

このアルゴリズムの速さは黄金分割比によって定まっている。実際、k 回目の反復のときに探索している区間の長さは β^k であり、$k+1$ 回目の区間の長さは $\gamma^k := \beta^k - \tau^k$ である。よって、区間の長さは、黄金分割比の逆数の割合だけ短くなっていく。もし、$\beta^k < \varepsilon$ を終了条件にすれば、終了までかかる反復回数は,

$$\left(\frac{2}{\sqrt{5}+1}\right)^k \beta^0 < \varepsilon$$

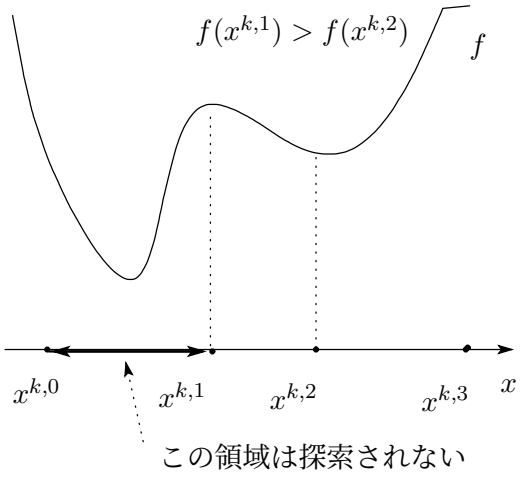


図 6.4: 凸関数でない場合

となり、

$$k > \log\left(\frac{\sqrt{5}+1}{2}\right)\log\left(\frac{\beta^0}{\varepsilon}\right)$$

を満たす最初の k でアルゴリズムは終了する。なお、この反復回数は終了するために必ずかかる回数であり、これより速くなることはない。これは後の章で紹介する微分を使った手法に比べると、大きな欠点である。

今までは、目的関数 f が凸関数であることを仮定していた。それでは、凸関数でないときにはどうなるのであろうか？図 6.4 にあるように、f が凸関数でない場合は、最小解がある区間を棄ててしまい、それ以外の区間の探索をすることがありえる。このため、凸関数でない場合は、黄金分割法は (局所的) 最小解を見つけられないこともある。

6.2 直接探索法

この節では、n 次元の決定変数を持つ目的関数 $f : R^n \to R$ の制約なし最小化問題

$$\min \quad f(x)$$

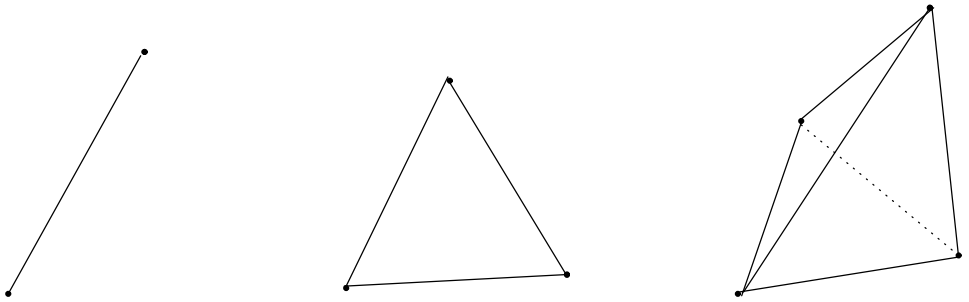
に対する、直接探索法の説明をする。黄金分割法と同様に、直接探索法は目的関数の値しか用いない反復法である。

直接探索法では、最小解に収束するような n 次元単体を生成する手法である．このため、直接探索法は、単体法 (シンプレックス法) と呼ばれることもある。ここで，単体とは，$n+1$ 個の点で表された体積を持つ[1] 物体である．具体的には，1 次元では半直線を表し，2 次元では 3 角形，3 次元では 4 面体.... である (図 6.5).

それでは、直接探索法において、現在与えられている単体から次の単体を生成する方法を説明しよう。まず，n 次元単体、つまり、$n+1$ 個の点が $x^0, x^1, x^2, \ldots, x^n$ が与えられているとする．このとき，便宜上

$$f(x^0) \le f(x^1) \le f(x^2) \le \ldots \le f(x^n)$$

[1] n このベクトル $x^i - x^0, i = 1, \ldots, n$ が 1 次独立になることと等価である



1 次元単体 (半直線)　　2 次元単体 (3 角形)　　3 次元単体 (4 面体)

図 6.5: 単体の例

と番号づけされているとする．さらに、この中で最小点と最大点を明確にするために $x^{\max} := x^n$, $x^{\min} := x^0$ とする。直接探索法では，一番悪い点 $x^{\max}$ を動かすことによって，新しい点 x^{new} を生成することを行なう．このとき、x^{new} は次のような性質を持つことが望ましい。

性質 (i) 新しい点の集合 $\{x^0, x^1, \cdots, x^{n-1}, x^{\text{new}}\}$ が単体になる。

性質 (ii) 単体全体 (単体を構成する全ての点) の関数値の減少する。

性質 (iii) 単体の体積の縮小する。

性質 (i) は、直接探索法において、必ず必要とされる性質である。性質 (ii) は、最小化が成功していることを意味し、性質 (ii) は、直接探索法が収束へ向かっていることを意味している。つまり、ある程度の単体の体積が小さくなったとき、このアルゴリズムによって、最小化問題の”解”が求まっていることが期待できる。

$x^{\max}$ の目的関数値はそれ以外の単体の点 $\{x^0, x^1, \cdots, x^{n-1}\}$ よりも大きいので，他の点 $\{x^0, x^1, \cdots, x^{n-1}\}$ の方向に動かせば，目的関数値が $x^{\max}$ よりも小さくなる x^{new} が見つかりそうである．そこで，$x^{\max}$ を他の点 $\{x^0, x^1, \cdots, x^{n-1}\}$ の平均点 x^c

$$x^c = \frac{1}{n}\sum_{i=0}^{n-1} x^i$$

の方向へ動かすことを考える。これは、他の点で構成される平面の反対側に映すことを意味している (図 6.6)．x^n を x^c の反対側に移すわけだから，反対側に移された点を反射点と呼ぶ。この反射点を x^{ref} とすると、

$$x^{\text{ref}} = 2x^c - x^{\max}$$

となる．これで，$f(x^{\text{ref}}) < f(x^{\max})$ になれば, 次の点 x^{new} を反射点 x^{ref} とする。このとき、点 $\{x^0, x^1, \ldots, x^{n-1}, x^{\text{new}}\}$ もまた、単体 (体積がある) になっている。また、この新しい単体の点の関数値の和は、前の単体の点の関数値の和よりも小さくなっている。つまり、性質 (i), (ii) が成り立つ。しかし、新しい単体の体積は、前の単体の体積と同じになるため、性質 (iii) は成り立たない。

ここで、もし、x^{ref} がすべての点の関数値よりも小さいとき、つまり、$f(x^{\text{ref}}) < f(x^{\min})$ のときは、もっと先に行った方がよりよい点 x^{new} を見つけられる可能性がある。そこで、そのときは、その先の点も調べてみる。これは、反射点をさらにその先に伸ばした点で、拡張点と言う．拡張点を x^{exp} とすると、

$$x^{\text{exp}} = 3(x^c - x^{\max}) - x^{\max} = 3x^c - 4x^{\max}$$

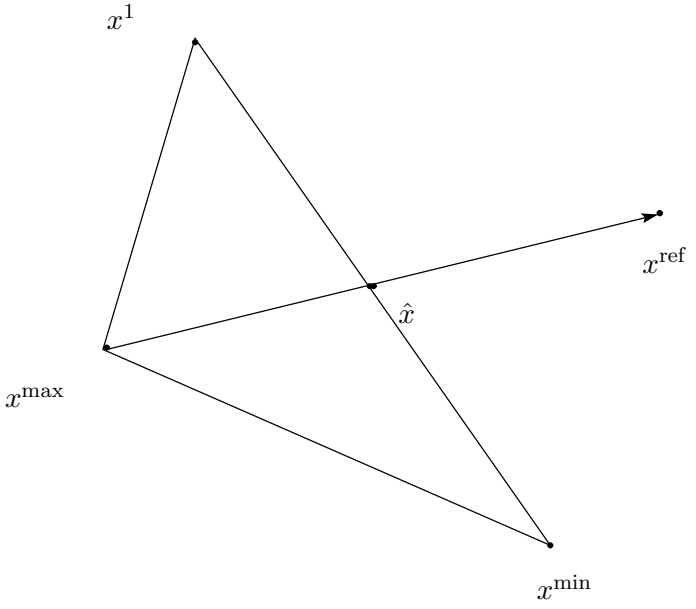


図 6.6: 反射点

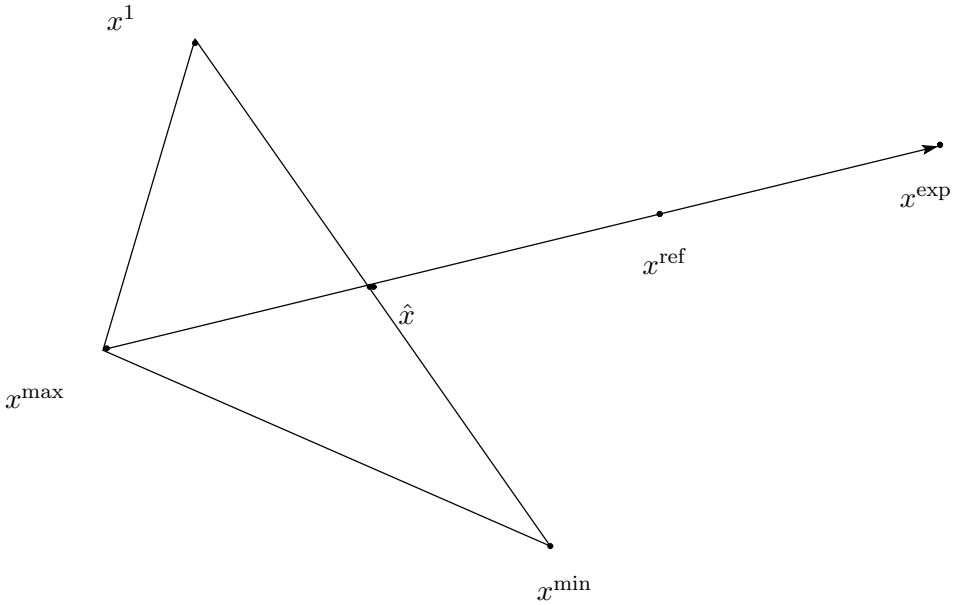


図 6.7: 拡張点

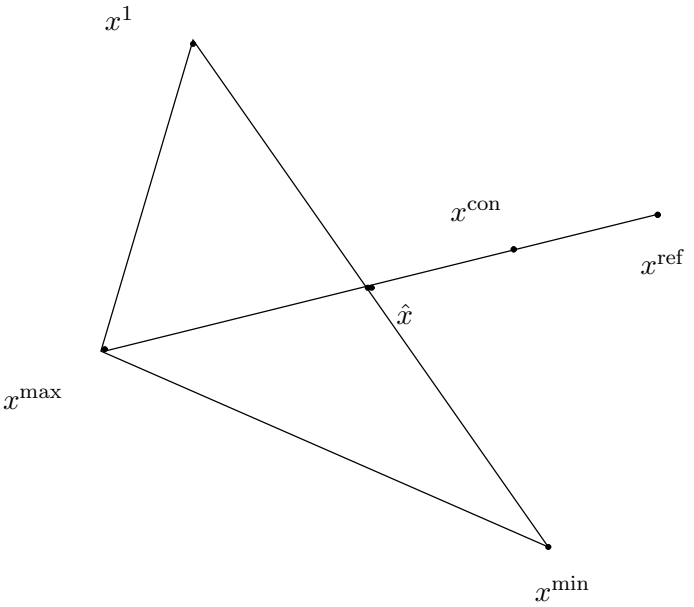


図 6.8: 縮小点

となる (図 6.7)。そして、$f(x^{\mathrm{exp}}) < f(x^{\mathrm{ref}})$ であれば、$x^{\mathrm{new}} = x^{\mathrm{exp}}$ とする。このときも、反射点の場合と同様、性質 (i), (ii) が成り立つ。しかしながら、新しい単体の体積は前の単体の体積に比べて大きくなっている。

今までの議論は、点 x^{ref} が良い点、つまり $f(x^{\mathrm{ref}}) < f(x^{\mathrm{max}})$ を満たしているときであった。しかしながら、必ずしもその不等式が成り立つわけではない。そのときは、x^c はよい点である可能性が高いので，x^{ref} を x^c に近づければ，関数値が下がることが期待できる．そのような点が，縮小点とよび、x^{con} と表すことにする．ここでは、x^{con} は、x^{ref} と x^c の中点にあるとして、

$$x^{\mathrm{con}} := \frac{1}{2}(x^{\mathrm{ref}} + x^c) = \frac{1}{2}(3x^c - x^{\mathrm{max}})$$

とする (図 6.8)。ここで、もし、$f(x^{\mathrm{con}}) < f(x^{\mathrm{max}})$ であれば、次の点 $x^{\mathrm{new}} = x^{\mathrm{ref}}$ を x^{max} の代わりにすればよい。このとき、性質 (i),(ii),(iii) の全てが成り立つことに注意しよう。

反射点 x^{ref} も縮小点 x^{con} も関数値が x^{max} の関数値より小さくならない場合もある．例えば、図 6.9 のように等高線が与えられているとき、点 x^{max} から x^c の方向に進んでも関数値は増えるだけであり、これまでの手法では単体の関数値を減らすことが期待できない。そのようなときは，x^{max} の移動だけではなく、単体全体の移動を考える．現在、x^{min} が単体の端点の中で、一番、関数値の小さいわけであるから，すべての点をその点に近づければ、単体の点の関数値の和は小さくできそうである．そうすることによって，図 6.9 のように x^c 付近に存在する”山の稜線”から遠ざかることが期待できる．そのような操作を単体縮小と呼ぶ．これは、すべての点をその点と x^{min} との中点に移してやればよい (図 6.9)。(このとき、x^0 は動かないことに注意)

$$\bar{x}^i := \frac{1}{2}(x^i + x^{\mathrm{min}})$$

この操作によって、生成された新しい点 $\{\bar{x}^0, \ldots, \bar{x}^n\}$ は、性質 (i) と (iii) を満たす。

このように、反射点、拡張点、縮小点、単体縮小、のどれかを繰り返すと、性質 (i) は必ず満たされ、性質 (ii) か性質 (iii) のどちらかがおきることが期待できる。

これまでに説明した手法をまとめると，直接探索法は以下のように記述できる．

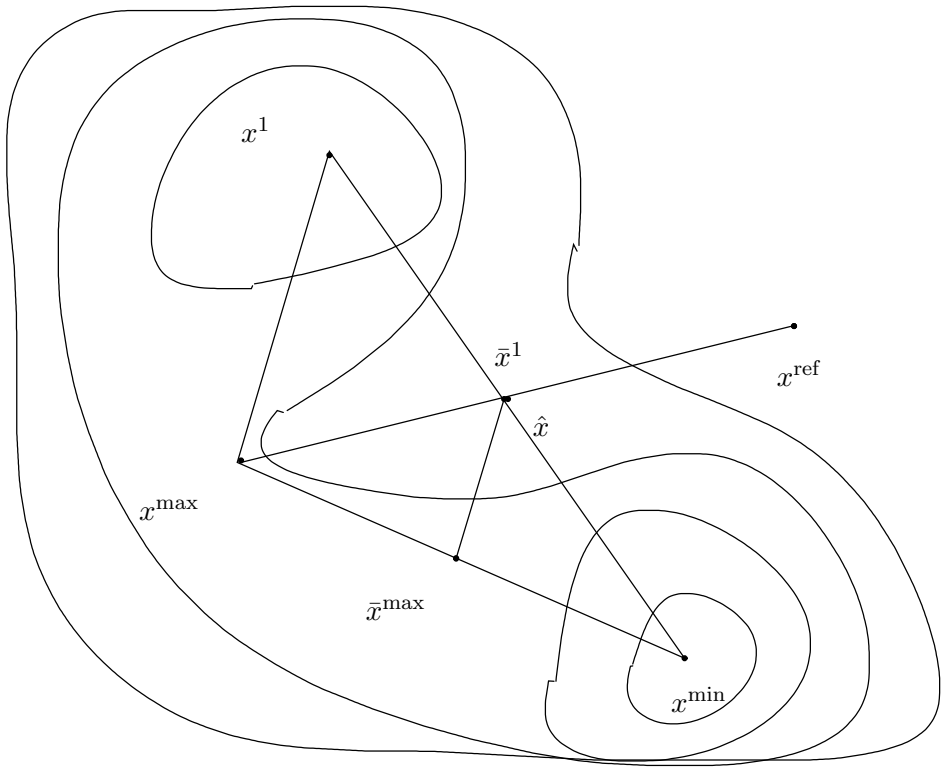


図 6.9: 単体縮小

Algorithm 6.2.1 ステップ 0：初期化　n 次元単体を構成する端点 $[x^0, x^1, \ldots, x^n]$ を選ぶ。

ステップ 1：並び替え

$$f(x^0) \leq f(x^1) \leq f(x^2) \leq \ldots \leq f(x^n)$$

となるよう端点の並び替えを行う。$x^{\min} = x^0, x^{\max} = x^n$ とする。

ステップ 1：反射 $x^c = \frac{1}{n}\sum_{i=0}^{n-1} x^i$ を求め、反射点 $x^{\mathrm{ref}} = 2x^c - x^{\max}$ を計算する。もし、$f(x^{\mathrm{ref}}) \leq f(x^{\max})$ ならばステップ 2 へ。そうでなければステップ 3 へ。

ステップ 2：拡張　拡張点 $x^{\exp} = 3x^c - 4x^{\max}$ を求める。もし、$f(x^{\mathrm{ref}}) \leq f(x^{\exp})$ であれば、x^n を x^{ref} に置き換えて、ステップ *1* へ。そうでなければ、x^n を $x^{\exp}$ に置き換えて、ステップ *1* へ。

ステップ 3：縮小 縮小点 $x^{\mathrm{con}} := \frac{1}{2}(3x^c - x^{\max})$ を求める。もし、$f(x^{\mathrm{con}}) \leq f(x^{\max})$ であれば、x^n を x^{con} に置き換えて、ステップ *1* へ。そうでなければ、ステップ *4* へ。

ステップ 4：単体縮小　単体縮小

$$x^i := \frac{1}{2}(x^i + x^{\min})$$

を行い、ステップ *1* へ。

直接探索法は、関数値を計算するだけで実行できる非常に単純なアルゴリズムである。微分やヘシアンなどの計算もする必要もないので、計算機に実装することは非常に容易である。また、アルゴリズムが簡単なことから、微分やヘシアンが存在しない (または計算できない) 問題や、連続ですらない問題に対しても適応できる。

しかしながら、収束性の観点から言えば、大きな問題がある。まず、局所的最小値にすら収束することが保証されていない。例えば、次の問題を考えよう。

$$\min \ \ \sin(x)$$

このとき、明らかに、任意の整数 i に対して、$x = \frac{(4i-1)\pi}{2}$ がこの問題の解となり、そのときの目的関数値は、-1 である。一方、この問題の初期単体 (1 次元であるため線分) として、$[0, \frac{\pi}{2}]$ を与える。このとき直接探索法を用いると、絶えず拡張点を求めることになり、単体は収束しない。また、関数値は 0 のままである。そこでこのような欠点を補うために、いくつかの修正方法が提案されている。その中のひとつに乱数を用いる方法がある。反射点 x^{ref} は、

$$x^{\mathrm{ref}} = x^{\max} + 2(x^c - x^{\max})$$

と変形すると、$x^{\max}$ から探索方向 $x^c - x^{\max}$、ステップサイズ 2 で移った点と考えることができる。このとき、ステップサイズ 2 の代わりに、$[1, 2]$ の間でランダムに選ばれた t を用いることが考えられる。このとき、先ほどの問題で収束しない欠点は解消することができる。ただし、このようにしても、大域的収束性は保証されない。これは、直接探索法が、目的関数の微分等の情報を用いていないからである。次回から、微分を用いた大域的収束が保証されたアルゴリズムを紹介する。