

目的

与えられた関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ の最小値または最大値を与える点 $x \in \mathbb{R}^n$ を求める手法を学び、実際にプログラミング言語Pythonを用いた演習を行うことによって、その手法を理解する。特に、最適化の手法の根幹をなす、最急降下法、ニュートン法、射影勾配法を学ぶ。

アルゴリズム

目的関数の最小化問題のアルゴリズムである降下法の主な流れは、関数の値を減らす点列 $\{x^k\}$ を x^k が停留点となるまで生成していくことである。点列 $\{x^k\}$ は $x^{k+1} := x^k + t_k d^k$ で定義される。

d^k は降下方向と呼ばれ、条件 $\langle d^k, \nabla f(x^k) \rangle < 0$ を満たしていればよい。

t^k はステップサイズと呼ばれ、 t^k を決めるときにアルミホ条件 $f(x^k + t d^k) \leq f(x^k) + \xi t \langle d^k, \nabla f(x^k) \rangle$ を用いたバックトラック法を用いる。このバックトラック法には2つあり、1つ目は $\rho = 0.5$ に固定したバックトラック法、2つ目は二次補間法によるバックトラック法である。この2つについて説明する。

バックトラック法 ($\rho = 0.5$)

- (0) $\xi, \rho \in (0, 1)$, 初期ステップサイズ \bar{t} を決める。 $t := \bar{t}$ とおく。
- (1) アルミホ条件 $f(x^k + t d^k) \leq f(x^k) + \xi t \langle d^k, \nabla f(x^k) \rangle$ を満たしているならば、 $t_k := t$ として終了。
- (2) $t := \rho t$ としてステップ (1) に戻る。

バックトラック法 (二次補間法)

二次補間法では新たに関数 $\phi(t) := f(x^k + t d^k)$ を定義し、この関数を用いてステップサイズを計算する。

- (0) $\xi, \rho \in (0, 1)$, 初期ステップサイズ \bar{t} を決める。 $t^{(0)} := \bar{t}$, $m := 0$ とおく。
- (1) アルミホ条件 $f(x^k + t^{(m)} d^k) \leq f(x^k) + \xi t^{(m)} \langle d^k, \nabla f(x^k) \rangle$ を満たしているならば、 $t_k := t^{(m)}$ として終了
- (2) $t^{(m+1)} := -\frac{\phi'(0)(t^{(m)})^2}{2(\phi(t^{(m)}) - \phi(0) - \phi'(0)t^{(m)})}$ とする。ただし、 $t^{(m+1)} < 0.1t^{(m)}$ 又は $t^{(m+1)} > 0.9t^{(m)}$ ならば $t^{(m+1)} := \frac{t^{(m)}}{2}$ とする。ステップ (1) に戻る。

続いて、点列を生成するアルゴリズムについて説明する。アルゴリズムの主な流れは、降下方向を決定し、バックトラック法を用いてステップサイズを求めることである。降下方向の選び方は複数あり、この選び方によって異なるアルゴリズムになる。ここでは、最急降下法とニュートン法について説明する。

最急降下法

最急降下法では降下方向 d^k を $d^k := -\nabla f(x^k)$ で定める。最急降下法の流れは以下である。

- (0) バックトラック法のパラメータ $\xi, \rho \in (0, 1)$ と初期点 x^0 を選び、 $k := 0$ とおく。
- (1) $\|\nabla f(x^k)\| \leq \varepsilon$ ならば、 x^k を解として計算終了。
- (2) $d^k := -\nabla f(x^k)$ とおく。

- (3) バックトラック法を用いて, t_k を求める.
- (4) $x^{k+1} := x^k + t_k d^k$, $k := k + 1$ としてステップ (1) へ戻る.

ニュートン法

ニュートン法では $\nabla^2 f(x)^{-1}$ が一様に正定値であるとき, 降下方向 d^k を $d^k := -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$ で定める. $\nabla^2 f(x)^{-1}$ が正定値でないときは, $B_k := \nabla^2 f(x^k) + \tau I$ (I は単位行列) が正定値となるように $\tau > 0$ をおいて, 降下方向 d^k を $d^k := -B_k^{-1} \nabla f(x^k)$ で定める. 以上を踏まえてニュートン法の流れは以下である.

- (0) バックトラック法のパラメータ $\xi, \rho \in (0, 1)$ と初期点 x^0 を選び, $k := 0$ とおく.
- (1) $\|\nabla f(x^k)\| \leq \varepsilon$ ならば, x^k を解として計算終了.
- (2) $B_k := \nabla^2 f(x^k) + \tau I$ として, $d^k := -B_k^{-1} \nabla f(x^k)$ とおく. ただし, $\nabla^2 f(x)^{-1}$ が一様に正定値ならば, $\tau = 0$ とし, $\nabla^2 f(x)^{-1}$ が正定値でないならば, B_k が正定値となるまで $\tau = 2, 4, 8, 16, 32 \dots$ と増やす.
- (3) バックトラック法を用いて, t_k を求める.
- (4) $x^{k+1} := x^k + t_k d^k$, $k := k + 1$ としてステップ (1) へ戻る.

ここまで説明してきた最急降下法とニュートン法は制約付き問題のときには用いることができない. 続いて, 制約付き問題で用いられるアルゴリズムである射影勾配法について説明する.

射影勾配法

ここでは簡単のために制約集合 X を上下限制約 $X = \{x \in \mathbb{R}^n : l \leq x \leq u\}$ とする. 降下方向 d^k は $d^k := P_X(x^k - s_k \nabla f(x^k)) - x^k$ で定める. ただし, $P_X(x)$ とは x の X への射影であり, X が上下限制約であれば, $P_X(x) = \max\{l, \min\{x, u\}\}$ で求まる. また, $s_k > 0$ である. 降下方向 d^k が 0 であることと x^k が停留点であることは等価であるので終了条件は $\|d^k\| \leq \varepsilon$ とする. また, 点列が十分更新されないことがあるので, 十分更新されてないときに計算終了する. 以上を踏まえて射影勾配法の流れは以下である.

- (0) バックトラック法のパラメータ $\xi, \rho \in (0, 1)$ と初期点 x^0 を選び, $k := 0$ とおく.
- (1) $s_k > 0$ を選び, $d^k := P_X(x^k - s_k \nabla f(x^k)) - x^k = \max\{l, \min\{x^k - s_k \nabla f(x^k), u\}\}$ とおく.
- (2) $\|d^k\| \leq \varepsilon$ ならば, x^k を解として計算終了.
- (3) バックトラック法を用いて, t_k を求める.
- (4) $t_k \|d^k\|_\infty \leq \varepsilon_{\text{small}} \max\{1, \|x^k\|_\infty\}$ を満たすとき (点列が十分更新されないとき), x^k を解として計算終了. ただし, $\varepsilon_{\text{small}} := 10^{-16}$, $\|z\|_\infty := \sup_{i=0, \dots, n-1} |z_i|$ である.
- (5) $x^{k+1} := x^k + t_k d^k$, $k := k + 1$ としてステップ (1) へ戻る.

パラメータ

各アルゴリズムにおいていくつかのパラメータが存在する. 今回はバックトラック法でのパラメータ ξ, ρ, \bar{t} をそれぞれ $\xi := 10^{-4}$, $\rho := 0.5$, $\bar{t} = 1$ とし, 終了条件でのパラメータ ε を $\varepsilon := n \times 10^{-6}$ (n は各課題における次元数) と定める. ε に次元数を掛けることで, 次元が大きいほど誤差が大きくなることを許容している. これらのパラメータをこのように定めたのは一般的に用いられている数値だからである. また, 射影勾配法でのパラメータ s_k を $s_k := 1$ と定める. このように定めた理由は簡単にするためである.

また, 初期点 x_0 を課題 3 では $x_0 := (1, 1)^\top$, $(10, -20)^\top$, $(-36, 114)^\top$, 課題 4 では $x_0 := (1, 1)^\top$, 課題 5 で

は全ての要素を 1, 課題 6 では $x_0 := (0, 0)^\top$, 課題 7 では $x_0 := (2, 1, -2, -1)^\top$ とする. 次元数 n を課題 5 では $n = 2, 5, 10, 20$, 課題 7 では $n = 4$ とする. 初期点や次元数をこのように定めた理由は数理工学実験冊子で指定されているからである.

問題の性質

今回取り扱う 5 つの最小化問題について説明する. ただし, 課題 5 と課題 7 については既に数理工学実験冊子に記載されているので, ここでは省略する.

課題 3

初期点: $x_0 := (1, 1)^\top, (10, -20)^\top, (-36, 114)^\top$

目的関数: $f(x) := x_0^2 + e^{x_0} + x_1^4 + x_1^2 - 2x_0x_1 + 3$

勾配: $\nabla f(x) = \begin{pmatrix} 2x_0 + e^{x_0} - 2x_1 \\ 4x_1^3 + 2x_1 - 2x_0 \end{pmatrix}$

ヘッセ行列: $\nabla^2 f(x) = \begin{pmatrix} 2 + e^{x_0} & -2 \\ -2 & 12x_1^2 + 2 \end{pmatrix}$

課題 4

初期点: $x_0 := (1, 1)^\top$

目的関数: $f(x) := \sum_{i=0}^2 f_i(x)^2$ ただし, $f_i(x) := y_i - x_0(1 - x_1^{i+1})$, $y_0 = 1.5, y_1 = 2.25, y_2 = 2.625$

勾配: $\nabla f(x) = 2 \sum_{i=0}^2 f_i(x) \nabla f_i(x)$ ただし, $\nabla f_i(x) = \begin{pmatrix} x_1^{i+1} - 1 \\ (i+1)x_0x_1^i \end{pmatrix}$

ヘッセ行列: $\nabla^2 f(x) = 2 \sum_{i=0}^2 (f_i(x) \nabla^2 f_i(x) + \nabla f_i(x) \nabla f_i(x)^\top)$ ただし, $\nabla^2 f_i(x) = \begin{pmatrix} 0 & (i+1)x_1^i \\ (i+1)x_1^i & i(i+1)x_0x_1^{i-1} \end{pmatrix}$

課題 6

初期点: $x_0 := (0, 0)^\top$

目的関数: $f(x) := \left(x_1 - \frac{5.1x_0^2}{4\pi^2} + \frac{5x_0}{\pi} - 6 \right)^2 + 10 \left(1 - \frac{1}{8\pi} \right) \cos x_0 + 10$

制約条件: $-5 \leq x_0 \leq 10, 0 \leq x_1 \leq 15$

勾配: $\nabla f(x) = \begin{pmatrix} 2 \left(x_1 - \frac{5.1x_0^2}{4\pi^2} + \frac{5x_0}{\pi} - 6 \right) \left(-\frac{5.1x_0}{2\pi^2} + \frac{5}{\pi} \right) - 10 \left(1 - \frac{1}{8\pi} \right) \sin x_0 \\ 2 \left(x_1 - \frac{5.1x_0^2}{4\pi^2} + \frac{5x_0}{\pi} - 6 \right) \end{pmatrix}$

各課題の実験結果

課題 3

3 つの初期点 $(1, 1)^\top, (10, -20)^\top, (-36, 114)^\top$ における各手法・直線探索での実験結果をまとめた表が表 1 である. なお, 計算が終了した理由は「 $\|\nabla f(x^k)\| \leq \varepsilon$ 」である.

初期点	手法	直線探索の 手段	反復回数	目的関数の 評価回数	最後の 3 回の $\ \nabla f(x^k)\ $	最適値	目的解
$(1, 1)^\top$	最急降下法	$\rho = 0.5$	30	114	(4.791e-06, 2.470e-06, 1.281e-06)	3.597	$(-0.7335, -0.4933)^\top$
		二次補間法	24	68	(6.051e-06, 2.050e-06, 1.676e-06)	3.597	$(-0.7335, -0.4933)^\top$
	ニュートン法	$\rho = 0.5$	6	12	(2.987e-03, 2.593e-06, 3.619e-12)	3.597	$(-0.7335, -0.4933)^\top$
		二次補間法	6	12	(2.987e-03, 2.593e-06, 3.619e-12)	3.597	$(-0.7335, -0.4933)^\top$
$(10, -20)^\top$	最急降下法	$\rho = 0.5$	36	166	(4.367e-06, 2.250e-06, 1.166e-06)	3.597	$(-0.7335, -0.4933)^\top$
		二次補間法	28	100	(4.652e-06, 3.603e-06, 1.683e-06)	3.597	$(-0.7335, -0.4933)^\top$
	ニュートン法	$\rho = 0.5$	14	29	(4.546e-02, 8.112e-04, 3.477e-07)	3.597	$(-0.7335, -0.4933)^\top$
		二次補間法	15	31	(4.053e-02, 8.099e-04, 3.405e-07)	3.597	$(-0.7335, -0.4933)^\top$
$(-36, 114)^\top$	最急降下法	$\rho = 0.5$	41	206	(2.227e-06, 3.027e-06, 1.562e-06)	3.597	$(-0.7335, -0.4933)^\top$
		二次補間法	35	142	(3.285e-06, 2.354e-06, 1.279e-06)	3.597	$(-0.7335, -0.4933)^\top$
	ニュートン法	$\rho = 0.5$	20	44	(9.805e-03, 4.951e-05, 1.274e-09)	3.597	$(-0.7335, -0.4933)^\top$
		二次補間法	23	53	(2.563e-03, 3.425e-06, 6.095e-12)	3.597	$(-0.7335, -0.4933)^\top$

表 1 課題 3 の問題:実験結果

課題 4

各手法・直線探索での実験結果をまとめた表が表 2 である.なお,計算が終了した理由はいずれも「 $\|\nabla f(x^k)\| \leq \varepsilon$ 」である.

手法	直線探索の手段	反復回数	目的関数の評価回数	最後の 3 回の $\ \nabla f(x^k)\ $	最適値	目的解
最急降下法	$\rho = 0.5$	740	4811	(4.113e-06, 2.466e-06, 1.733e-06)	3.019e-12	$(3.000, 0.5000)^\top$
	二次補間法	453	2054	(2.085e-06, 4.924e-06, 1.983e-06)	5.608e-12	$(3.000, 0.5000)^\top$
ニュートン法	$\rho = 0.5$	7	14	(5.942e-04, 1.173e-05, 3.743e-11)	4.842e-22	$(3.000, 0.5000)^\top$
	二次補間法	7	14	(5.943e-04, 1.173e-05, 3.743e-11)	4.842e-22	$(3.000, 0.5000)^\top$

表 2 課題 4 の問題:実験結果

課題 6

各手法・直線探索での実験結果をまとめた表が表 3 である.なお,計算が終了した理由はいずれも「 $\|d^k\| \leq \varepsilon$ 」である.

手法	直線探索の手段	反復回数	目的関数の評価回数	最後の 3 回の $\ d^k\ $	最適値	目的解
射影勾配法	$\rho = 0.5$	45	203	(2.136e-06, 2.907e-06, 1.253e-06)	0.3979	$(3.142, 2.275)^\top$
	二次補間法	39	118	(4.003e-06, 2.006e-06, 1.926e-06)	0.3979	$(3.142, 2.275)^\top$

表 3 課題 6 の問題:実験結果

課題 7

各手法・直線探索での実験結果をまとめた表が表 4 である.なお,計算が終了した理由はいずれも「点列が十分更新されないこと」である.

手法	直線探索の手段	反復回数	目的関数の評価回数	最後の 3 回の $\ d^k\ $	最適値	目的解
射影勾配法	$\rho = 0.5$	30	787	(2.000, 2.000, 2.000)	4.441e-16	(-1.176e-16, -1.166e-16, 1.176e-16, 1.166e-16) ^T
	二次補間法	18	219	(2.000, 2.000, 2.000)	4.441e-16	(1.995e-16, -2.613e-18, -1.995e-16, 2.613e-18) ^T

表 4 課題 7 の問題:実験結果

課題 5

4 つの次元数における各手法・直線探索での 5 回の計算の反復回数の中央値をまとめた表が表 5 である.なお,計算が終了した理由はいずれも「 $\|\nabla f(x^k)\| \leq \varepsilon$ 」である.

表 5 を見ればわかるように最急降下法は次元が大きいくほど反復回数も大きくなっている.また,ニュートン法ではどの次元数でも 2 回目の反復で計算が終了している.この理由を以下に述べる.

(理由)

$\nabla^2 f(x^0) = A$ が正定値であるので,降下方向は

$$d^0 = -\nabla^2 f(x^0)^{-1} \nabla f(x^0) = -A^{-1} A x^0 = -x^0$$

で定まる.バクトラック法では, $t = \bar{t} = 1$ においてアルミホ条件の左辺は

$$f(x_0 + t d^0) = f(x_0 - x_0) = f(0) = \frac{1}{2} \langle 0, 0 \rangle = 0$$

一方,アルミホ条件の右辺は

$$\begin{aligned}
 f(x^0) + \xi t \langle d^0, \nabla f(x^0) \rangle &= f(x^0) + 10^{-4} \times \langle -x^0, A x^0 \rangle \\
 &= \frac{1}{2} \langle A x^0, x^0 \rangle - 10^{-4} \times \langle A x^0, x^0 \rangle \quad [\because A \text{ が対称行列}] \\
 &= (0.5 - 10^{-4}) \langle A x^0, x^0 \rangle \\
 &\geq 0 \quad [\because A \text{ が正定値なので } \langle A x^0, x^0 \rangle \geq 0]
 \end{aligned}$$

以上から $t = 1$ でアルミホ条件が満たされるのでステップサイズは $t_0 = 1$ となる.すると更新された点列は $x^1 = x^0 + t_0 d^0 = x^0 - x^0 = 0$ となり, $d^1 = -A^{-1} A x^1 = -x^1 = 0$ で終了条件 $\|d^1\| \leq \varepsilon$ が満たされるので,計算を終了する.よって,ニュートン法では 2 回目の反復で計算が終了する.

手法	直線探索の手段	次元	反復回数の中央値
最急降下法	$\rho = 0.5$	2	195
		5	1210
		10	20072
		20	54200
	二次補間法	2	123
		5	611
		10	1018
		20	4478
ニュートン法	$\rho = 0.5$	2	1
		5	1
		10	1
		20	1
	二次補間法	2	1
		5	1
		10	1
		20	1

表 5 課題 5 の問題:実験結果

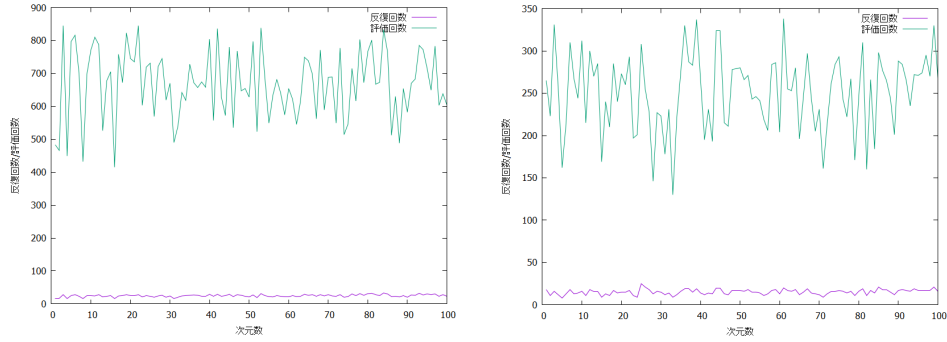
手法の比較

実験結果から分かるように,最急降下法とニュートン法を比較すると反復回数や目的関数の評価回数はニュートン法の方が少ない.また,直線探索について, $\rho = 0.5$ で固定した場合と二次補間法を比較すると,反復回数や目的関数の評価回数はほとんどの場合で二次補間法の方が少なくなっているが,ニュートン法の一部において $\rho = 0.5$ に固定した場合の方が少なくなっていることがある.これは二次補間法で $t^{(m+1)}$ と $t^{(m)}$ の差が小さすぎたり,大きすぎたりするときに点列の更新が適切に進められないからである.このようなことを避けるために「 $t^{(m+1)} < 0.1t^{(m)}$ 又は $t^{(m+1)} > 0.9t^{(m)}$ ならば $t^{(m+1)} := \frac{t^{(m)}}{2}$ 」とする補正を行ったが,例えば $t^{(m+1)} = 0.89t^{(m)}$ だったりした場合,点列の更新が適切に進められない.このような理由で二次補間法のほうが反復回数や評価回数が多い場合がある.

課題 8

課題 7 の問題について,反復回数や目的関数の評価回数が次元数に依存するかどうかを調べる.次元数を 1 から 100 まで増やしながら,反復回数と評価回数を記録した.なお,初期点は平等にするため,かつ,大域的最小解である原点に収束するためにすべての要素を $\frac{1}{3}$ とした.いずれの次元数においても,計算が終了した理由は「点列が十分更新されないこと」である.これらの結果を表したグラフが図 1 である.

図 1 から分かるように,バラつきはあるが反復回数,評価回数ともに次元が増えるほど増えたりすることはなく,ほぼ一定である.



$\rho = 0.5$

二次補間法

図1 課題8の問題:各直線探索における次元数に対する反復回数(紫線)と評価回数(緑線)の関係

課題9

今回実装した射影勾配法は降下方向を $d^k := P_X(x^k - s_k \nabla f(x^k)) - x^k$ と定めた。これは、まず最急降下法と同様に点 x^k から勾配の逆方向 $-s_k \nabla f(x^k)$ に進み、その点を制約集合 X に射影することで実行可能にしているのである。ここで、 $-\nabla f(x^k)$ とは最急降下法における降下方向であり、これを $d'^k := -\nabla f(x^k)$ とおくと、射影勾配法の降下方向は $d^k = P_X(x^k + s_k d'^k) - x^k$ と表すことができる。

さて、ここまでは最急降下法の降下方向を用いて射影勾配法の降下方向を決めていたが、今までの実験結果からわかるように最急降下法はニュートン法に比べ、収束が遅い。なので、最急降下法のかわりにニュートン法を用いて射影勾配法の降下方向を決めることにする。 $B_k := \nabla^2 f(x^k) + \tau I$ (I は単位行列) が正定値となるように $\tau \geq 0$ をおいて、ニュートン法と同様に降下方向 d'^k を $d'^k := -B_k^{-1} \nabla f(x^k)$ として、射影勾配法の降下方向を $d^k = P_X(x^k + s_k d'^k) - x^k$ とする。

この工夫が上手くいったかどうかを具体的な問題に対して実行することで考察する。射影勾配法を用いる課題は課題6と課題7があるが、ニュートン法を用いて降下方向を決める際にヘッセ行列が必要となる。課題6のヘッセ行列は課題7に比べ簡単に求めることができるので、課題6を用いて考察する。ここで、課題6のヘッセ行列は以下である。

$$\text{ヘッセ行列: } \nabla^2 f(x) = \begin{pmatrix} 2\left(-\frac{5.1x_0}{2\pi^2} + \frac{5}{\pi}\right)^2 - \frac{5.1}{\pi^2} \left(x_1 - \frac{5.1x_0^2}{4\pi^2} + \frac{5x_0}{\pi} - 6\right) - 10\left(1 - \frac{1}{8\pi}\right) \cos x_0 & 2\left(-\frac{5.1x_0}{2\pi^2} + \frac{5}{\pi}\right) \\ 2\left(-\frac{5.1x_0}{2\pi^2} + \frac{5}{\pi}\right) & 2 \end{pmatrix}$$

この工夫を追加して課題6について計算した実験結果が表6である。なお、比較のために最急降下法を用いたときの実験結果(表3の内容)も表に含めている。また、計算が終了した理由はいずれも「 $\|d^k\| \leq \varepsilon$ 」である。

目的解が最急降下法を使った場合とニュートン法を使った場合とで異なっているが、どちらも大域的最小解であるので、最小化問題は解けている。表6からわかるように、勾配方向の決定の際に最急降下法よりもニュートン法を使った方が反復回数、評価回数ともに少なくなった。これは最急降下法よりもニュートン法のほうがより適切な降下方向を選んでいるからである。射影勾配法においても同様のことがいえる。

手法	直線探索の手段	反復回数	目的関数の評価回数	最後の3回の $\ d^k\ $	最適値	目的解
射影勾配法 (ニュートン法)	$\rho = 0.5$	4	8	(3.349e-02, 9.876e-05, 3.804e-10)	0.3979	$(9.425, 2.475)^\top$
	二次補間法	4	8	(3.349e-02, 9.876e-05, 3.804e-10)	0.3979	$(9.425, 2.475)^\top$
射影勾配法 (最急降下法)	$\rho = 0.5$	45	203	(2.136e-06, 2.907e-06, 1.253e-06)	0.3979	$(3.142, 2.275)^\top$
	二次補間法	39	118	(4.003e-06, 2.006e-06, 1.926e-06)	0.3979	$(3.142, 2.275)^\top$

表 6 課題 6 の問題:工夫を加えたときの実験結果と元の降下法による実験結果

まとめと今後

最急降下法, ニュートン法, 射影勾配法, バックトラック法の原理と方法を理解することができた. 最適化においては, どの手法も一長一短であり, 問題によって適切な手法を選ぶことが大事であることが分かった.

今回の方法では局所的最小解に収束してしまうことがあるため, 必ず大域的最小解を求めることができるようにすることが今後の課題である.