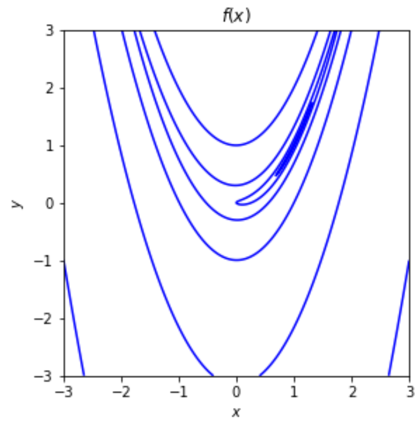
問 1 に取り組んだ。以下、

$$f(x_1, x_2) = 100(x_2 - {x_1}^2)^2 + (1 - x_1)^2$$

の大域的最適解は 2 つの 2 乗の項が 0 となる $(x_1, x_2) = (1,1)$ であり、$f(1,1) = 0$ であることに注意する。下図に f の等高線を示した。最適解周辺で細長い等高線ができているところが特徴的である。また、初期点として $x_0 = (1.2, 1.2)^{\mathrm{T}}$、$x_1 = (-1.2, 1)^{\mathrm{T}}$ の他に $x_2 = (1, -1.2)^{\mathrm{T}}$、$x_3 = (-1.2, -1.2)^{\mathrm{T}}$ の場合を試した（補足）。全ての手法について、停止条件を $\|\nabla f(x_k)\| \leq \varepsilon = 0.0001$ とし、レポート問題文に記載のあった標準的なパラメータを用いて実装した。



1. 最急降下法

Armijo 条件と Wolfe 条件を使った場合の 2 つを実装した。

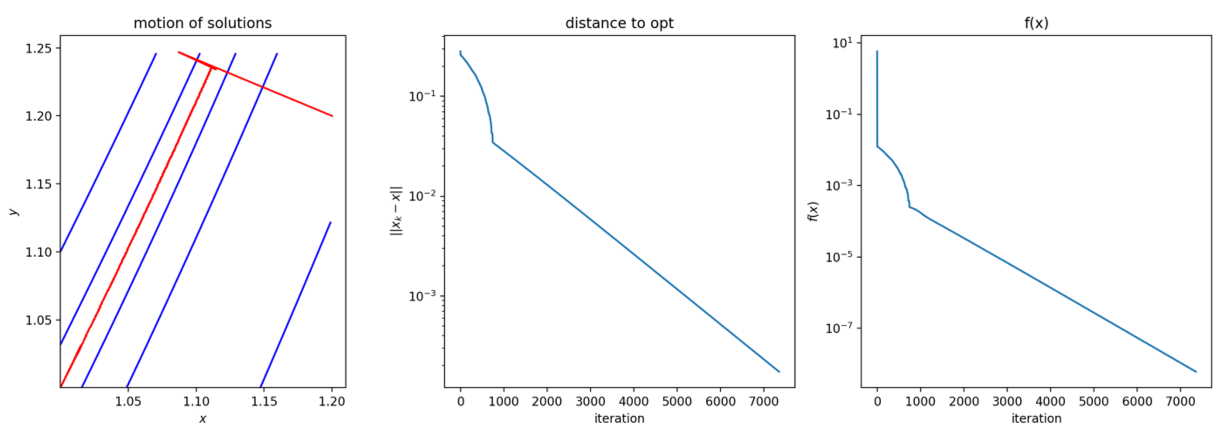
1.1. Armijo 条件（バックトラック法）

結果は次の通りになった。以下、一番左の図には青で等高線、赤で解の挙動を示す。また、関数 f は引数をベクトルと見て $f(x)$ と表す。

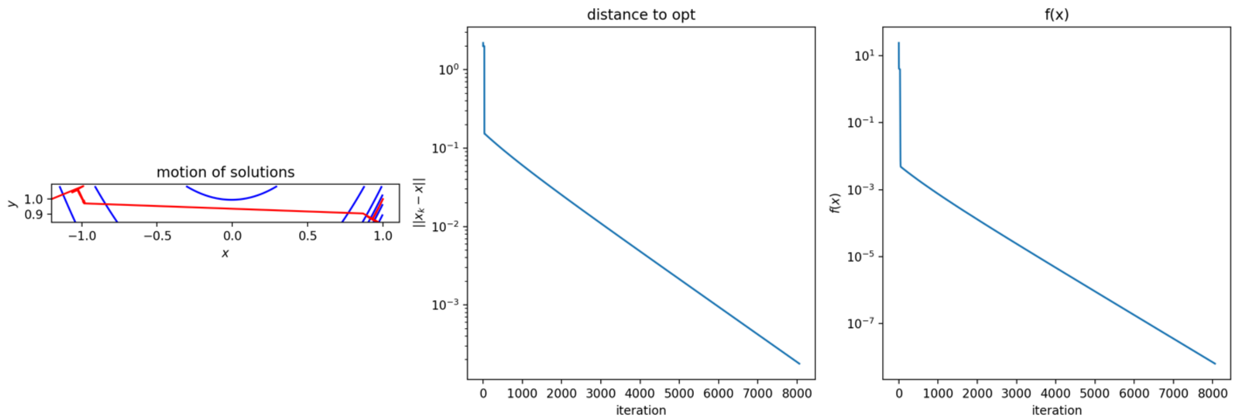
・初期点 $x_0 = (1.2, 1.2)^{\mathrm{T}}$ のとき

解：[1.00007748 1.00015543]

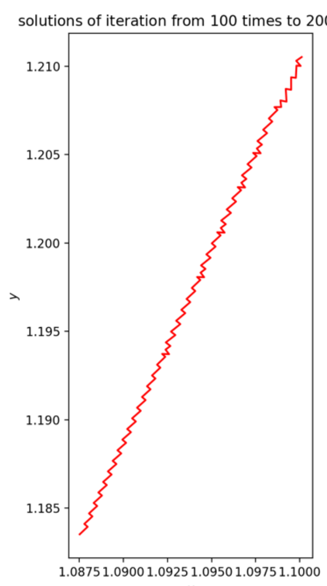
グラフ：（$f(1,1) = 0$ より一番右のグラフには $f(x)$ を示している。以下同様。）

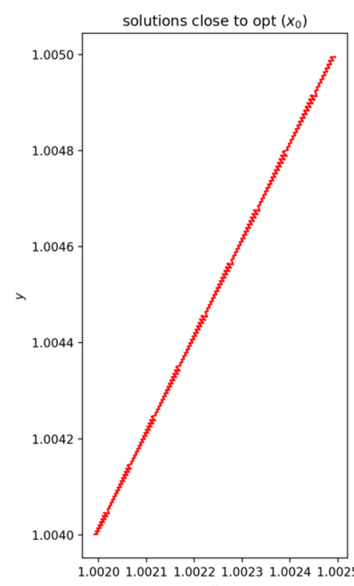


・初期点 $x_1 = (-1.2, 1)^{\mathrm{T}}$ のとき
解：[0.99992058 0.9998407]
グラフ：



・初期点 $x_0 = (1.2, 1.2)^{\mathrm{T}}$ のときの解の挙動を拡大したもの
　上図の解の挙動では最初の何回かの解の挙動が大きく、他の部分での解の動きが分かりにくいため、さらに反復の最初の部分の解の挙動と最適解近傍での解の挙動を調べると下のようになった。左図は反復が 100 回から 200 回までの解の挙動、右図は$|x-1|$、$|y-1|$ が(0.004,0.005) に含まれる範囲での解の挙動を表す。

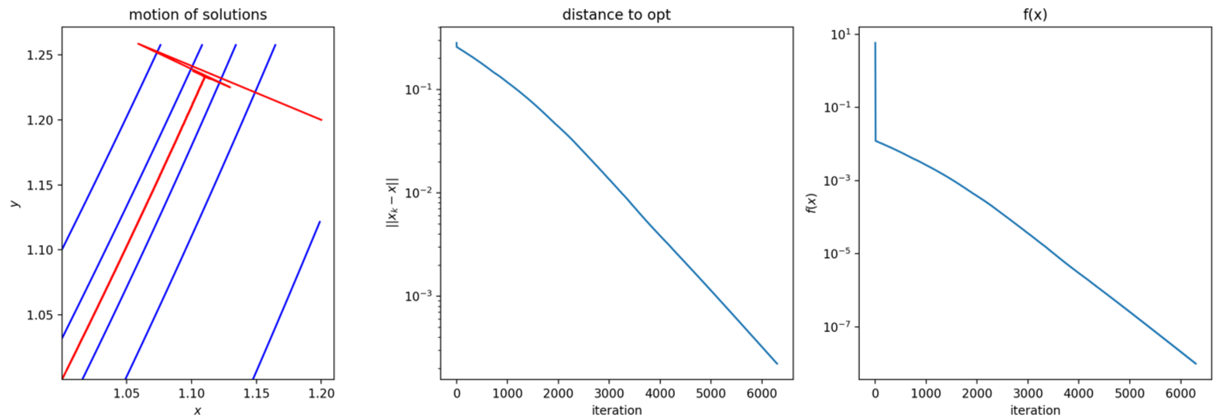




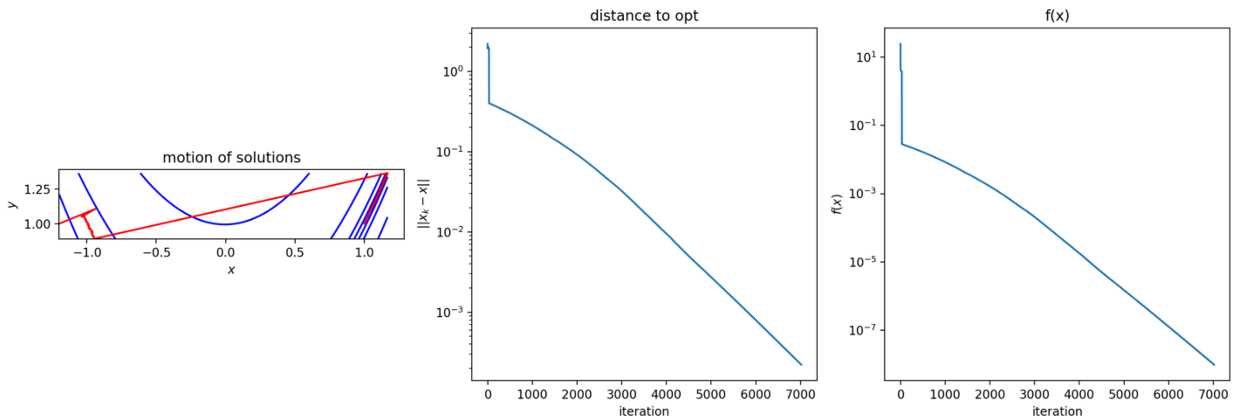
1.2. Wolfe 条件

　レジュメを参考にして Wolfe 条件を用いた場合の実装をした。最初の$\alpha_{\max}$ （Wolfe 条件のレジュメと同じもの）の取り方によって収束の速さが変わる様子が見られたが、今回は試した中で収束が速かった$\alpha_{\max} = 10$ の場合の結果を以下に示す。（$\alpha_{\max}$ の良い取り方は関数と初期点に依存すると考えられる。）

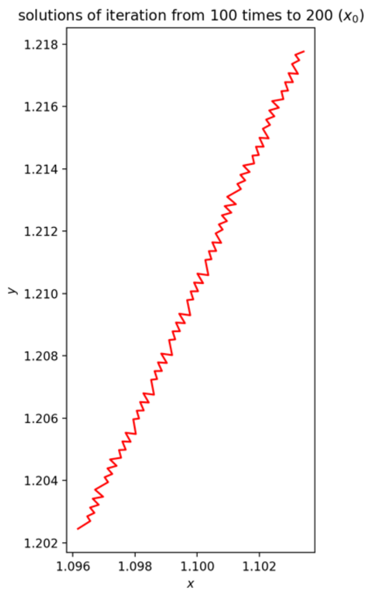
・初期点 $x_0 = (1.2, 1.2)^{\mathrm{T}}$ のとき
解：[1.00009968 1.00019986]
グラフ：

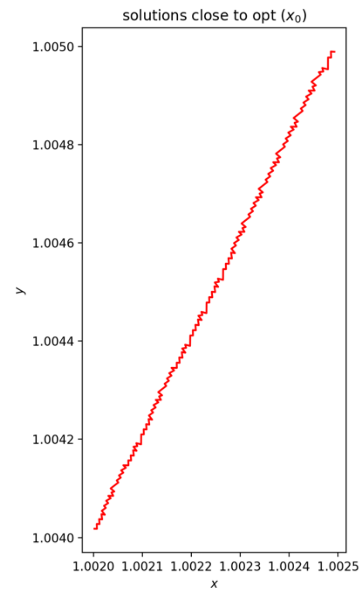


・初期点 $x_1 = (-1.2, 1)^{\mathrm{T}}$ のとき
解：[1.00009904 1.00019858]
グラフ：



・初期点 $x_0 = (1.2, 1.2)^{\mathrm{T}}$ のときの解の挙動を拡大したもの





1.3.考察（最急降下法）

補足に示した場合も含めて最適解から遠い初期点の方が反復の回数が多くなる傾向が見られた。これについて上の 2 つの初期点の場合を比較するとx_0の方は 10 回程度で最終的な解の改善方向に入っていたのに対し、x_1では 100 回程度かかっていたことが$f(x)$ のグラフからも分かり、最適解から遠ざかることで最適解でない場所で解の小さい改善を行ってしまうため、反復の回数が増え得ると考えた。しかし今回の関数で反復回数を大きく左右するのは、解の挙動を全体的に見た時の最終的な解の改善方向（$f(x)$ の等高線が細長くなる部分）に入る際にできるだけ最適解に近い位置で入れるどうかであると考えられるため、初期点の位置によっては少し遠ざけると反復が減る場合があると考えた。

また、$f(x)$、最適解への距離ともに基本的に単調減少だが、log に対してほぼ線形であり、反復が進むにつれ減少速度が遅くなっていることがグラフから読み取れる。このことは反復の最初の部分の解の挙動と最適解近傍での解の挙動を比較した図からも理解でき、反復が進むにつれてステップ幅が小さくなっていることが分かる。

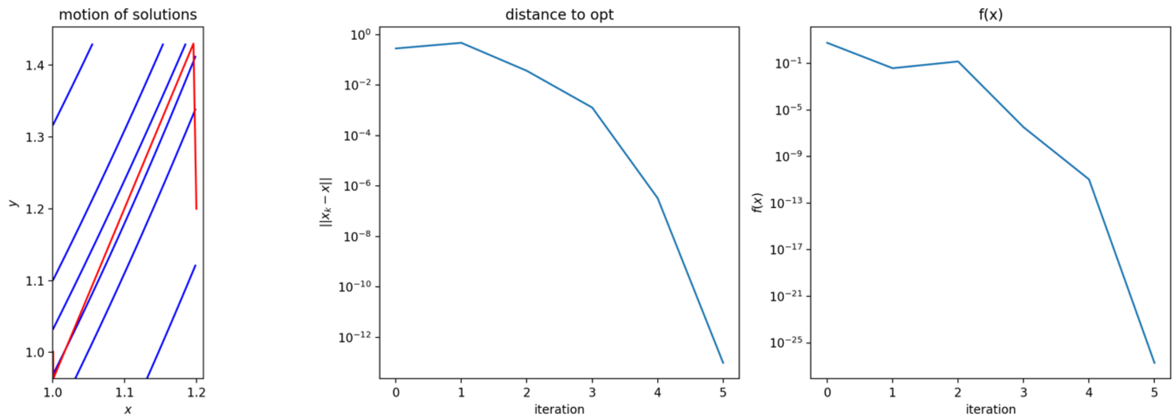
バックトラック法と Wolfe 条件を利用した場合の違いについては、グラフの傾きや解の挙動から見られるステップ幅の違いから見て取れる。Wolfe 条件を利用した場合ではステップ幅が大きくとられていたことから、曲率条件がきちんと動いていることが確認できた。また、補足に挙げた他の初期点の場合も含め、バックトラック法の途中で見られる上に凸のコブ状の解の改善が、Wolfe 条件を実装した場合はほとんど見られないが、これは曲率条件のおかげでステップ幅を大きく取れるために急に解の改善が悪くなる状況が起こりにくいからだと考えられる。

2. ニュートン法

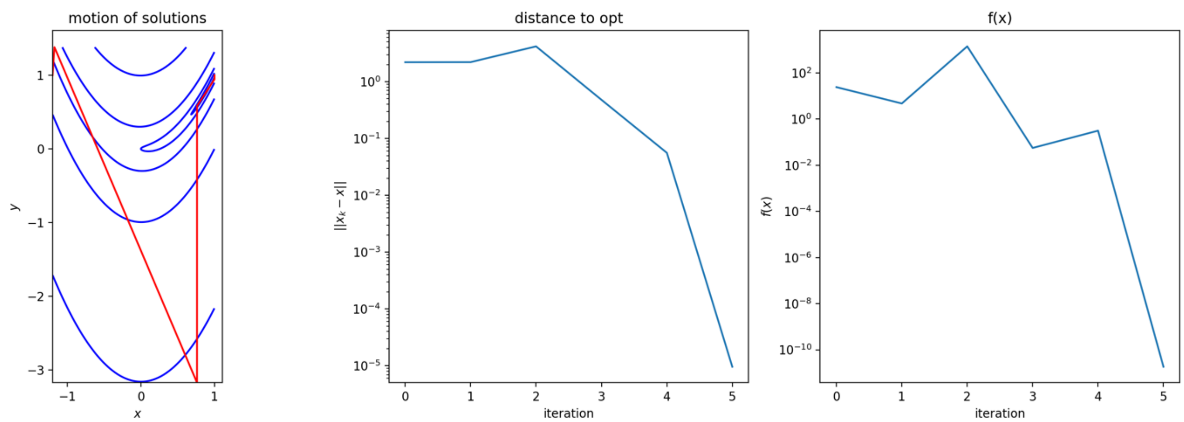
次にニュートン法による計算結果を以下に示す。

・初期点 $x_0 = (1.2, 1.2)^{\mathrm{T}}$ のとき

解：[1. 1.]
グラフ：



・初期点 $x_1 = (-1.2, 1)^{\mathrm{T}}$ のとき
解：[0.9999957 0.99999139]
グラフ：



2.1 考察（ニュートン法）

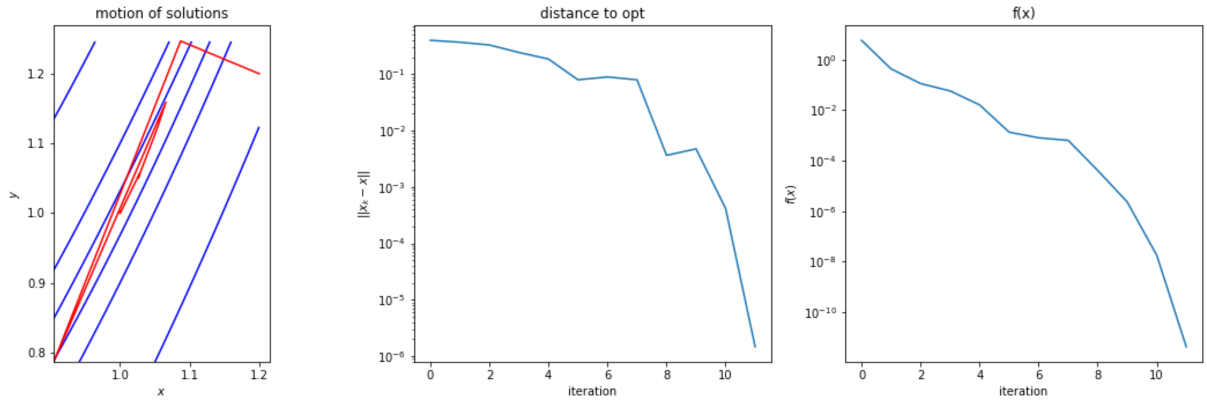
2 次近似を利用していることで収束が桁違いに速くなっている。$f(x)$ 最適解への距離が最初の 1,2 回では減少しにくい様子が見られるが、いずれについてもその後すぐに値を下げ、特に最後の 1,2 回で急激にオーダーを下げている。ここに 2 次収束の速さが現れていると考えられる。最初の方で解が改善しないことがあるのは、改善方向には向かっているものの、ステップ幅を 1 で固定しているために行き過ぎてしまうからだと考えられる。

3. 準ニュートン法

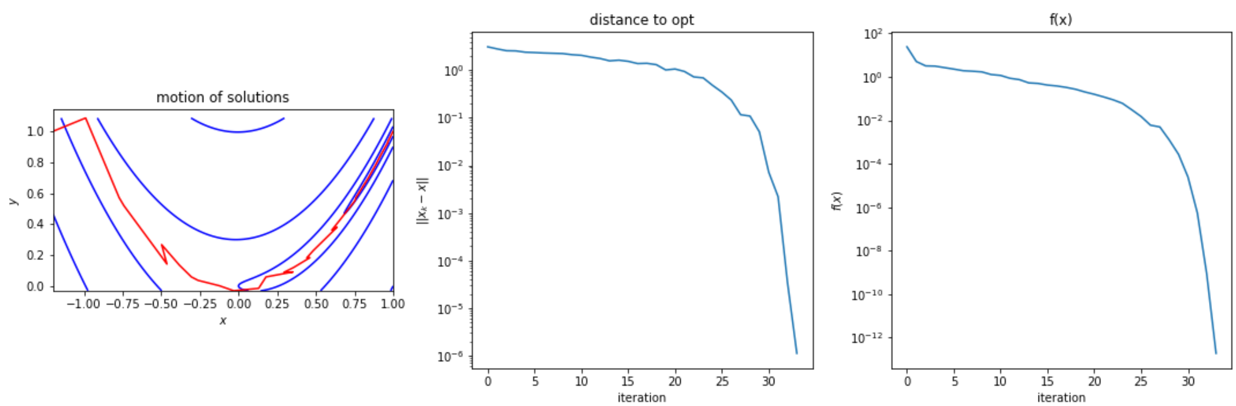
最後に、準ニュートン法による計算結果を以下に示す。（逆）行列の更新式は BFGS 公式、ステップ幅の計算は Armijo 条件を課し、バックトラック法により行った。

・初期点 $x_0 = (1.2, 1.2)^{\mathrm{T}}$ のとき

解：[[1.00000054] [1.00000088]]
グラフ：



・初期点 $x_1 = (-1.2, 1)^{\mathrm{T}}$ のとき
解：[[0.99999966] [0.99999928]]
グラフ：



3.1 考察（準ニュートン法）

　最急降下法より格段に収束が速く、ニュートン法の数倍の反復回数で収束している。ニュートン法の時と同様に、最適解近傍に至ると最適解との距離や $f(x)$ が急激にオーダーを下げる様子が見られ、ここに準ニュートン法の超一次収束性が現れていると考えられる。収束の早さの他にも、ここではステップ幅の改善に Armijo 条件を課しているために解が必ず改善されているところがニュートン法との違いとして挙げられる。

4. 補足（異なる初期点の場合）

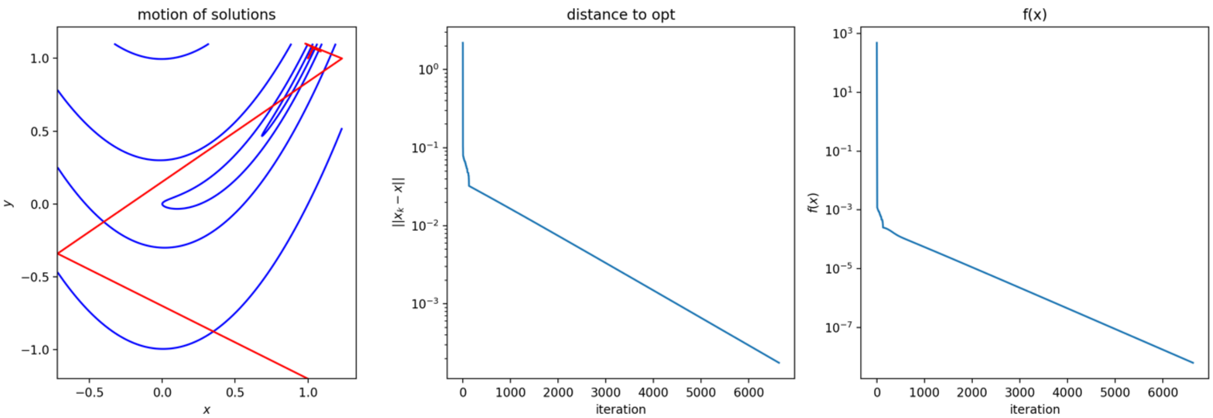
　以下、比較のために他の初期点での実行結果を示す。初期点x_0 において最急降下法を実行した場合、バックトラック法の方が Wolfe 条件を実装したときよりも反復回数が少なく済んだこと以外は初期点x_0、x_1 の場合と同様の結果が得られた。これは、x_0 の時には最終的な解の改善方向（$f(x)$ の等高線が細長くなる部分）に最適解から遠い位置で入ってしまったためだと考えられる。

4.1.1.最急降下法（Armijo 条件）

・初期点 $x_2 = (1, -1.2)^{\mathrm{T}}$ のとき

解：[1.00007856 1.0001576]

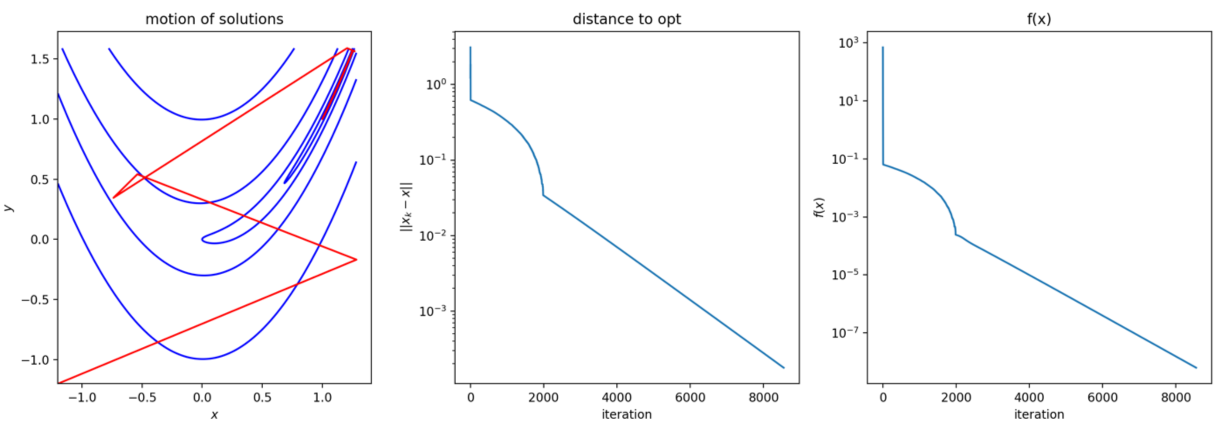
グラフ：



・初期点 $x_3 = (-1.2, -1.2)^{\mathrm{T}}$のとき

解：[1.00007916 1.00015879]

グラフ：

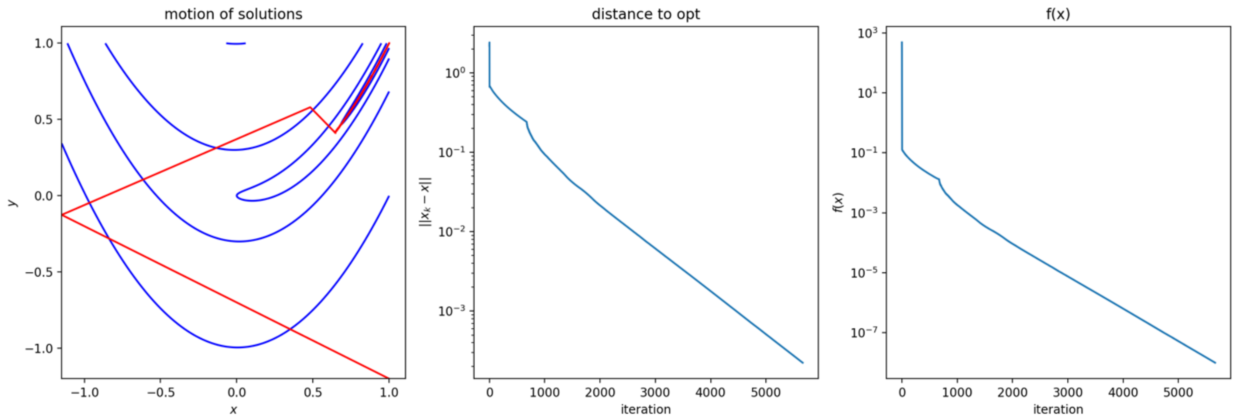


4.1.2 最急降下法（Wolfe 条件）

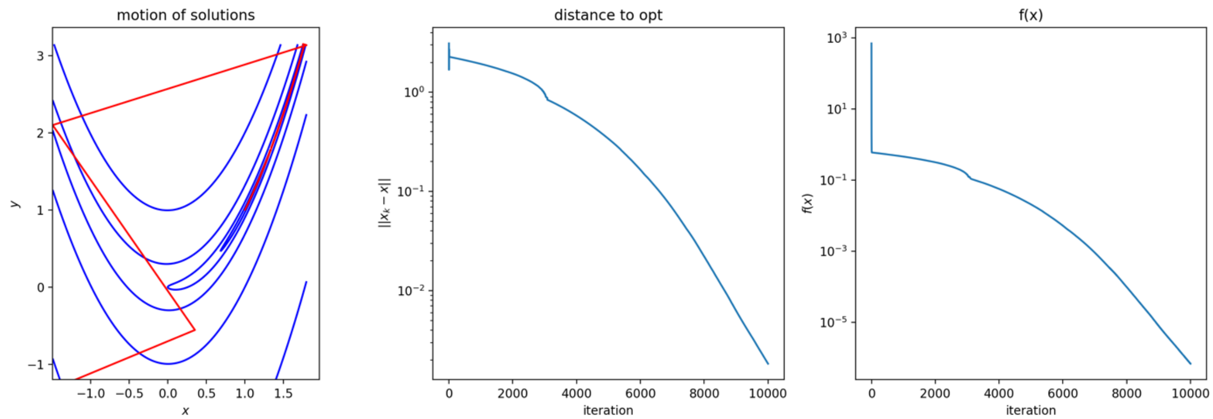
・初期点 $x_2 = (1, -1.2)^{\mathrm{T}}$ のとき

解：[0.99990012 0.99979975]

グラフ：



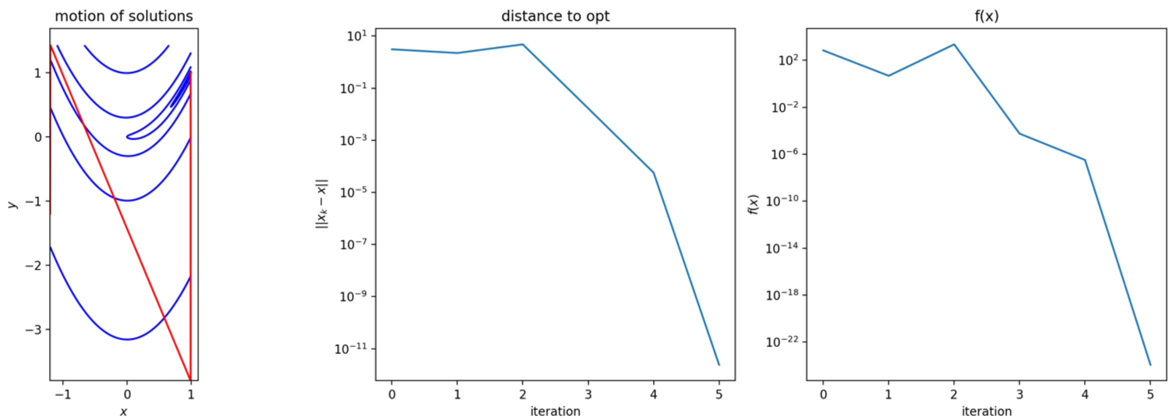
初期点 $x_3 = (-1.2, -1.2)^{\mathrm{T}}$のとき
解：[1.00081738 1.00164364]
グラフ：



4.2. ニュートン法
・初期点 $x_2 = (1, -1.2)^{\mathrm{T}}$ のとき
解：[1. 1.]
　この時は、$(\nabla^2 f(x_2))^{-1} = \frac{1}{176400}\begin{pmatrix} 200 & 400 \\ 400 & 1682 \end{pmatrix}$、$\nabla f(x_2) = \begin{pmatrix} 880 \\ -440 \end{pmatrix}$ となり、$-(\nabla^2 f(x_2))^{-1}\,\nabla f(x_2) = \begin{pmatrix} 0 \\ 2.2 \end{pmatrix}$ より 1 回で最適解が求まってしまうため、グラフは省略する。

初期点 $x_3 = (-1.2, -1.2)^{\mathrm{T}}$のとき
解：[1. 1.]

グラフ：

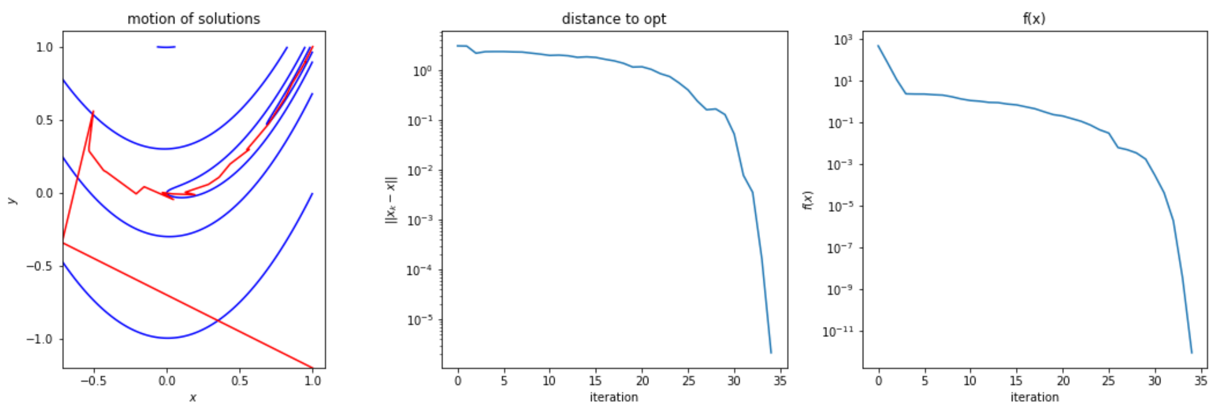


4.3. 準ニュートン法

・初期点 $x_2 = (1, -1.2)^{\mathrm{T}}$ のとき

解：[[0.99999935] [0.99999863]]

グラフ：



初期点 $x_3 = (-1.2, -1.2)^{\mathrm{T}}$のとき

解：[[0.99999995] [0.9999999]]

グラフ：