情報工学実験２

「数理計画法」

第２回：非線形最適化のアルゴリズム

数理計画法（数理最適化）

数理最適化における作業：

- モデリング
- 定式化
- アルゴリズムを用いた求解
- 得られた解の吟味，モデルの再検討

数学＋計算機による作業

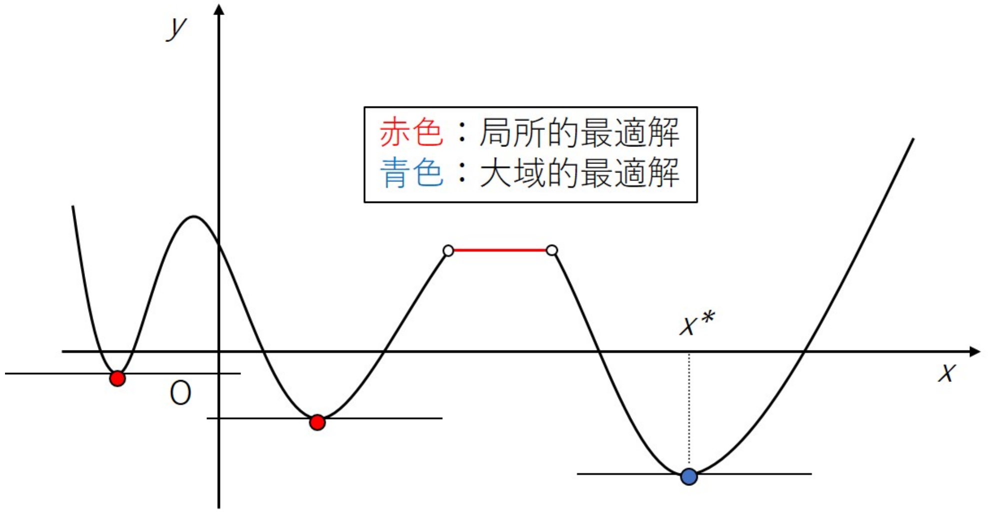
多くの場合ソルバーを活用
ソルバーは万能でない
ソルバーが求める解は最適解の保証なし

第２回の目的：
非線形計画法に対する最急降下法とニュートン法の実装
アルゴリズムの特性を把握することの大切さを学ぶ

無制約最小化問題と最適解

$f: \mathbb{R}^n \to \mathbb{R}$, n変数関数,　目的関数

無制約最小化問題　$$\min_{(x_1,\dots,x_n)\in\mathbb{R}^n} f(x_1,\dots,x_n) = \min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$$



大域的最適解　$\boldsymbol{x}^* = (x_1^*,\dots,x_n^*)$:

$f(\boldsymbol{x}^*) \le f(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \mathbb{R}^n$

局所的最適解　$\boldsymbol{x}^* = (x_1^*,\dots,x_n^*)$: $\quad \exists \epsilon > 0$

$f(\boldsymbol{x}^*) \le f(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \mathbb{R}^n, \|\boldsymbol{x}^* - \boldsymbol{x}\| < \epsilon$

非線形計画問題は局所最適解が求まれば「御の字」

勾配ベクトル，ヘッセ行列

$f: \mathbb{R}^n \to \mathbb{R}$, n変数関数

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{pmatrix}, \quad \nabla^2 f(\boldsymbol{x}) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right) = \begin{pmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}$$

勾配ベクトル

ヘッセ行列

勾配ベクトルは目的関数が最も増加する方向

$$f(x_1, x_2) = 3x_1^2 + 4x_1x_2 + 5x_2^2 + \sin x_1 + e^{3x_2}$$

$$\nabla f(x) = \begin{pmatrix} 6x_1 + 4x_2 + \cos x_1 \\ 4x_1 + 10x_2 + 3e^{3x_2} \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} 6 - \sin x_1 & 4 \\ 4 & 10 + 9e^{3x_2} \end{pmatrix}$$

最適性条件

(1) $\boldsymbol{x}^* \in \mathbb{R}^n$が局所最適解

- $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ （１次の必要条件）　停留点
- $\nabla^2 f(\boldsymbol{x}^*)$:半正定値 （２次の必要条件） ※ f : 2 回連続的微分可能

(2) $\boldsymbol{x}^* \in \mathbb{R}^n$が以下の条件を満たすならば局所最適解

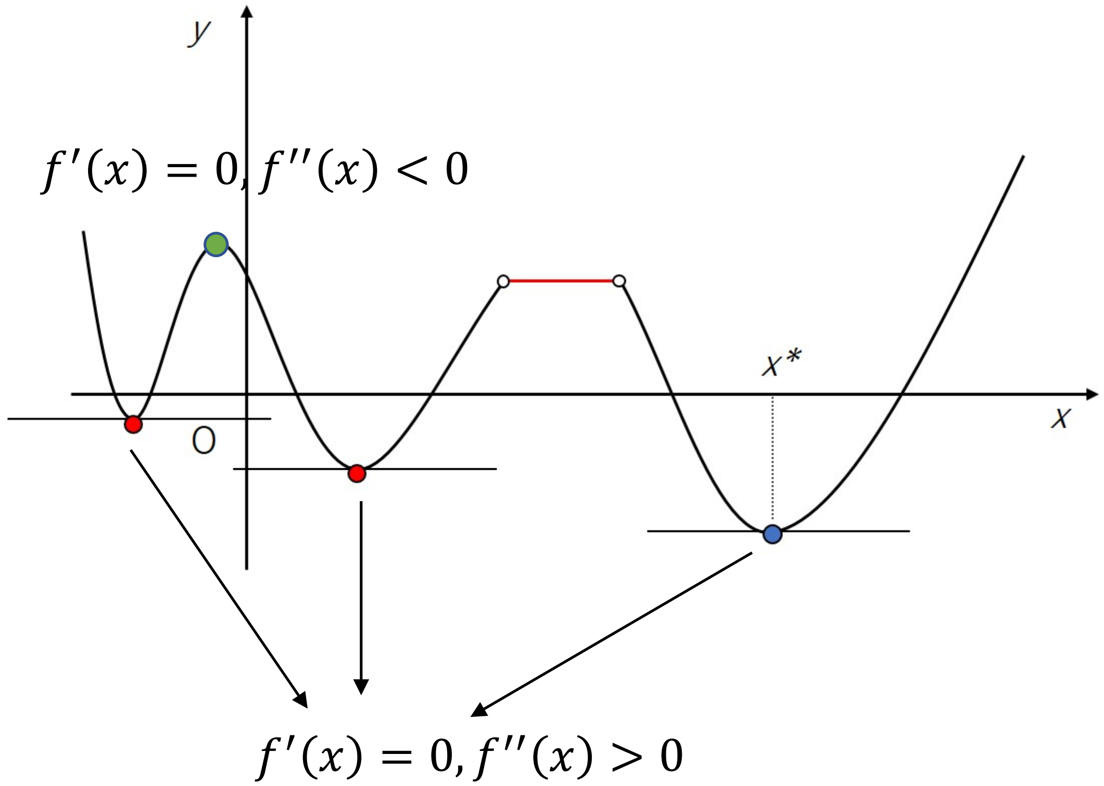
- $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$
- $\nabla^2 f(\boldsymbol{x}^*)$:正定値 （２次の十分条件） ※ f : 2 回連続的微分可能

対称 $n \times n$ 行列 X

半正定値： $d^T X d \geq 0 \ (\forall d \in \mathbb{R}^n) \Leftrightarrow X$ の固有値がすべて非負

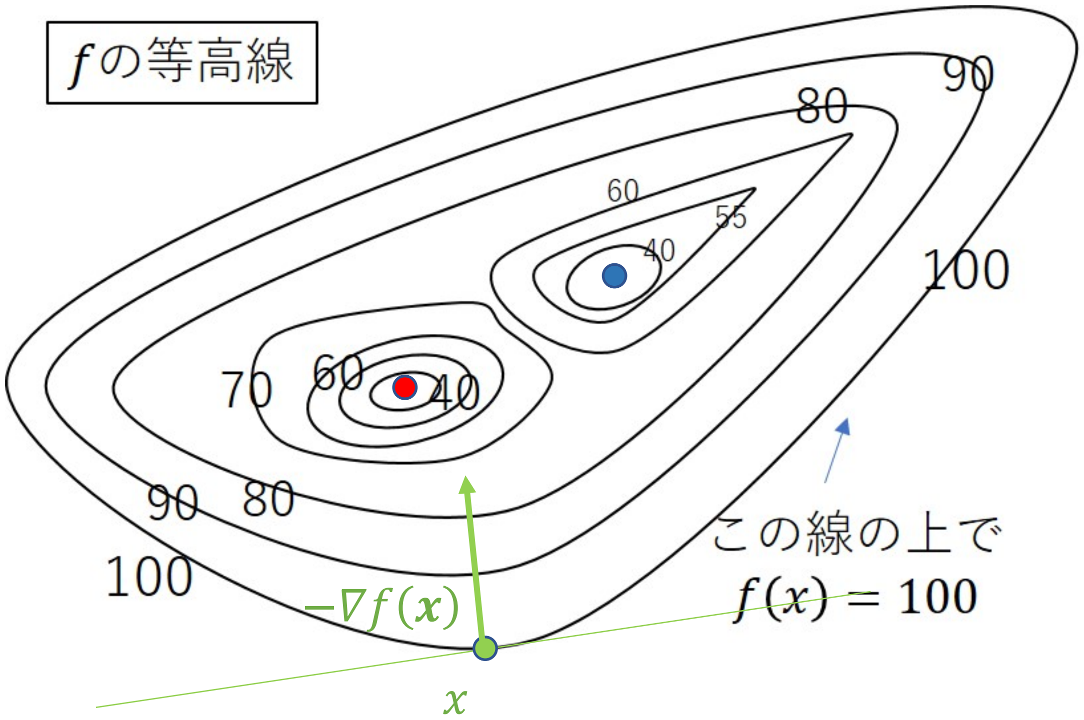
正定値：$d^T X d > 0 \ (\forall d \in \mathbb{R}^n, d \neq 0) \Leftrightarrow X$ の固有値がすべて正

1変数関数



y
$f'(x) = 0, f''(x) < 0$
O
x^*
x
$f'(x) = 0, f''(x) > 0$

2変数関数



fの等高線
90
80
60
55
40
100
70
60
40
90
80
100
$-\nabla f(x)$
x
この線の上で
$f(x) = 100$

凸関数とその性質

$f(x)$が凸関数

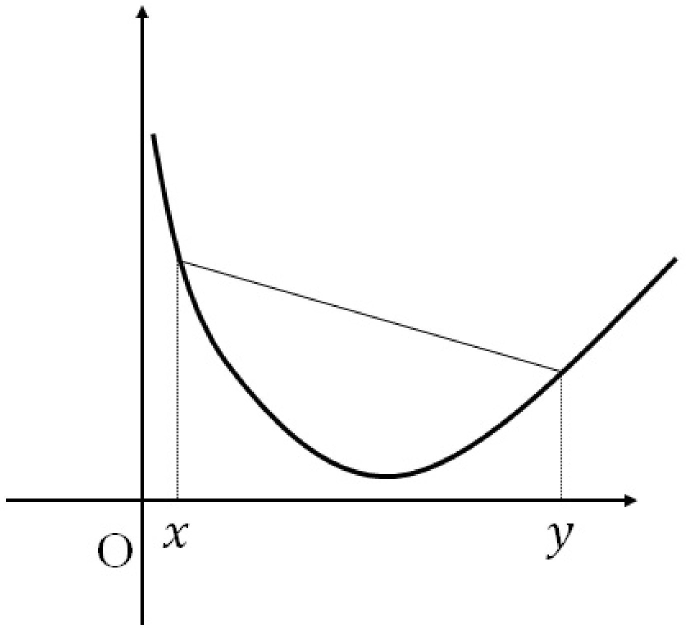
$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \le \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, \forall \lambda \in (0,1)$$

$f(x)$が凸関数$\Leftrightarrow$ $\nabla^2 f(x)$ が半正定値 $(\forall x \in \mathbb{R}^n)$

※ f : 2 回連続的微分可能

$f(x)$が凸関数のとき

- $\boldsymbol{x}^*$が局所最適解 $\Rightarrow$ $\boldsymbol{x}^*$が大域的最適解
- $\boldsymbol{x}^*$が停留点 $\Rightarrow$ $\boldsymbol{x}^*$が大域的最適解



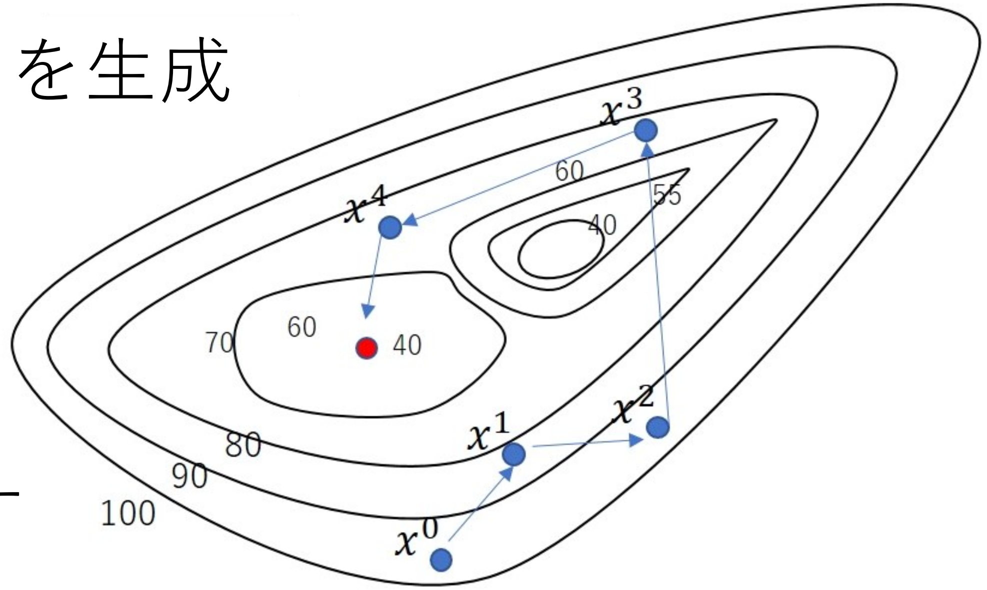
$f(x)$が凸関数のとき，ソルバーが求める解は大域的最適解

反復法

- 多くの非線形最適化アルゴリズムが用いる枠組み
- 初期点 $\boldsymbol{x}_0$ からスタート，点列 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ を生成

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

ステップ幅　探索方向



- 最終的に局所最適解 $\boldsymbol{x}^*$ への収束を目指す
- 探索方向は降下方向が望ましい
- ステップ幅は直線探索（または信頼領域法）で求める

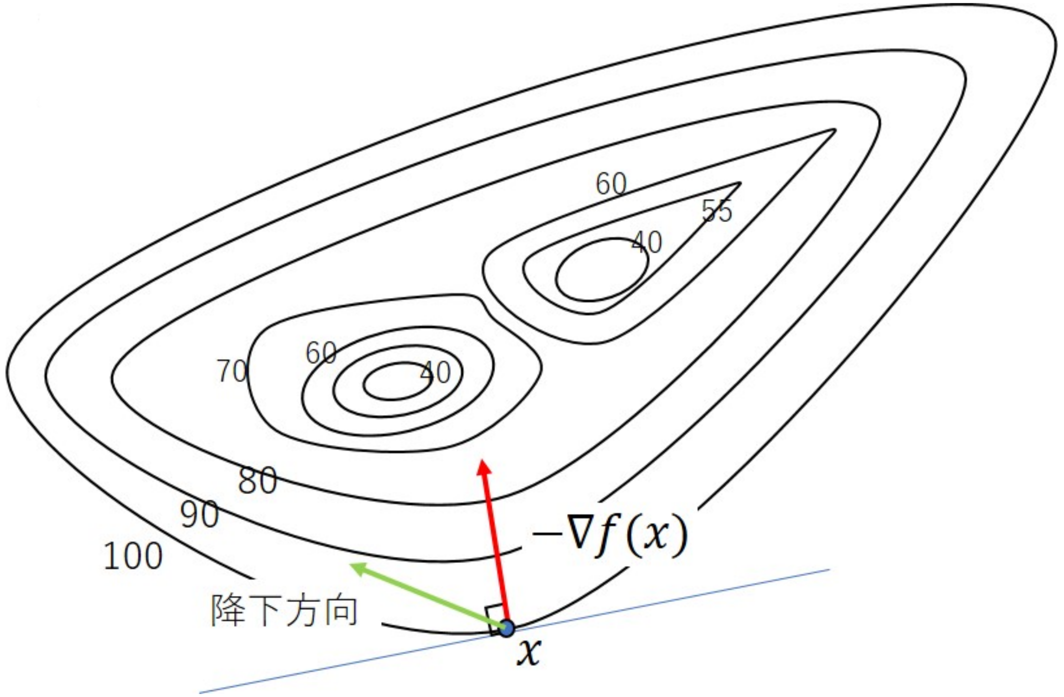
降下方向

- $\boldsymbol{x}$ における降下方向：$\nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d} < 0$ を満たす $\boldsymbol{d} \in \mathbb{R}^n$

降下方向：目的関数が減少する方向
勾配ベクトルの反対方向となす角が 90度未満

※勾配ベクトル：目的関数が最も急激に増加する方向

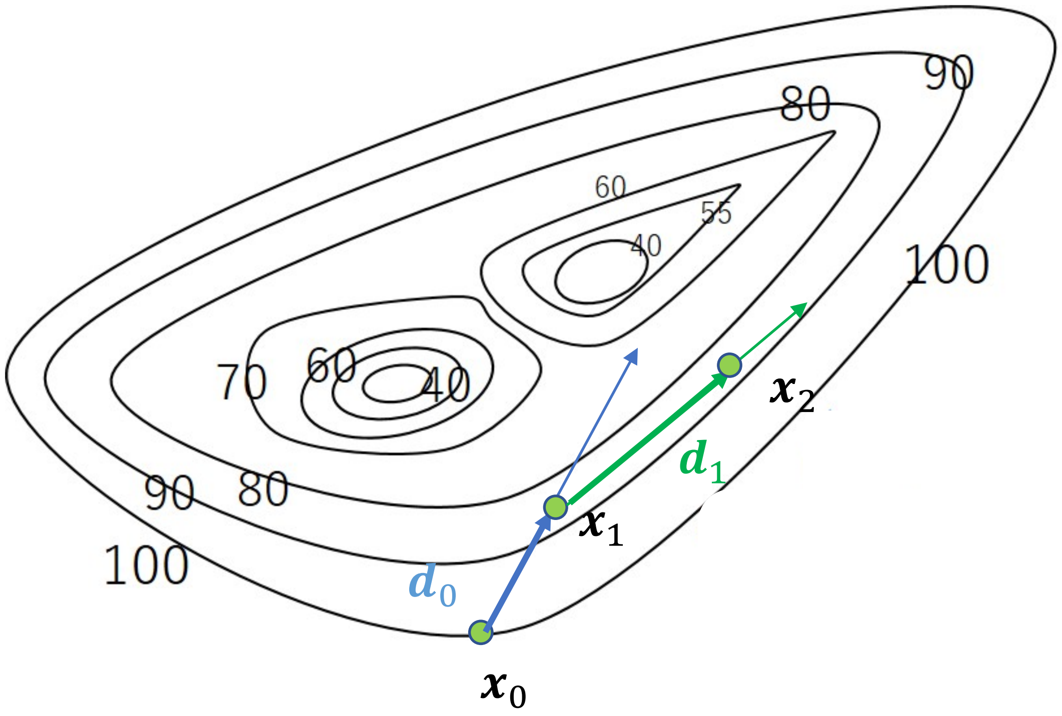
- $\boldsymbol{x}$ における最急降下方向：$-\nabla f(x)$



直線探索を用いた反復法

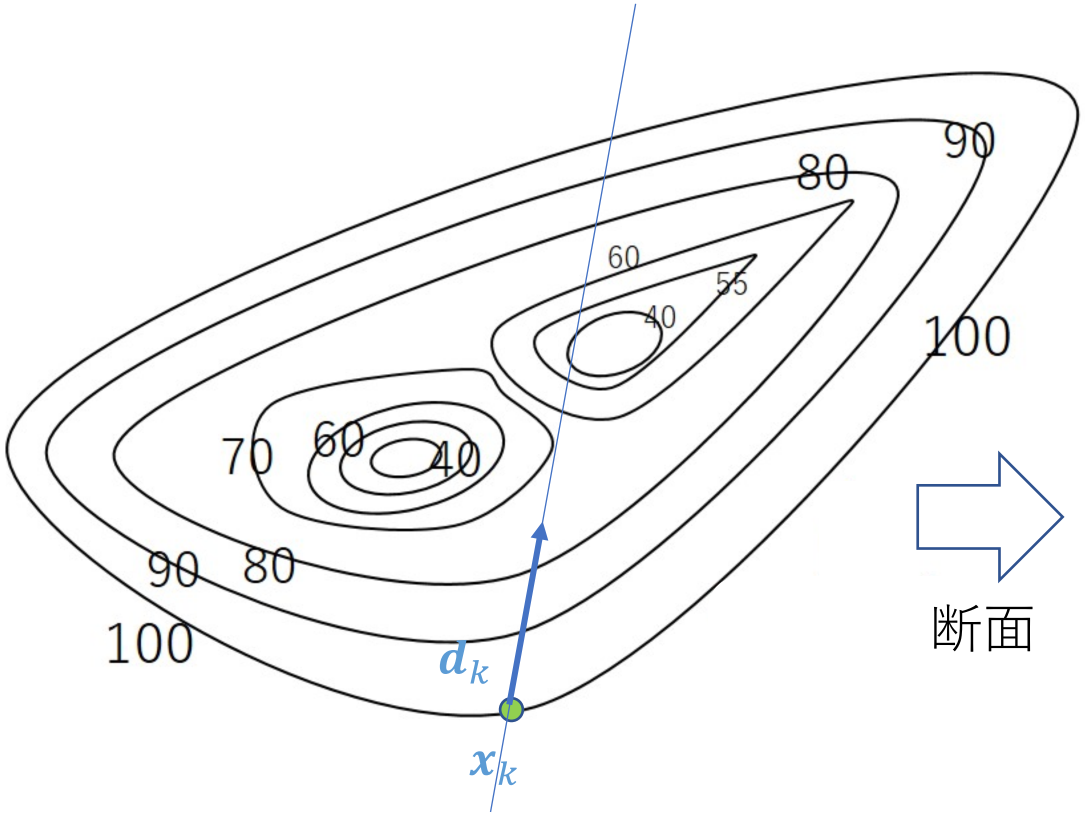
0: 初期点 $\boldsymbol{x}_0$ を選び， $k \coloneqq 0$ とする
1: 停止条件が満たされれば終了
2: 探索方向 $\boldsymbol{d}_k$ を定める
3: 直線探索でステップ幅α_kを計算
4: $\boldsymbol{x}_{k+1} \coloneqq \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
5: $k \coloneqq k+1$ として **1**へ

停止条件の例：
- $\|\nabla f(x_k)\|$が十分小さい
- 更新幅$\|x_{k+1} - x_k\|$が十分小さい
- 十分小さい：10^{-8} あるいは10^{-15}未満

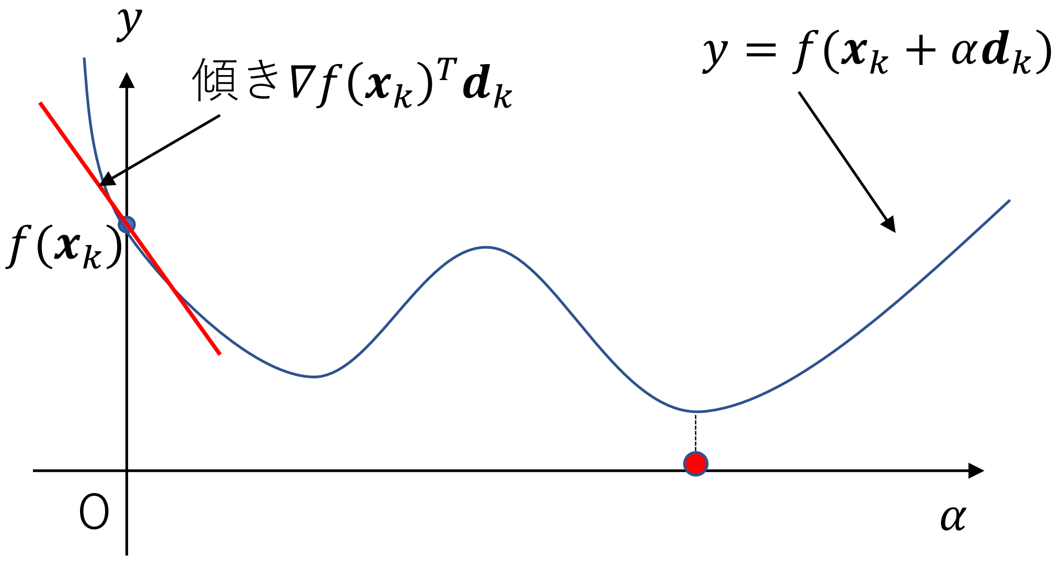


直線探索

- 現在点 $\boldsymbol{x}_k$ と探索方向 $\boldsymbol{d}_k$ が所与，ステップ幅を選ぶ
- 理想：α_k は $f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) = \min\{f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) | \alpha \geq 0\}$ を満たす



断面



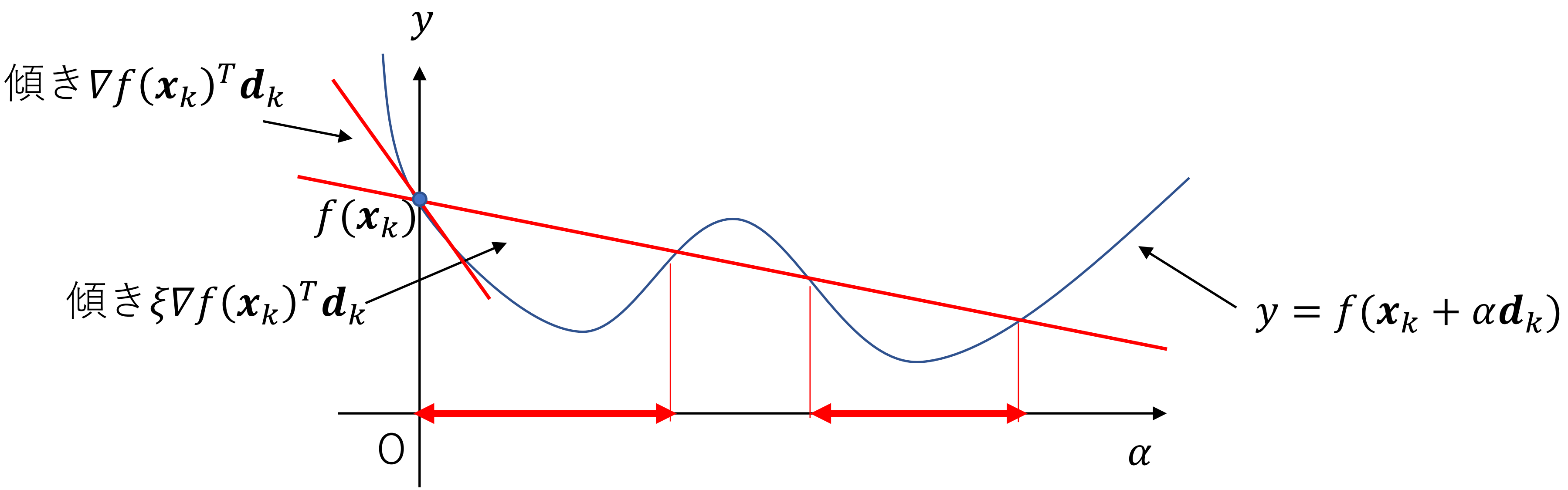
Armijo(アルミホ)条件

Armijo 条件：定数 ξ： $0 < \xi < 1$ に対して

$$f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \leq f(\boldsymbol{x}_k) + \left(\xi \nabla f(\boldsymbol{x}_k)^{\mathrm{T}} \boldsymbol{d}_k\right)\alpha$$ を満たす $\alpha > 0$

関数値のある程度の減少を保証する目的

Armijo条件を強化したWolfe条件もある



Armijo条件に対する直線探索法

現在点 $\boldsymbol{x}_k$ と探索方向 $\boldsymbol{d}_k$ は所与

0: パラメータ $0 < \xi < 1, 0 < \tau < 1$を選び，$\alpha \coloneqq 1$とする

1: Armijo 条件

$$f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \leq f(\boldsymbol{x}_k) + (\xi \nabla f(\boldsymbol{x}_k)^{\mathrm{T}} \boldsymbol{d}_k)\alpha \quad (+10^{-6})$$

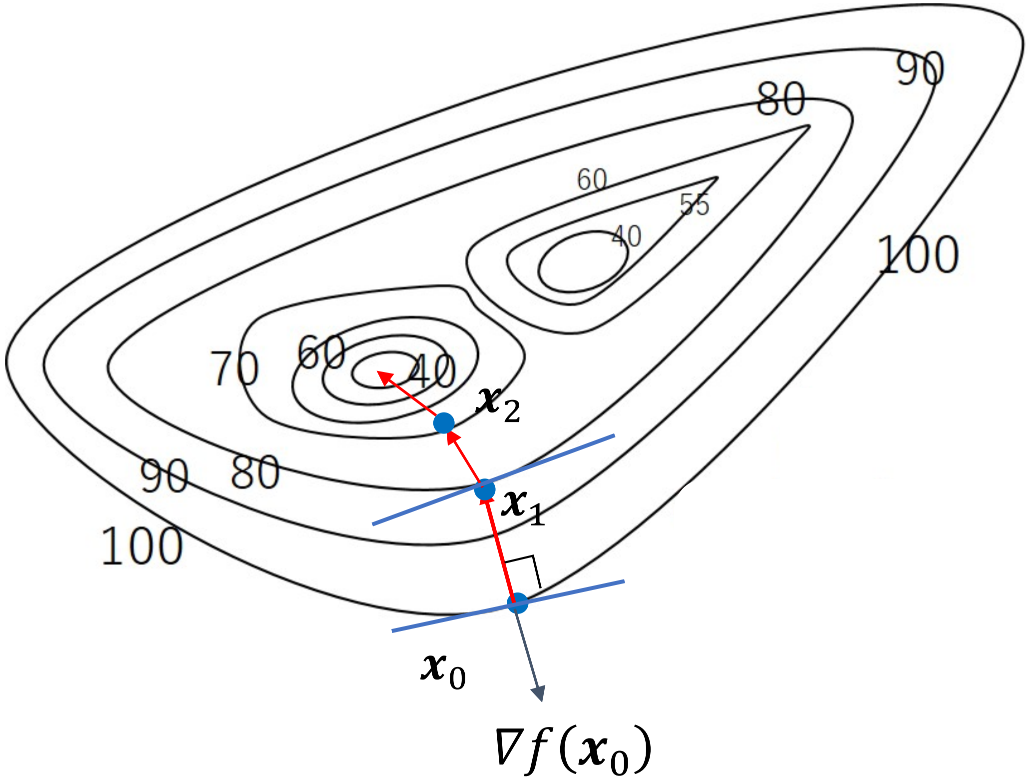
が満たされれば $\alpha_k \coloneqq \alpha$ として終了

2: $\alpha \coloneqq \tau\alpha$ として**1**へ

最急降下法のアルゴリズム

探索方向 $\boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$ とする反復法

0: 初期点 $\boldsymbol{x}_0$ を選び，$k := 0$ とする
1: 停止条件が満たされれば終了
2: 探索方向 $\boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$ を計算
3: 直線探索(Armijo条件 or Wolfe条件)
でステップ幅 α_k を計算
4: $\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
5: $k := k + 1$ として **1**へ



最急降下法の性質

関数 f がいくつかの仮定を満たすとき

- 大域的収束性をもつ

  直線探索にWolfe条件を用いると，任意の初期点に対して

$$\lim_{k\to\infty}\|\nabla f(\boldsymbol{x}_k)\| = 0$$

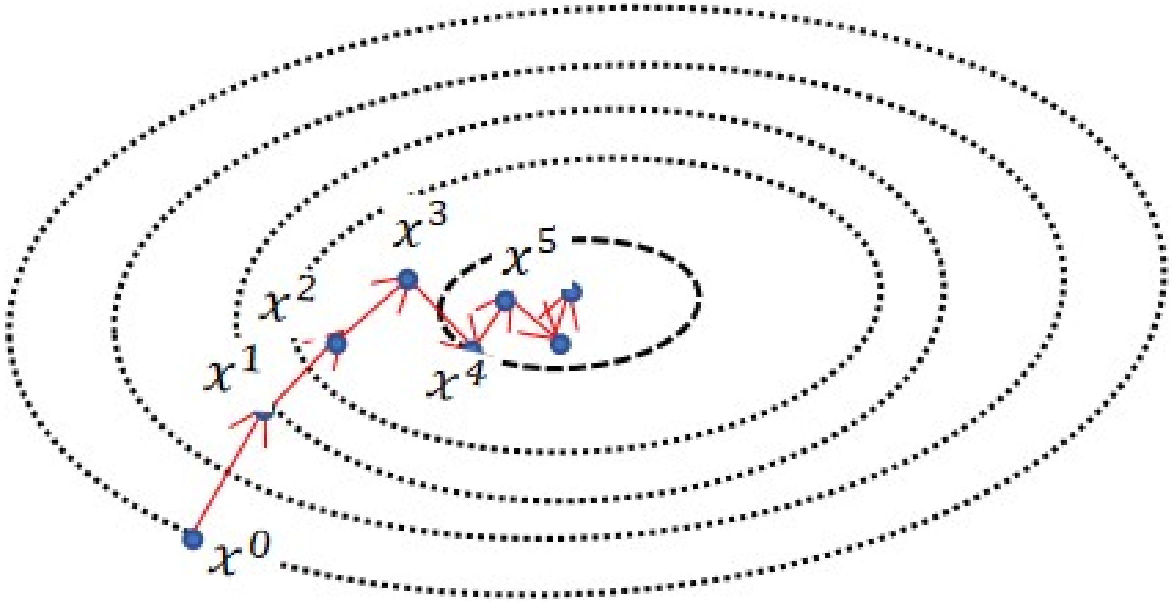
（生成点列は停留点に収束する）

- 生成点列は 1 次収束する

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\| \leq c\,\|\boldsymbol{x}_k - \boldsymbol{x}^*\|$$

を満たす定数$0 < c < 1$がある

$\boldsymbol{x}^*$ 付近では収束するまでに要する反復回数が非常に多い



$\boldsymbol{x}^*$ 付近では点列がジグザグに生成される

ニュートン法

- 現在点 $\boldsymbol{x}_k$ における$f(x)$の２次近似を利用した反復法

- 点 $\boldsymbol{x}_k$ における２次までの Taylor展開：

$$f(\boldsymbol{x}_k + \boldsymbol{d}) = f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d} = q(\boldsymbol{d})$$

において探索方向 $\boldsymbol{d}$ を次の関係を満たすように選ぶ

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

$$\boxed{\nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d} = -\nabla f(\boldsymbol{x}_k)}$$ ニュートン方程式

$\boldsymbol{d}$ ：ニュートン方向

ニュートン法のアルゴリズム

0: 初期点 $\boldsymbol{x}_0$ を選び， $k \coloneqq 0$ とする

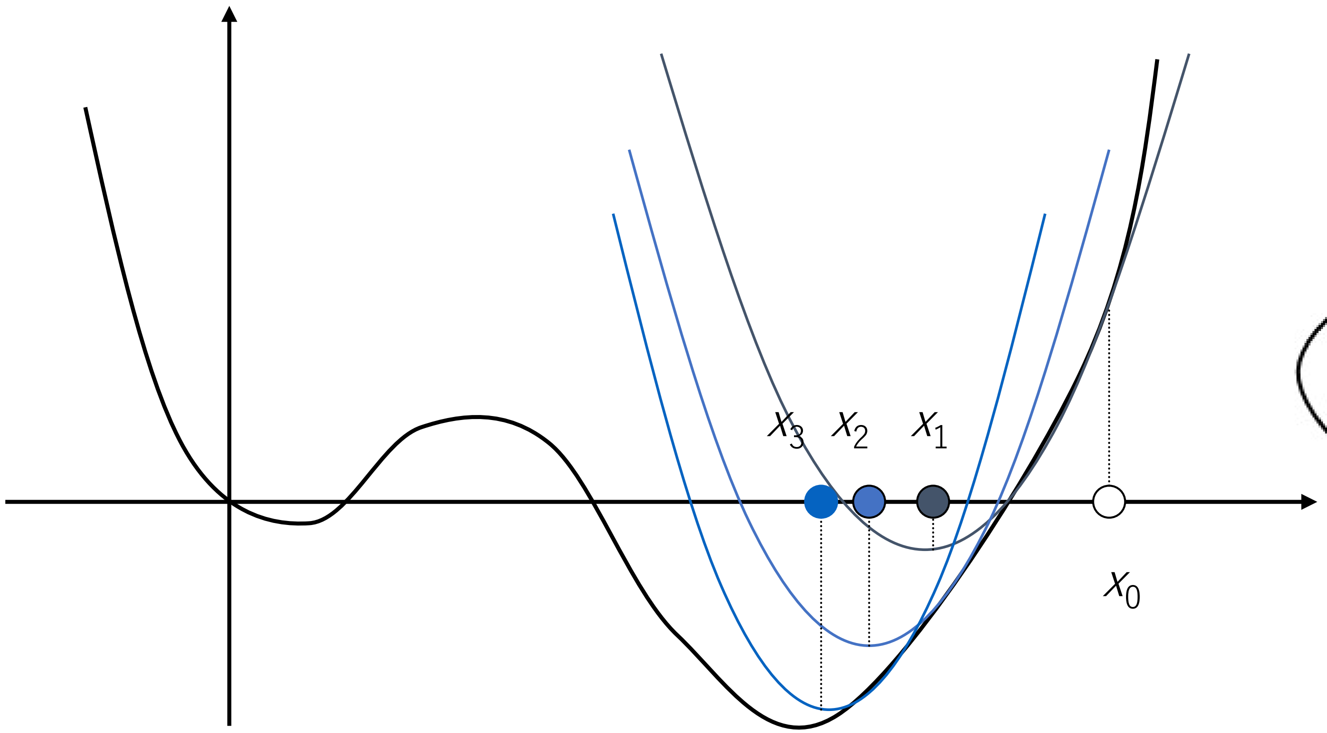
1: 停止条件が満たされれば終了

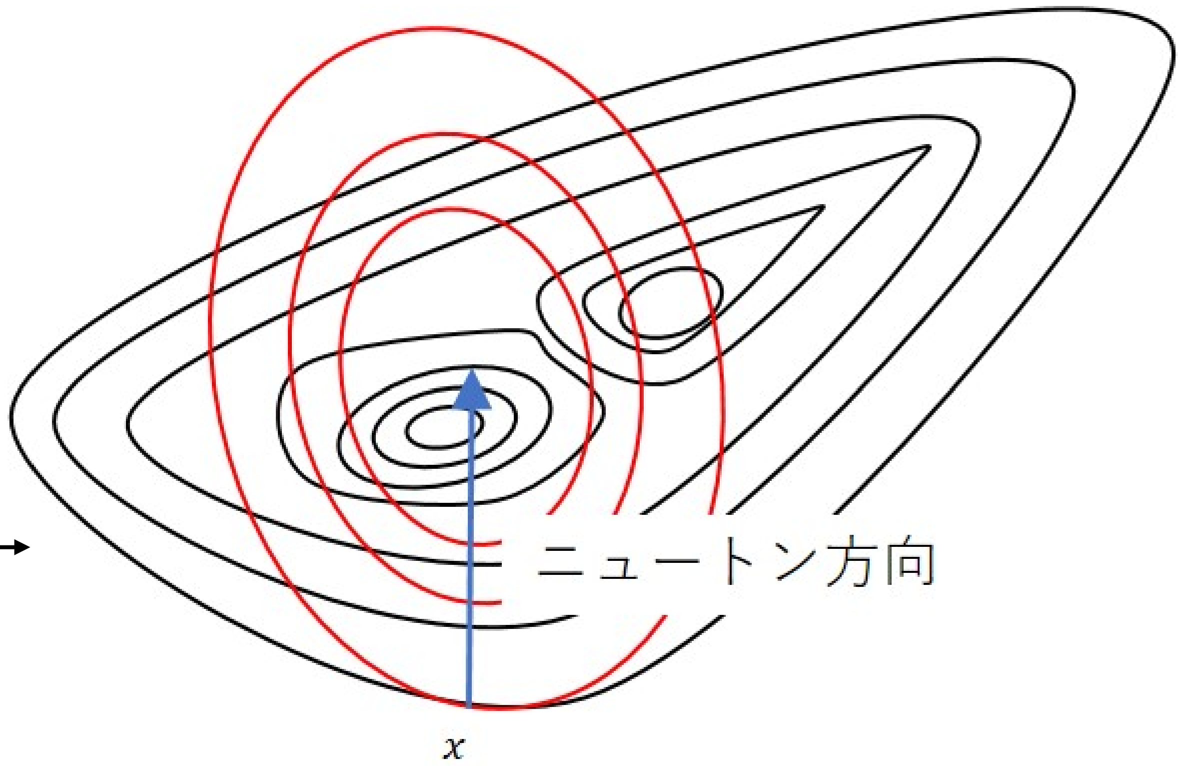
2: ニュートン方程式 $\nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d} = -\nabla f(\boldsymbol{x}_k)$ を解いて $\boldsymbol{d}_k$ を計算

3: $\boldsymbol{x}_{k+1} \coloneqq \boldsymbol{x}_k + \boldsymbol{d}_k$

4: $k \coloneqq k + 1$ として **1**へ

ステップ幅はつねに $\alpha_k = 1$ にとるので，直線探索は行わない


x_3 x_2 x_1
x_0


ニュートン方向
x

ニュートン法の性質

- 局所2次収束する：初期点$\boldsymbol{x}_0$を$\boldsymbol{x}^*$ の十分近くに取れば

  $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\| \leq c\,\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2$　　を満たす定数$c > 0$がある

  $\boldsymbol{x}^*$ 付近では収束スピードが非常に速い

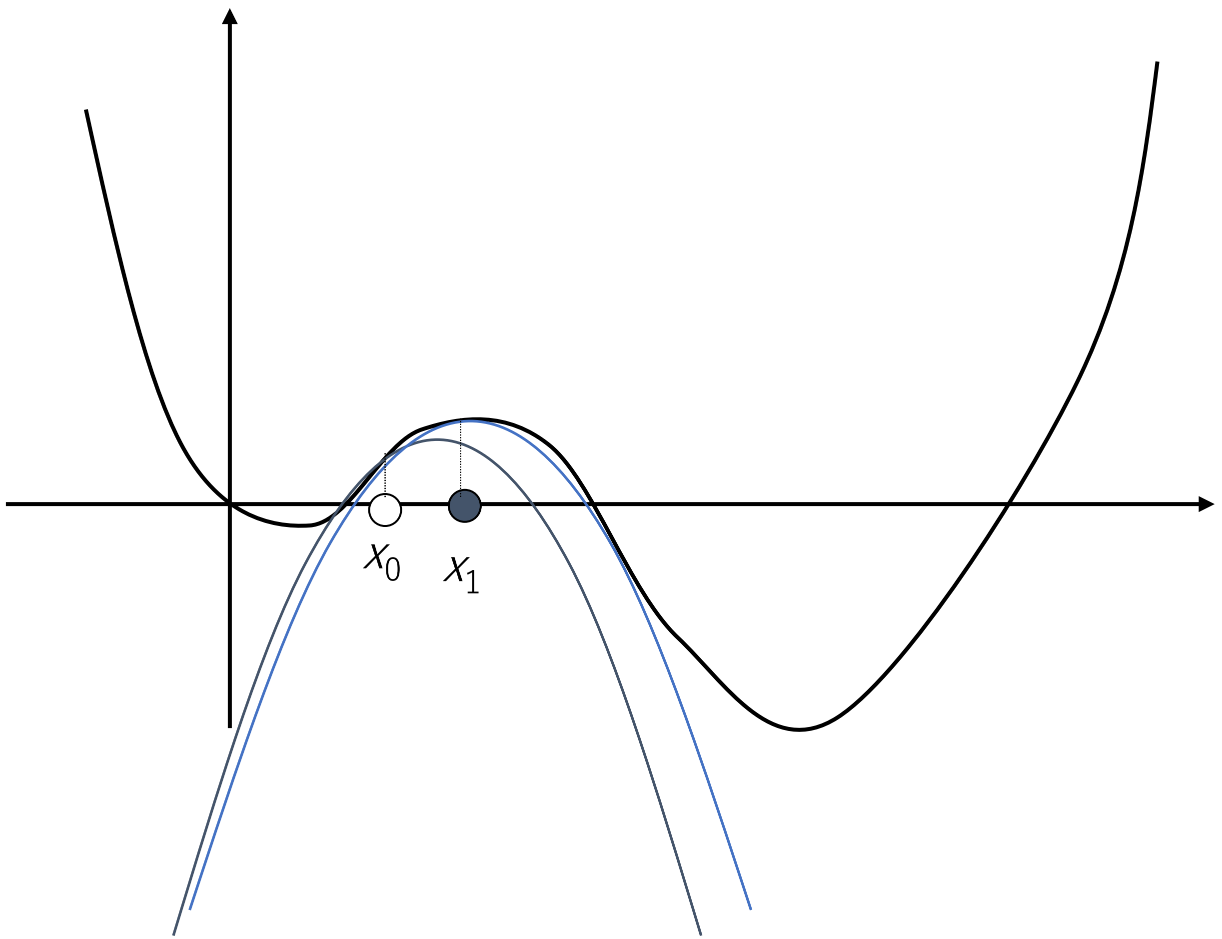
- 大域的収束性は持たない

  初期点の取り方が悪いと収束しない可能性あり，
  初期点によっては局所最適解でない停留点（極大点）に収束

- ヘッセ行列 $\nabla^2 f(\boldsymbol{x}_k)$が正則でない場合は探索方向が作れない

  ⇒準ニュートン法

初期点によっては局所的最小解でない停留点に収束



本日の実験課題

課題① Armijo 条件による直線探索法を含む最急降下法のプログラムを以下の２つの関数に対して完成させる：

$$f_1(x_1, x_2) = \frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} p & q \\ q & r \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}^T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \exp((x_1 - x_2)^2)$$
$$= \frac{1}{2}(px_1^2 + 2qx_1x_2 + rx_2^2) + c_1x_1 + c_2x_2 + \exp((x_1 - x_2)^2)$$

$$f_2(x_1, x_2) = \sum_{i=0}^{4} \sum_{j=0}^{2} a_{ij} x_1^i x_2^j$$

Armijo条件における定数は $\xi = 0.1, \tau = 0.5$ とし，
アルゴリズムの停止条件は $\|\nabla f(x_k)\| < 10^{-8}$ とすること

課題② 完成したプログラムをランダムデータ（後述）に対して適用し，ふるまいを観察する．プログラムの言語は何でもよい．

本日の実験課題

課題③ ニュートン法のプログラムを，①と同じ２つの関数に対して完成させる．プログラムの言語は何でもよい．

アルゴリズムの停止条件は以下の３つのいずれかが起きたとき，とせよ．

$\|\nabla f(x_k)\| < 10^{-8}$，　$\nabla^2 f(\boldsymbol{x}_k)$ が非正則，反復回数 $\boldsymbol{k}$ が 500を超えた

（最後のケースはニュートン法が収束しない場合）

課題④ 完成したプログラムを③と同じランダムデータに対して適用し，ふるまいを観察する．

※データ発生プログラム等は 数理計画法（第２回）.zipを展開

実験で用いるデータについて

- $f_1(x_1, x_2) = \frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} p & q \\ q & r \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}^T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \exp((x_1 - x_2)^2)$

  プログラム rand_exp.c を用いて作成したもの

  ランダムな Q, c と３つの初期点が３組発生される

  乱数の初期化は学籍番号を用いる

```
matrix Q
   6.0    2.0
   2.0   10.0
vector c
   6.0   -8.0
three initial solutions
  -5.0   -3.0
  -1.0   -4.0
 -11.0   -4.0
```

- $f_2(x_1, x_2) = \sum_{i=0}^{4} \sum_{j=0}^{2} a_{ij} x_1^i x_2^j$ ：

  プログラム rand_quart.c を用いて作成したもの

  ランダムな 5×3行列 Aと３つの初期点が２組

  ＋LETUS上のデータ１組

  各自，学籍番号の下２桁のデータを使用

  学籍番号 4615018 → data18.txt

$a_{01} = 0,$
$a_{01} = 28,$
$\cdots,$
$a_{40} = 1$

```
coefficient matrix A
   0.0   28.0    1.0
 -11.0    6.0    0.0
  16.0    0.0    0.0
  10.0    0.0    0.0
   1.0    0.0    0.0
three initial solutions
  37.0  -13.0
 -19.0   28.0
 -13.0   -6.0
```

数理計画法（第２回）.zip

プログラムの作成にあたって

- 各反復において以下の情報を出力するとよい：
  - 反復の回数(k)
  - 現在点(x_k)の座標と目的関数値
  - 勾配ベクトルの成分，ノルム
  - 点の差(x_k+1 – x_k)
- 最急降下法は反復回数が非常に多くなるので（確認した後），出力は一定反復ごと(100回ごとなど）になるように制御する．
- レポートの作成上，ファイルにも出力するとよい．
- 注意：他人のプログラムを写すことは不正行為，処罰の対象になる．写したことが発覚した場合は，写した方だけでなく，元のプログラムを提供した方も罰せられる可能性がある．
他人にはプログラムを絶対に渡さないように！

レポートの作成

(1) 実験目的と数学的理論の概要を述べよ.

(2) **課題①**で作成したプログラムをLETUSの所定の場所に提出し, さらにレポートにおいては用意した関数やクラス，工夫した点について述べること.

(3) **課題②**でプログラムをデータ適用した結果に基づいて，アルゴリズムのふるまいや性能について考察せよ.

考察には次ページの項目を含めること.

レポートの作成

関数f_1, f_2それぞれに対して，いくつかのデータを選び，以下の項目を含めよ．

- 反復回数と目的関数値のグラフ
- 勾配ベクトルのノルムがある程度小さく（0.05程度）なってから，最急降下法が生成する点列を $x_1 - x_2$ 平面にプロットしたもの（１つのデータでよい，プロットする範囲の大きさは非常に小さい ：10^{-3}四方程度）
- 凸性に関する考察．関数が凸か否かや，凸の場合に保証される性質が成り立っているか，また，得られた解が局所最適解かそれとも大域的最適解かなど，各自の自由な観点で議論して良い．

レポートの作成

(4) **課題②**で作成したプログラムをLETUSの所定の場所に提出し，さらに (2) と同様に用意した関数やクラス，工夫した点について述べること．

(5) **課題④**でプログラムをデータに適用した結果について，(3) と同様に考察をすること．その際，ニュートン法と最急降下法との違いを意識すること．

レポートの作成

(4) 課題②で作成したプログラムをLETUSの所定の場所に提出し，さらに (2) と同様に用意した関数やクラス，工夫した点について述べること．

(5) **課題④**でプログラムをデータに適用した結果について，(3) と同様に考察をすること．その際，ニュートン法と最急降下法との違いを意識すること．

提出するプログラムの名前について

提出するプログラムは

学生番号_名前_アルゴリズム

とすること.

例：

4615001_〇〇〇〇_SD.c　（最急降下法）

4615001_〇〇〇〇_newton.py　（ニュートン法）