情報数学II
第11回 最適化

小野順貴
(ONO, Nobutaka)
東京都立大学　システムデザイン学部
情報科学科　教授
onono@tmu.ac.jp

最適化問題

- ある関数を最小（もしくは最大）にする変数を求める問題

$$\boldsymbol{x}^* = \mathrm{argmax}_{\boldsymbol{x}} f(\boldsymbol{x})$$

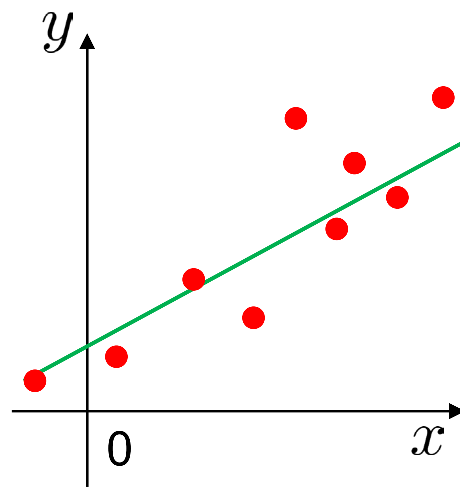
  - $\boldsymbol{x}$ は一般に、実数、または複素数のベクトル
  - $f(\boldsymbol{x})$ は実数（スカラー）。目的関数と呼ばれる。
  - 問題によっては何らかの制約がある場合も
    - 例） $|\boldsymbol{x}| = 1$ など。「制約付き最適化問題」と呼ばれる。
  - 工学的に様々な問題が最適化問題として定式化される
    - 誤差最小、コスト最小、利得最大、最尤推定、・・・

線形回帰分析

- 1変数の線形回帰問題とは、与えられたデータ (x_k, y_k) $(k = 1, \cdots, K)$に対し、線形回帰式 $y = ax + b$ が最もよくフィットするようにパラメータ (a, b) を求める問題。
- フィッティングの良さとしては二乗誤差がよく用いられる。

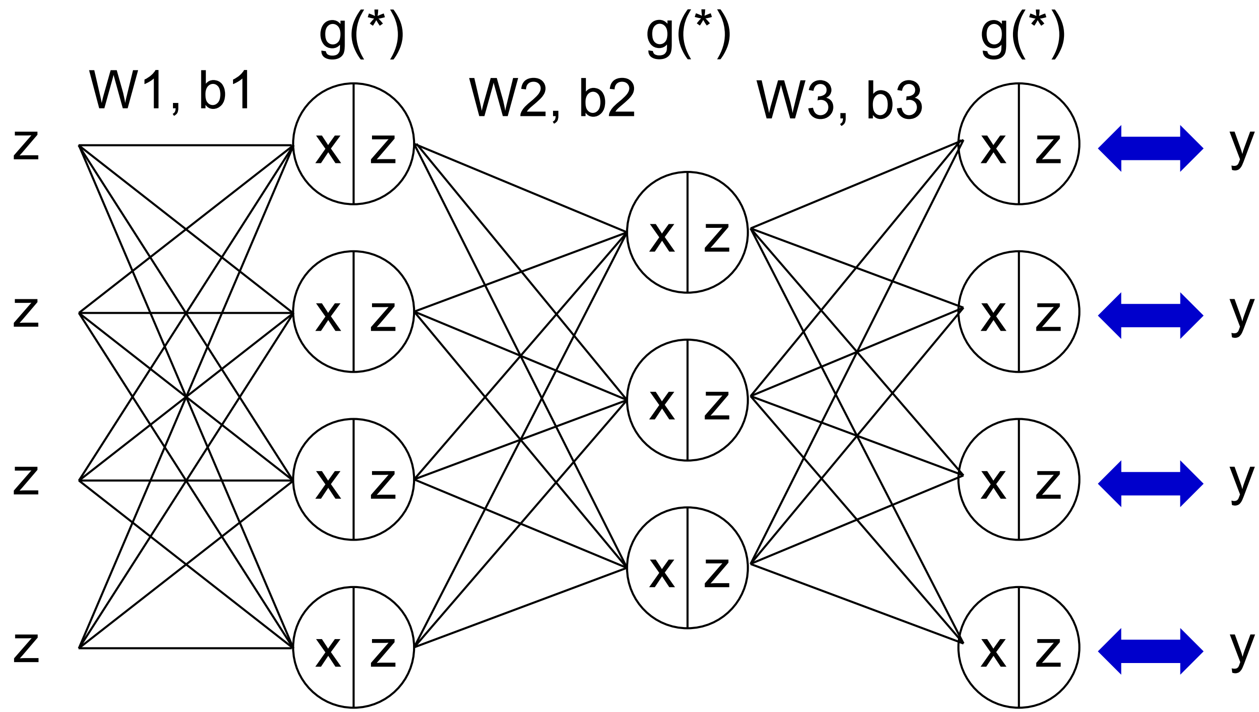
$$f(a,b) = \sum_{k=1}^{K} (\underbrace{ax_k + b}_{\text{線形回帰モデル}} - \underbrace{y_k}_{\text{データ}})^2$$

モデルとデータの二乗誤差



- どうやって二乗誤差を最小にする (a, b) を求めるか？

最適化問題の例:ニューラルネットワーク

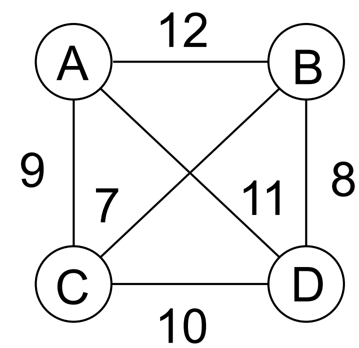


目的関数

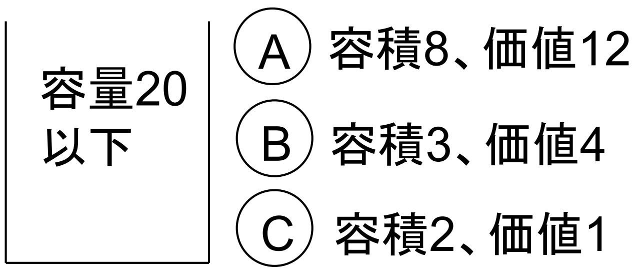
$$f(W_1, \boldsymbol{b}_1, W_2, \boldsymbol{b}_2, W_3, \boldsymbol{b}_3) = \sum_{i=1}^{n} \left(\boldsymbol{g}(W_3 \boldsymbol{g}(W_2 \boldsymbol{g}(W_1 \boldsymbol{z}_i + \boldsymbol{b}_1) + \boldsymbol{b}_2) + \boldsymbol{b}_3) - \boldsymbol{y}_i\right)^2$$

組み合わせ最適化問題

- 変数が連続値でなく離散である問題
  - 巡回セールスマン問題：全部の節点を1度ずつ通る巡回路の中でコスト最小のものを求める問題
    A→B→C→D→A：12+7+10+11=40
    A→B→D→C→A：12+8+10+9=39
    A→C→B→D→A：9+7+8+11=35 ← 最小



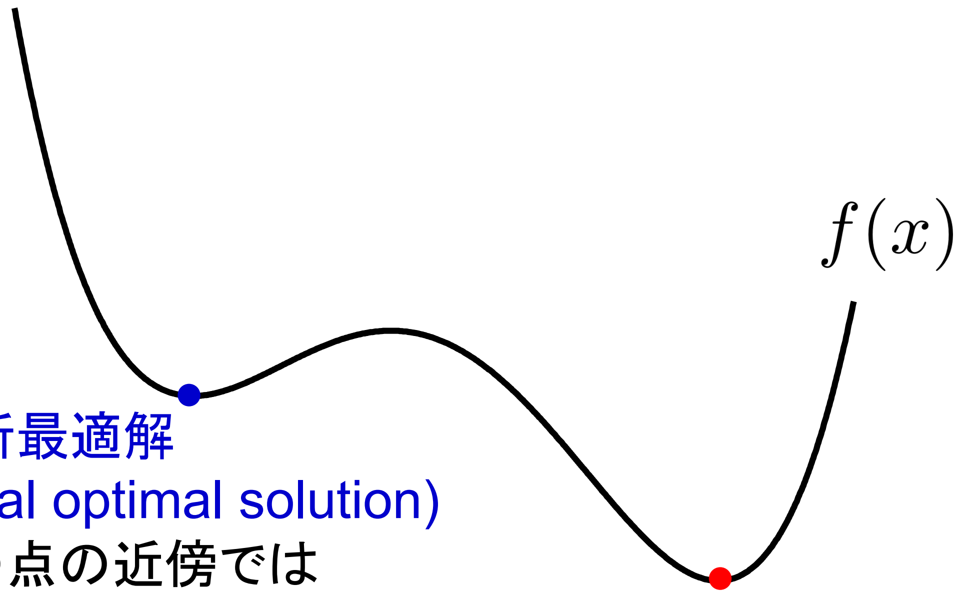
  - ナップサック問題：ある容量のナップサックに品物をいれ、価値を最大化する品物リストを求める問題
    Ax2+Cx2：容量20、価値26
    Ax1+Bx4：容量20、価値28 ←最大



- 本講義では詳細は扱わない
  - キーワード：動的計画法、線形計画法、分枝限定法など

大域的最適解と局所最適解(1/2)

- １次元の例



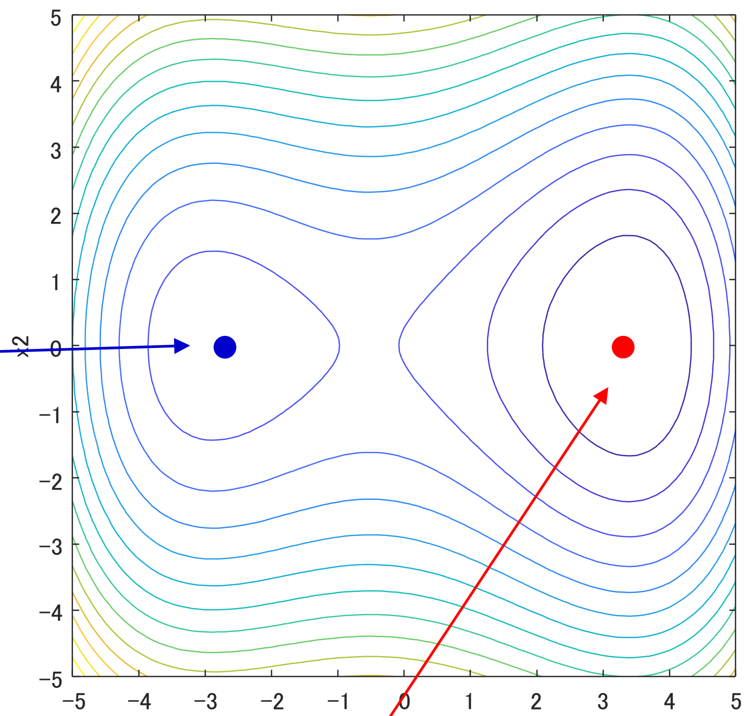
局所最適解
(local optimal solution)
この点の近傍では
最小（もしくは最大）

大域的最適解
(global optimal solution)
定義域（もしくは探索領域）全体
の中で最小（もしくは最大）

大域的最適解と局所最適解(2/2)

- 2次元の例
  （右図は等高線表示）

$f(x)$



局所最適解
(local optimal solution)
この点の近傍では
最小（もしくは最大）

大域的最適解
(global optimal solution)
定義域（もしくは探索領域）全体
の中で最小（もしくは最大）

局所最適解の必要条件

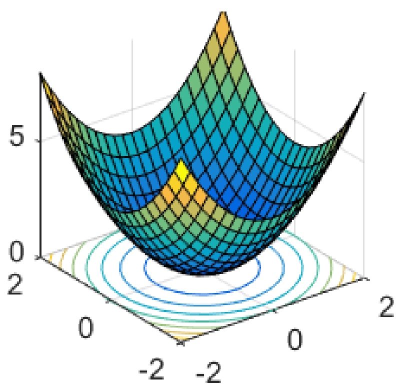
- 制約なし最適化で目的関数が微分可能な場合

$$\frac{\partial f}{\partial x_1} = \cdots = \frac{\partial f}{\partial x_N} = 0$$

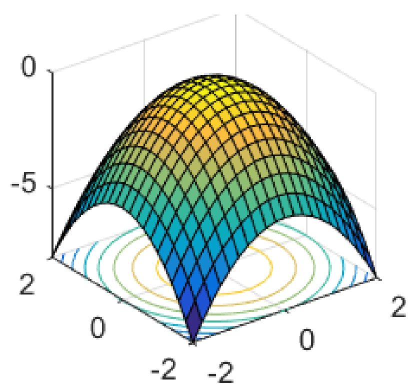
  - もし微分が0でないとすると、勾配方向に、より値が小さい（もしくは大きい）点が存在する

- 上記は必要条件。十分ではない

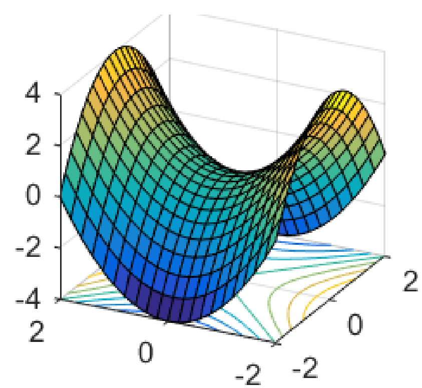
局所最小



局所最大



鞍点



https://stats.stackovernet.com/ja/q/66249

しらみつぶし法(exhaustive search)

- 解の候補となりうる $\boldsymbol{x}$ をしらみつぶしに調べる
  - 連続変数 $\boldsymbol{x}$ に対しては、真の意味でしらみつぶしはできないので、適当な範囲と探索幅を決めて解を探す
  - 例) $\boldsymbol{x} = (x_1, x_2)$ のとき、範囲と探索幅を例えば

$$-10 \le x_1 \le 10,\ -10 \le x_2 \le 10,\ \Delta x = 0.01$$

と決め、

$$(x_1, x_2) = (n_1 \Delta x,\ n_2 \Delta x)$$
$$-10/\Delta_x \le n_1 \le 10/\Delta_x$$
$$-10/\Delta_x \le n_2 \le 10/\Delta_x$$

の中から $f(\boldsymbol{x})$ を最大にするものを探す、など

  - 計算時間がかかるが、単純で $\boldsymbol{x}$ の次元が小さいときは有効

MATLAB練習19:しらみつぶし法

$$f(x_1, x_2) = x_1 e^{-(x_1^2 + x_2^2)}$$ の最小値をしらみつぶし法で求める

```
>> clear;                                  % 変数のクリア
>> x1=-5:0.1:5;                            % 探索範囲の設定
>> x2=-5:0.1:5;                            % 探索範囲の設定
>> [X1,X2]=meshgrid(x1,x2);
>> Z=X1.*exp(-(X1.^2+X2.^2));              % 関数値の計算
>> [C,I]=min(Z(:));                        % Z(:)で2次元配列を1次元化している
                                           % Cに最小値、Iにそのインデックスが返る
>> [x1idx,x2idx]=ind2sub(size(Z),I);       % Iは1次元化した配列上のインデックス
                                           % なので、2次元のインデックスに戻す
>> x1(x1idx)
ans =
    0
>> x2(x2idx)
ans =
   -0.7100
```

% $(x_1, x_2) = (0, -0.71)$ が解として求まった

探索範囲をもっと細かくしてやってみること

直線探索 (line search)

- 変数(ベクトル) $\boldsymbol{x}$ をある直線上で最適化する手法
- 直線探索を使った最適化の例

1. 探索方向 $\boldsymbol{p}$ を決める
   例えば勾配方向

$$\boldsymbol{p} = (\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_N})^t$$

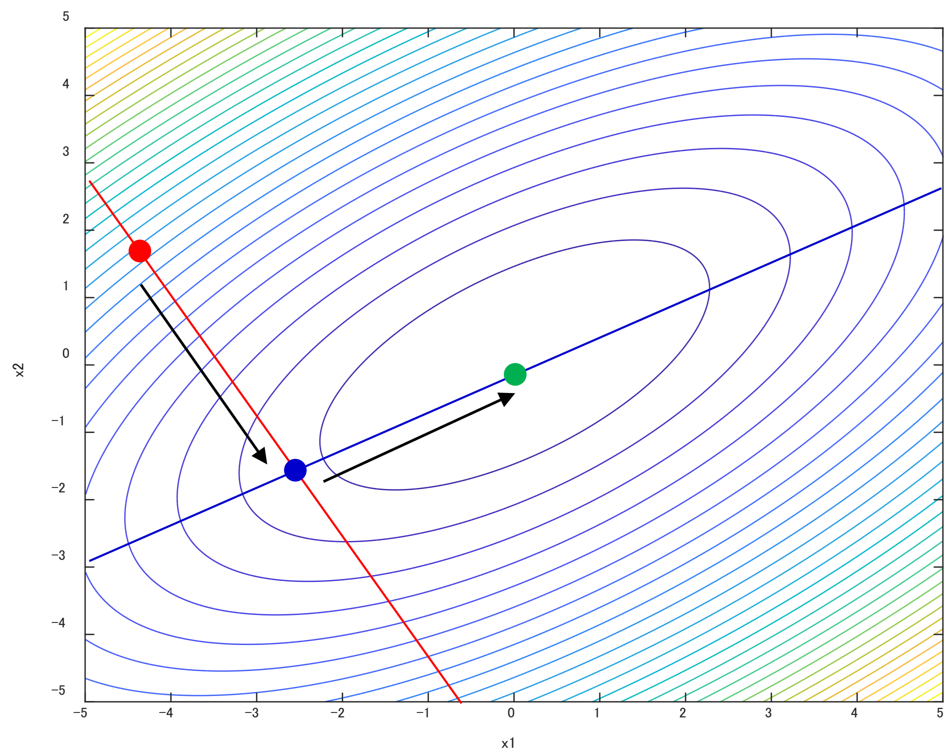
2. $\boldsymbol{p}$ 方向で $f(\boldsymbol{x})$を最適化
   (直線探索)

$$a^* = \mathrm{argmin}_a f(\boldsymbol{x} + a\boldsymbol{p})$$

3. $\boldsymbol{x}$ を更新し 1. に戻る

$$\boldsymbol{x} \leftarrow \boldsymbol{x} + a^*\boldsymbol{p}$$

2次元の例



座標降下法

- 変数(ベクトル) $\boldsymbol{x} = (x_1, \cdots, x_N)$ を各要素(座標)ごとに最適化する手法。直線探索の一種ともみられる。
  - 疑似コード

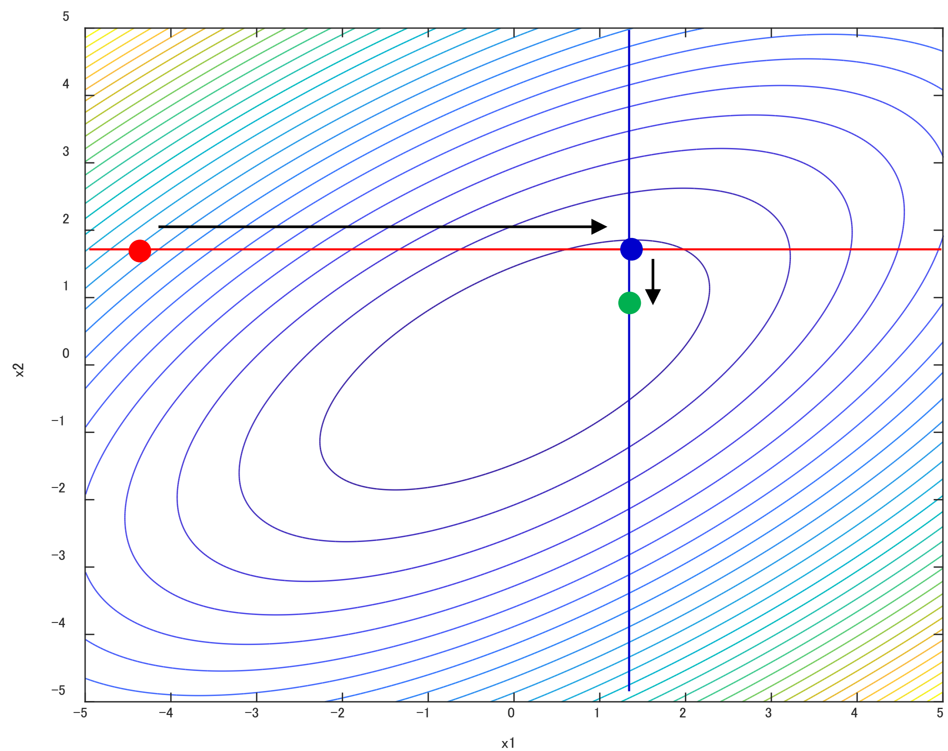
For $i = 1 : n$

$x_i \leftarrow \mathrm{argmin}_{x_i} f(\boldsymbol{x})$

End

ある x_i 以外の値は固定し、のみについて最適化する

2次元の例



直線探索の方法

- １次元なのでしらみつぶし法の適用も可能
- 黄金分割探索

MATLAB練習20：座標降下法

$f(x_1, x_2) = x_1 e^{-(x_1^2+x_2^2)}$ の最小値を座標降下法で求める

```
>> clear;                          % 変数のクリア
>> x2=-5;                          % 適当な初期値を設定
```

```
>> x1=-5:0.1:5;                    % 探索範囲の設定
>> z=x1.*exp(-(x1.^2+x2.^2));      % 関数値の計算
>> [c,idx]=min(z);                 % cに最小値、idxにそのインデックスが返る
>> x1=x1(idx)                      % x1を更新
```

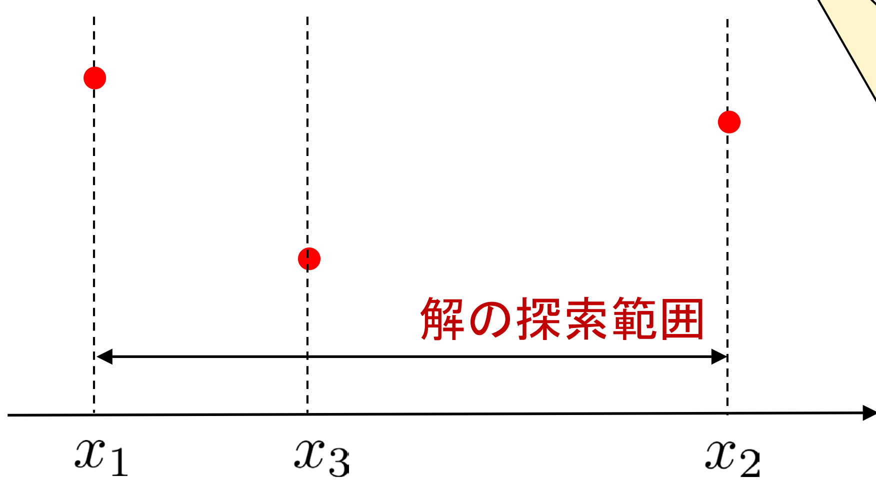
↕ 繰り返し行う

```
>> x2=-5:0.1:5;                    % 探索範囲の設定
>> z=x1.*exp(-(x1.^2+x2.^2));      % 関数値の計算
>> [c,idx]=min(z);                 % cに最小値、idxにそのインデックスが返る
>> x2=x2(idx)                      % x2を更新
```

黄金分割探索 (1/4)

- 単峰型関数の1次元最適化問題の効率的解法
  - 目的関数が単峰型（極小値が1つ）と仮定する
- 常に4点を評価しながら範囲を狭めていく
- $f(x_1) > f(x_3), f(x_3) < f(x_2)$ なら
  極小値は x_1, x_2 の間にある

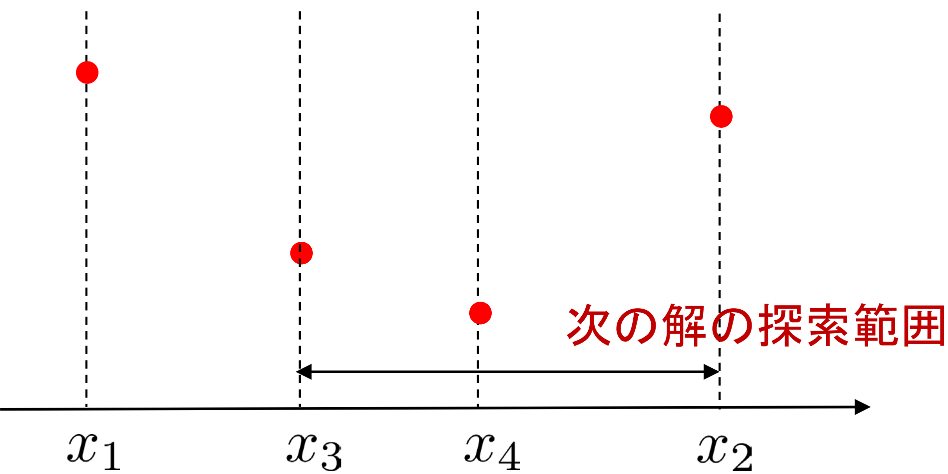
極小値が x_1, x_2 の外なら
$f(x_1) > f(x_3) > f(x_2)$
$f(x_1) < f(x_3) < f(x_2)$
のどちらかになっているはず



黄金分割探索 (2/4)

- ４点目で関数値を計算する。ここでは x_3, x_2 の間にとる

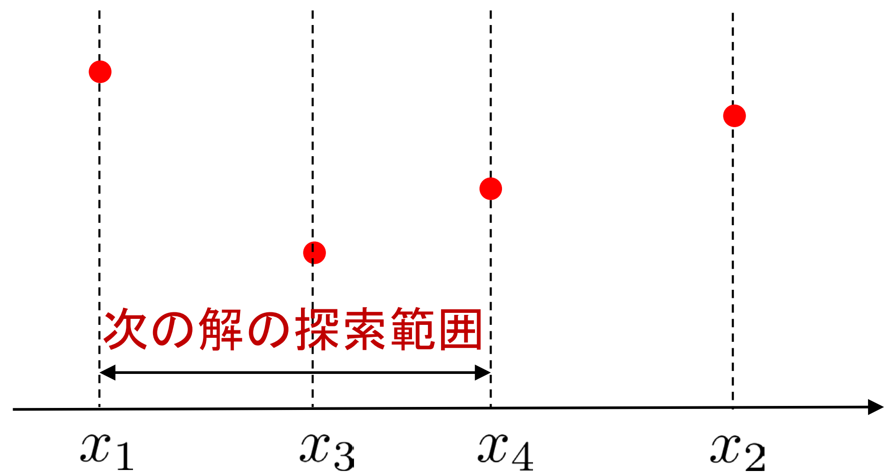
場合1: $f(x_3) > f(x_4)$



$$f(x_3) > f(x_4), f(x_4) < f(x_2)$$

なので、解は x_3, x_2 の間にある

場合2 $f(x_3) < f(x_4)$

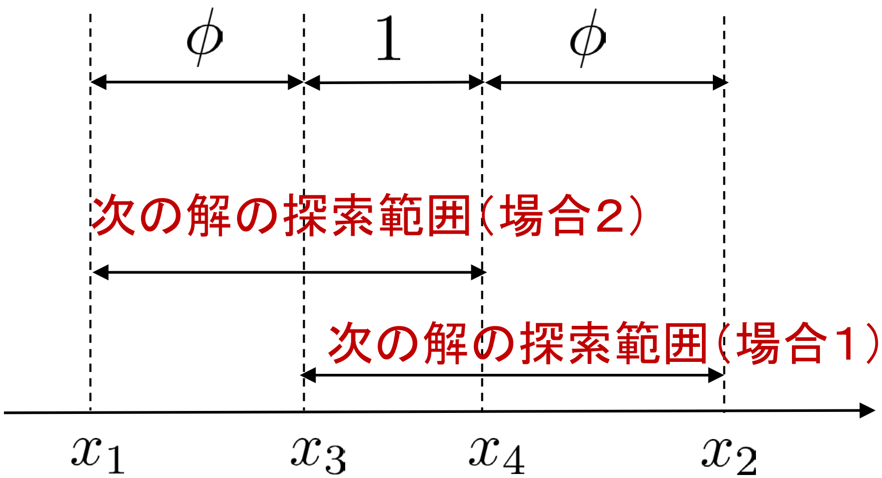


$$f(x_1) > f(x_3), f(x_3) < f(x_4)$$

なので、解は x_1, x_4 の間にある

黄金分割探索 (3/4)

- 探索幅の比率が一定であるための条件



$$\frac{\phi}{1+2\phi} = \frac{1}{1+\phi}$$

$$\phi^2 - \phi - 1 = 0$$

$$\phi = \frac{1+\sqrt{5}}{2}$$

黄金比

黄金分割探索 (4/4)

- アルゴリズムのまとめ

1. 最適値を含むと考えられる範囲 x_1, x_2 を定める
2. $$\begin{aligned} x_3 &= \gamma x_1 + (1-\gamma)x_2 \\ x_4 &= (1-\gamma)x_1 + \gamma x_2 \end{aligned} \qquad \gamma = \frac{1+\phi}{1+2\phi} = \frac{\sqrt{5}-1}{2}$$
3. If $f(x_3) > f(x_4)$

   then $$\begin{aligned} x_1 &\leftarrow x_3 \\ x_3 &\leftarrow x_4 \\ x_4 &\leftarrow (1-\gamma)x_1 + \gamma x_2 \end{aligned}$$

   else $$\begin{aligned} x_2 &\leftarrow x_4 \\ x_4 &\leftarrow x_3 \\ x_3 &\leftarrow \gamma x_1 + (1-\gamma)x_2 \end{aligned}$$
4. $|x_3 - x_4| \geq \varepsilon$ なら3へ。そうでなければ、$x^* = \dfrac{x_3 + x_4}{2}$

   適当な閾値

最急降下法

- 微分の値を用いて変数を更新する反復法
- N変数の目的関数: $f(\boldsymbol{x}) = f(x_1, \cdots, x_N)$
- 勾配ベクトル

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{pmatrix}$$

- 最急降下法の更新

$$\boldsymbol{x} \leftarrow \boldsymbol{x} - \mu \nabla f$$

再掲

MATLAB練習12:最急降下法

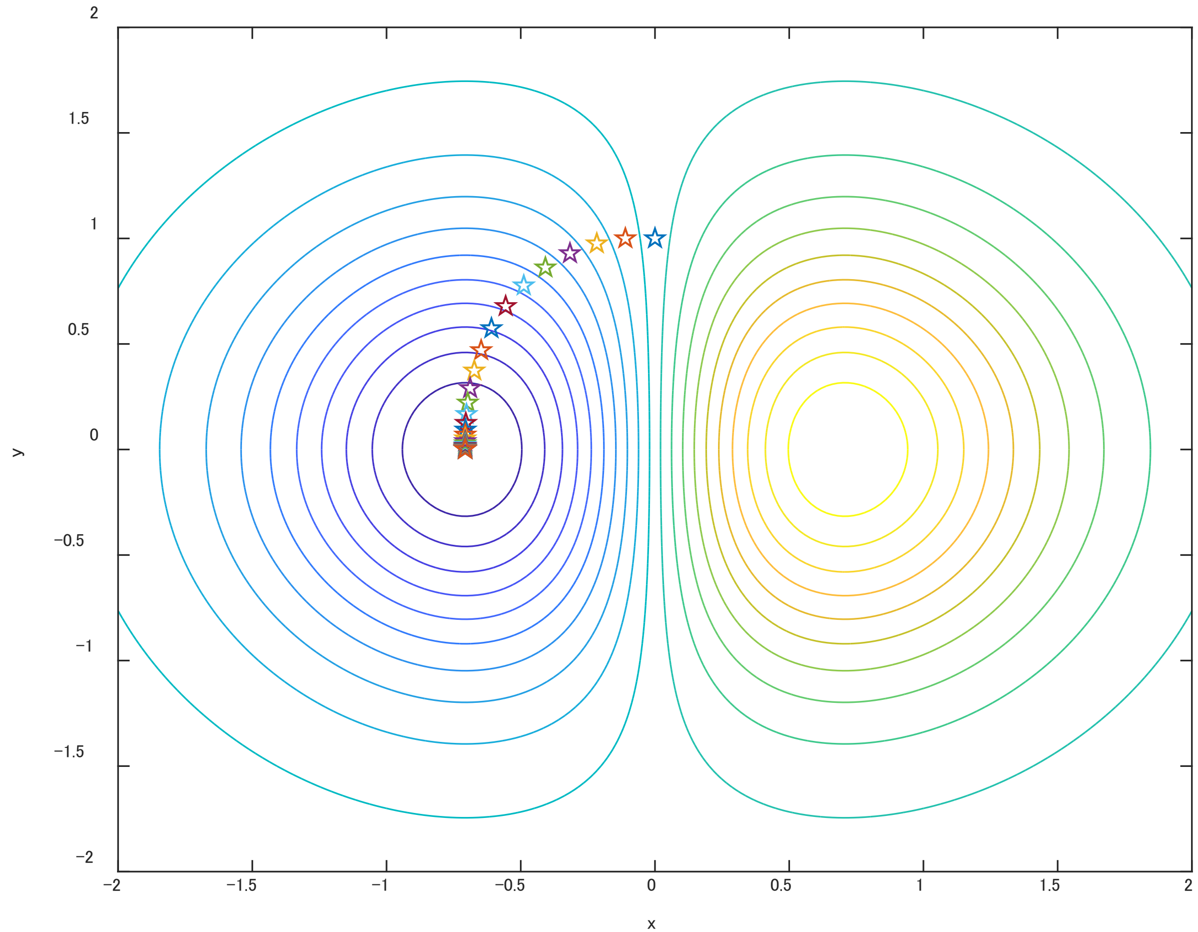
```
>> clear;                                % 変数のクリア
>> x=1; y=1;
>> x*exp(-x^2-y^2)                       % 関数の値を計算
ans =
   0.1353

>> dx=(1-2*x^2)*exp(-x^2-y^2);           % x微分の計算
>> dy=    -2*x*y*exp(-x^2-y^2);          % y微分の計算
>> mu=0.3;                               % ステップサイズ
>> x=x-mu*dx;                            % 勾配に沿って点を移動
>> y=y-mu*dy;                            % 勾配に沿って点を移動
>> x*exp(-x^2-y^2)                       % 関数の値を計算
ans =
   0.1095
```

勾配に沿って点を移動したことにより、確かに値が小さくなっている

再掲

MATLAB練習13：最急降下法2

- graddsc.m をダウンロードして実行
- 最急降下法により点が極小点に徐々に近づいて行く様子がわかる
- 初期値やステップサイズを変えてやってみること



等式制約付き最適化問題

- ある等式制約条件のもと、関数を最小(もしくは最大)にする変数を求める問題

$$\boldsymbol{x}^* = \mathrm{argmin}_{\boldsymbol{x}} f(\boldsymbol{x})$$

ただし　$g(\boldsymbol{x}) = 0$

- 以下では、$f(\boldsymbol{x}), g(\boldsymbol{x})$ は微分可能とする。

- 例）$f(x_1, x_2) = x_1 + x_2$ を最小とする x_1, x_2 を求めよ。

  ただし、$g(x_1, x_2) = x_1^2 + x_2^2 - 1 = 0$

ラグランジュの未定乗数法

- $g(\boldsymbol{x}) = 0$ の制約下で $f(\boldsymbol{x})$ を最大化（最小化）したい

- ラグランジュの未定乗数λを導入し

$$J(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$$

を考え、

$$\frac{\partial J}{\partial x_1} = 0,\ \frac{\partial J}{\partial x_2} = 0,\ \cdots,\ \frac{\partial J}{\partial x_N} = 0,\ \frac{\partial J}{\partial \lambda} = 0$$

を解く。

ラグランジュの未定乗数法の例題(1/2)

- $f(x_1, x_2) = x_1 + x_2$ を最小とする x_1, x_2 を求めよ。

ただし、$g(x_1, x_2) = x_1^2 + x_2^2 - 1 = 0$

- $$\begin{aligned} J(x_1, x_2, \lambda) &= f(x_1, x_2) + \lambda g(x_1, x_2) \\ &= (x_1 + x_2) + \lambda(x_1^2 + x_2^2 - 1) \end{aligned}$$

$$\frac{\partial J}{\partial x_1} = 1 + 2\lambda x_1 = 0 \Rightarrow x_1 = -\frac{1}{2\lambda}$$

$$\frac{\partial J}{\partial x_2} = 1 + 2\lambda x_2 = 0 \Rightarrow x_2 = -\frac{1}{2\lambda}$$

$$\frac{\partial J}{\partial \lambda} = x_1^2 + x_2^2 - 1 = 0 \Longrightarrow \left(-\frac{1}{2\lambda}\right)^2 + \left(-\frac{1}{2\lambda}\right)^2 - 1 = 0$$

ラグランジュの未定乗数法の例題(2/2)

$$x_1 = -\frac{1}{2\lambda}$$

$$x_2 = -\frac{1}{2\lambda}$$

$$\Rightarrow (x_1, x_2) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$$

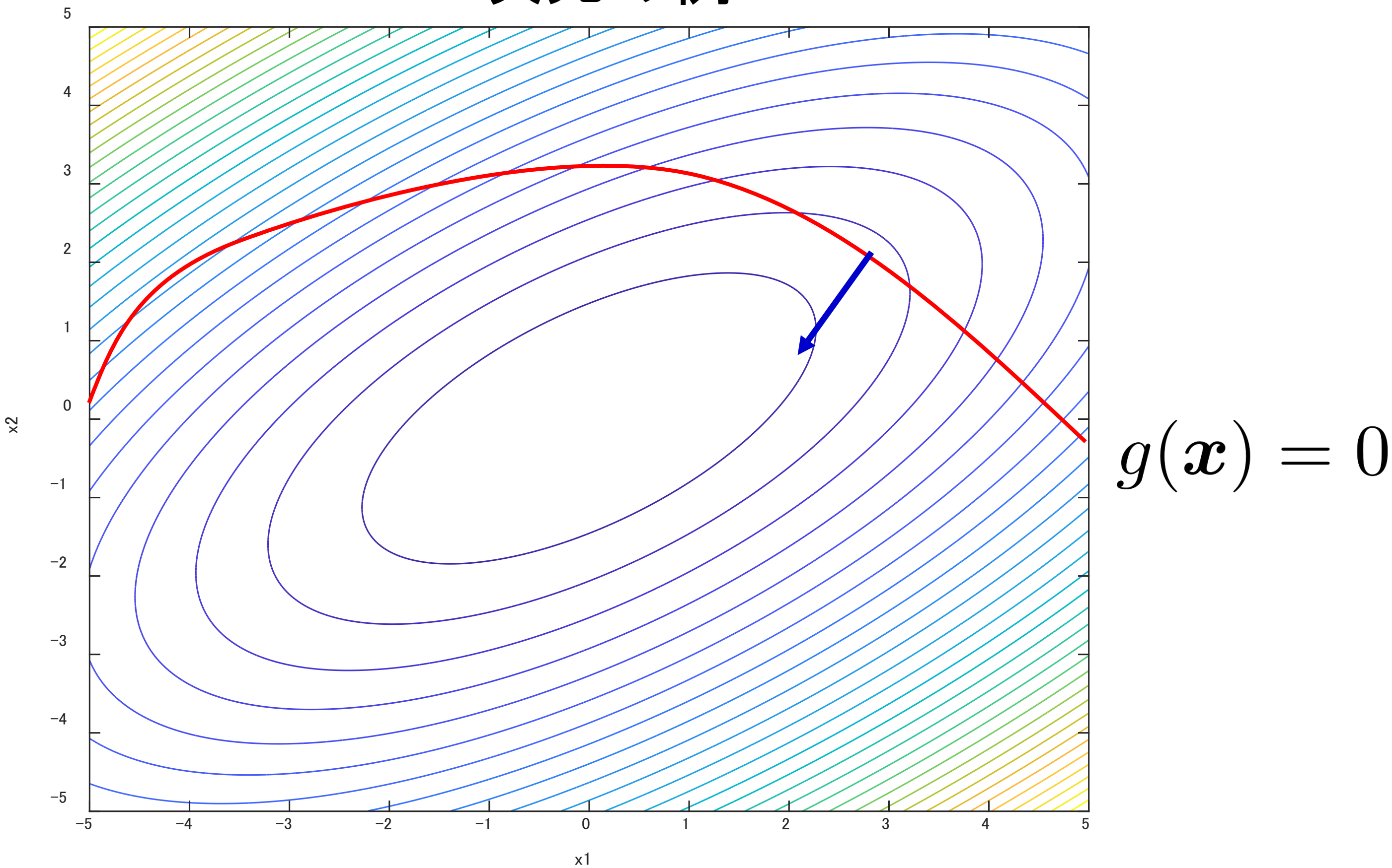
$$\left(-\frac{1}{2\lambda}\right)^2 + \left(-\frac{1}{2\lambda}\right)^2 - 1 = 0 \Rightarrow \lambda = \pm\frac{1}{\sqrt{2}}$$

片方が最大値、
もう片方が最小値

直感的な理解 (1/2)

探索領域が赤線に制約されているとき、どこが最小？

２次元の例

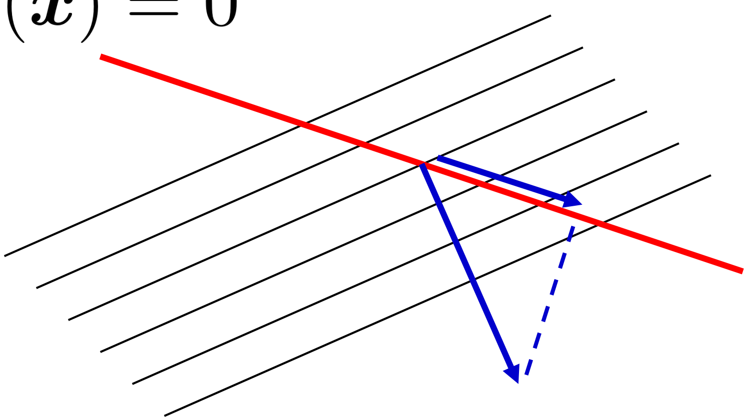


直感的な理解 (2/2)

ある点のまわりを拡大すると･･･

極小点でない例

$g(\boldsymbol{x}) = 0$

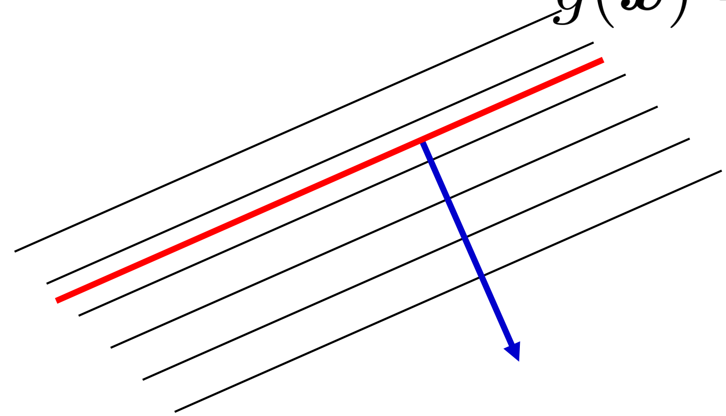


勾配ベクトル

制約条件の領域内で
勾配ベクトル成分が
0でない

極小点の例

$g(\boldsymbol{x}) = 0$



勾配ベクトル

制約条件の領域内で
勾配ベクトル成分が0
つまり $\nabla f(\boldsymbol{x}) = -\lambda \nabla g(\boldsymbol{x})$

勾配ベクトル

制約条件の領域の
法線ベクトル

接平面の法線ベクトルの説明

- 簡単のため2次元の例を示す

$g(x,y)=0$ という曲線(一般には曲面)の (x_0,y_0) での接線(接平面)を考える。(x_0,y_0) 周りでテーラー展開すると

$$g(x,y) = g(x_0,y_0) + \underbrace{g_x(x_0,y_0)}_{\text{gのxによる偏微分}}(x-x_0) + \underbrace{g_y(x_0,y_0)}_{\text{gのyによる偏微分}}(y-y_0) + \cdots$$

$$= \underbrace{\left(g_x(x_0,y_0)\ \ g_y(x_0,y_0)\right)}_{\text{接線と直交するベクトル(一般には法線ベクトル)}} \begin{pmatrix} x \\ y \end{pmatrix}$$

$$+ \left[\underbrace{g(x_0,y_0) - x_0 g_x(x_0,y_0) - y_0 g_y(x_0,y_0)}_{\text{定数項}} + \cdots\right]$$

よって、一般には $\nabla g(\boldsymbol{x})$ が制約条件gの節平面の法線ベクトル

ラグランジュの未定乗数法の例題2

- 周長が一定である長方形のうち、面積が最大のものは正方形であることを示せ。

目的関数: $f(x,y) = xy$

制約条件: $g(x,y) = 2x + 2y - L = 0$

$$J(x,y,\lambda) = f(x,y) + \lambda g(x,y)$$

$$\frac{\partial J}{\partial x} = y + 2\lambda = 0 \Rightarrow y = -2\lambda$$

$$\frac{\partial J}{\partial y} = x + 2\lambda = 0 \Rightarrow x = -2\lambda$$

$$\frac{\partial J}{\partial \lambda} = 2x + 2y - L = 0 \Rightarrow -8\lambda = L \Rightarrow \lambda = -\frac{L}{8}$$

$$x = y = \frac{L}{4}$$

他の例題

以下の制約付き最適化問題を解け。

1. $x - 2y = 3$ の制約の下、$f(x, y) = x^2 + y^2$ を最小化する (x, y) を求めよ。また、そのときの $f(x, y)$ の値を求めよ。

2. $x^2 - xy + y^2 = 1$ の制約の下、$f(x, y) = x + y$ を最小化する (x, y) を求めよ。また、そのときの $f(x, y)$ の値を求めよ。

まとめ

- 大域的最適解と局所最適解
- 最適化手法
  - しらみつぶし法
  - 直線探索
    - 座標降下法
    - 黄金分割探索
  - 最急降下法
  - ラグランジュの未定乗数法