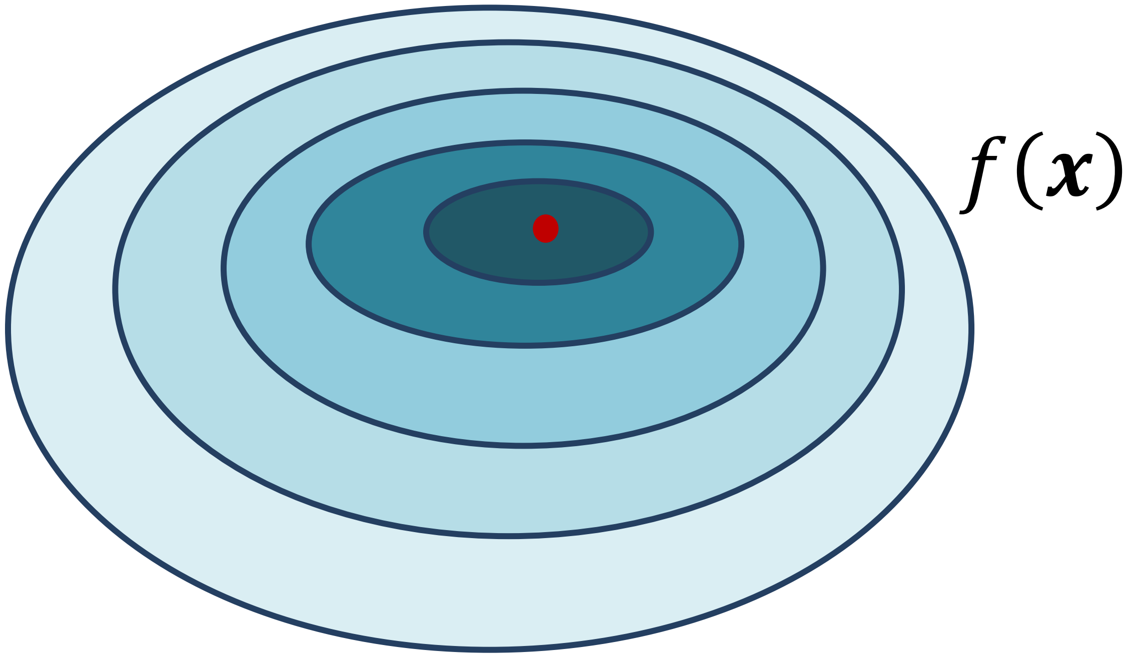
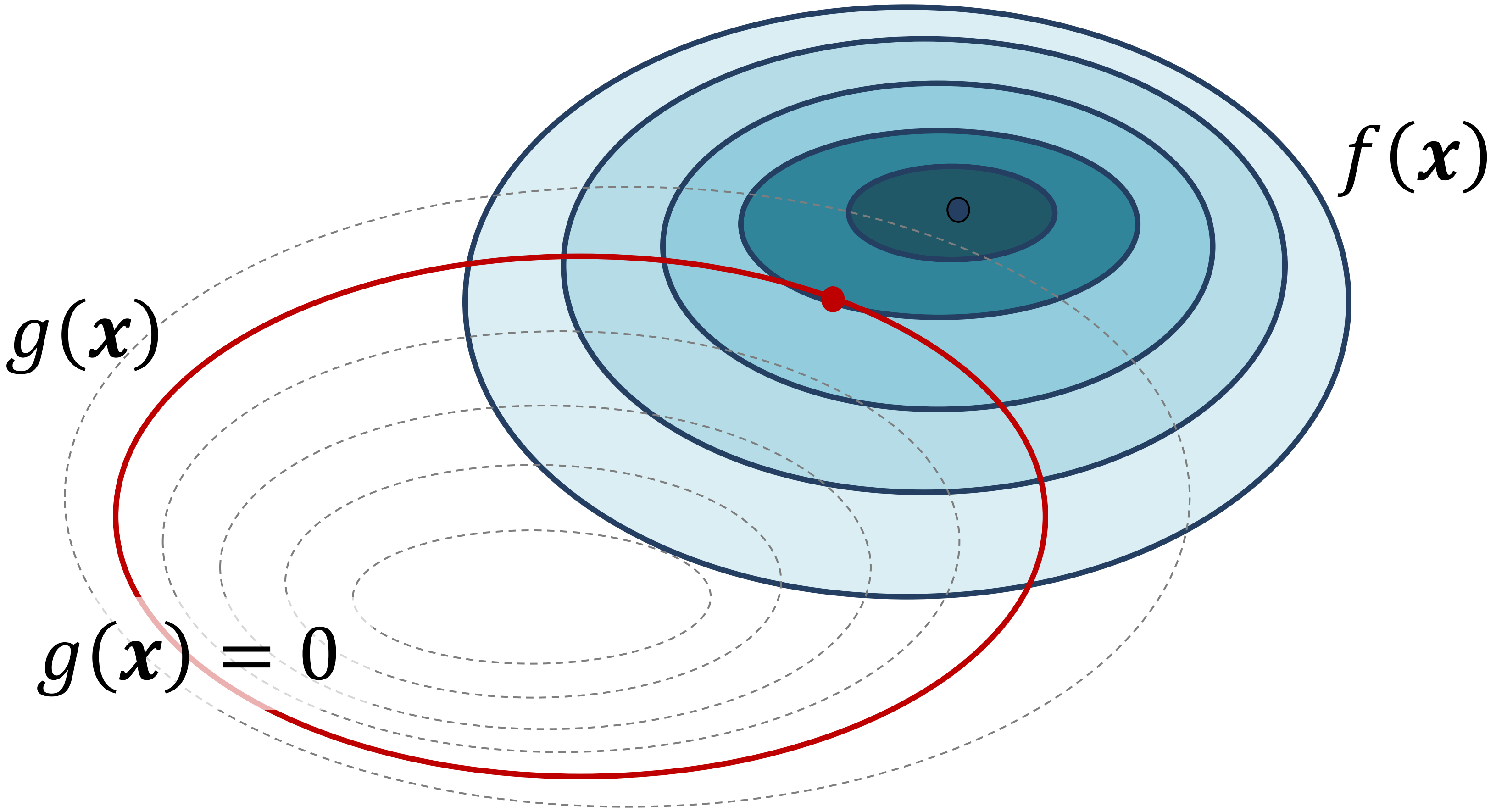
等式制約の最適化

前回まで：制約がない条件での最適化



今回 ：ある等式を満たす条件の下での最適化



上図，赤線上で $f(\boldsymbol{x})$ の最小値(最大値)を与える $\boldsymbol{x}$ は?

§ Lagrangeの未定乗数法

問題：

$$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$\text{s.t. } g(\boldsymbol{x}) = 0$$

$g(\boldsymbol{x}) = 0$ を満たすという条件での $f(\boldsymbol{x})$ の最大値を求める。

ラグランジュ関数

もともとの目的関数

常套手段：

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

制約

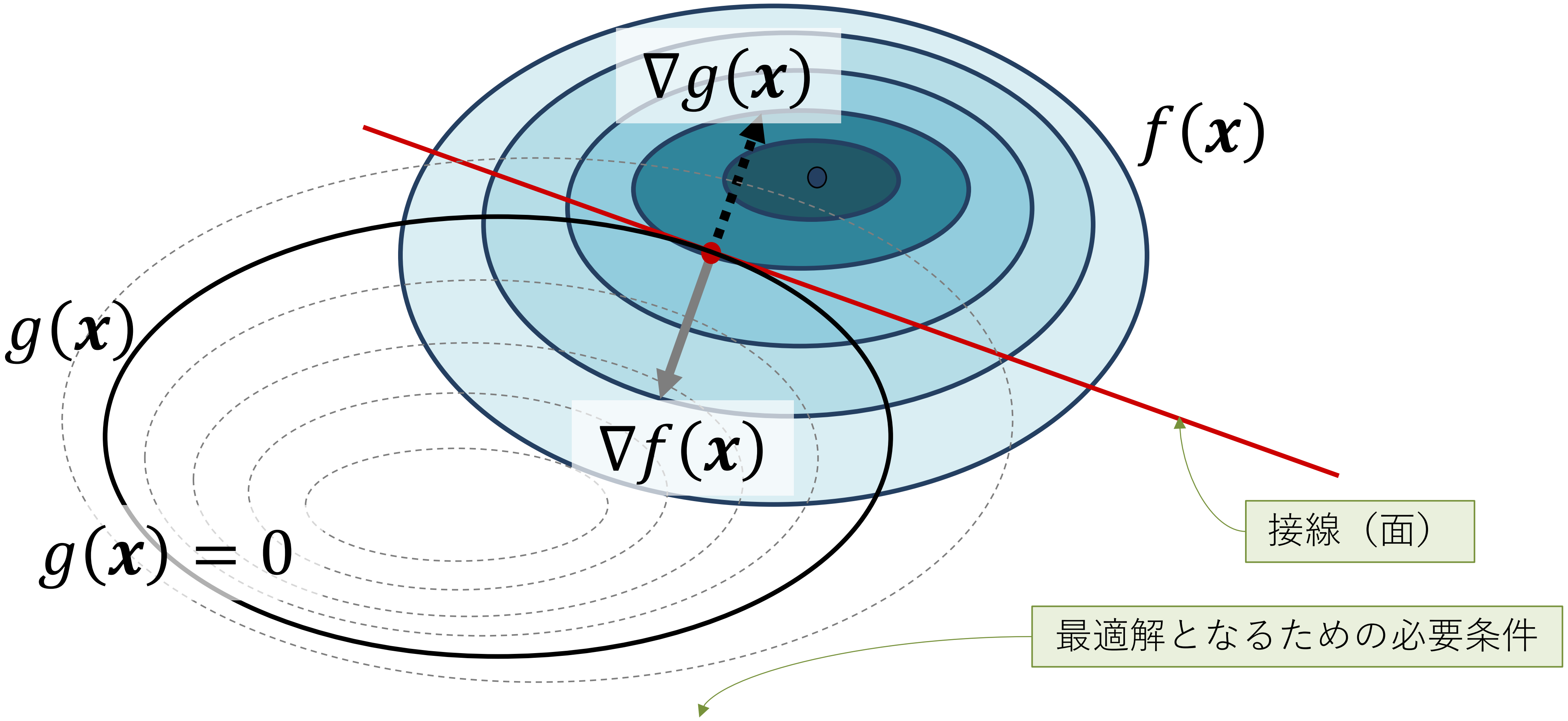
ラグランジュの未定定数

とおいて，

$$\frac{\partial}{\partial \boldsymbol{x}} L(\boldsymbol{x}, \lambda) = 0, \frac{\partial}{\partial \lambda} L(\boldsymbol{x}, \lambda) = g(\boldsymbol{x}) = 0$$

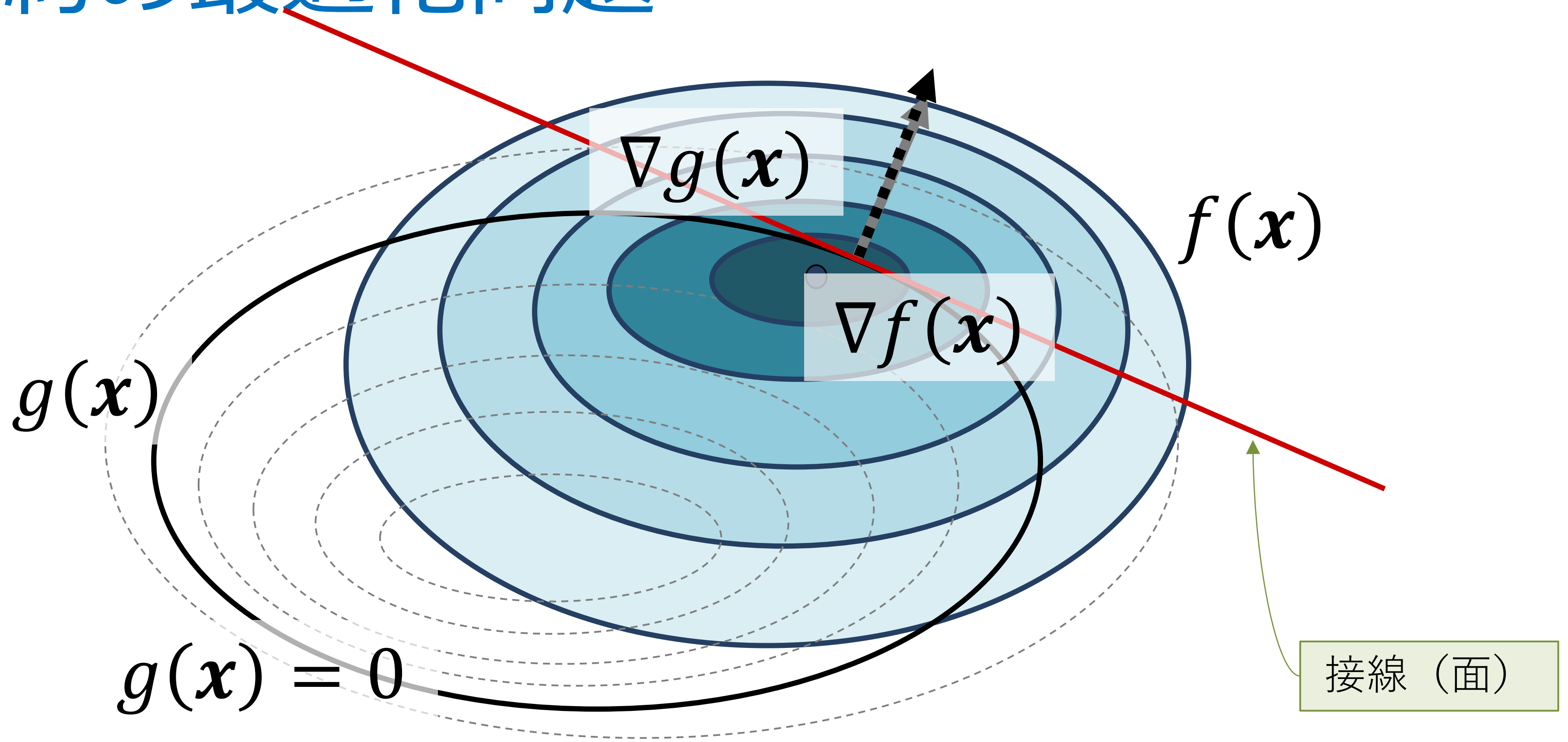
の連立を解く。

等式制約の最適化問題



$$\nabla g(\boldsymbol{x}) = -\lambda g(\boldsymbol{x}), \qquad g(\boldsymbol{x}) = 0$$

等式制約の最適化問題



最適解となるための必要条件

$$\nabla g(\boldsymbol{x}) = -\lambda g(\boldsymbol{x}), \qquad g(\boldsymbol{x}) = 0$$

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) \qquad \text{s.t.} \quad g(\boldsymbol{x}) = 0$$

$$\Rightarrow \quad \nabla f(\boldsymbol{x}) + \lambda \nabla g(\boldsymbol{x}) = 0, \qquad g(\boldsymbol{x}) = 0$$

$$\min_{\boldsymbol{x},\lambda} L(\boldsymbol{x},\lambda)\ , \qquad L(\boldsymbol{x},\lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$$

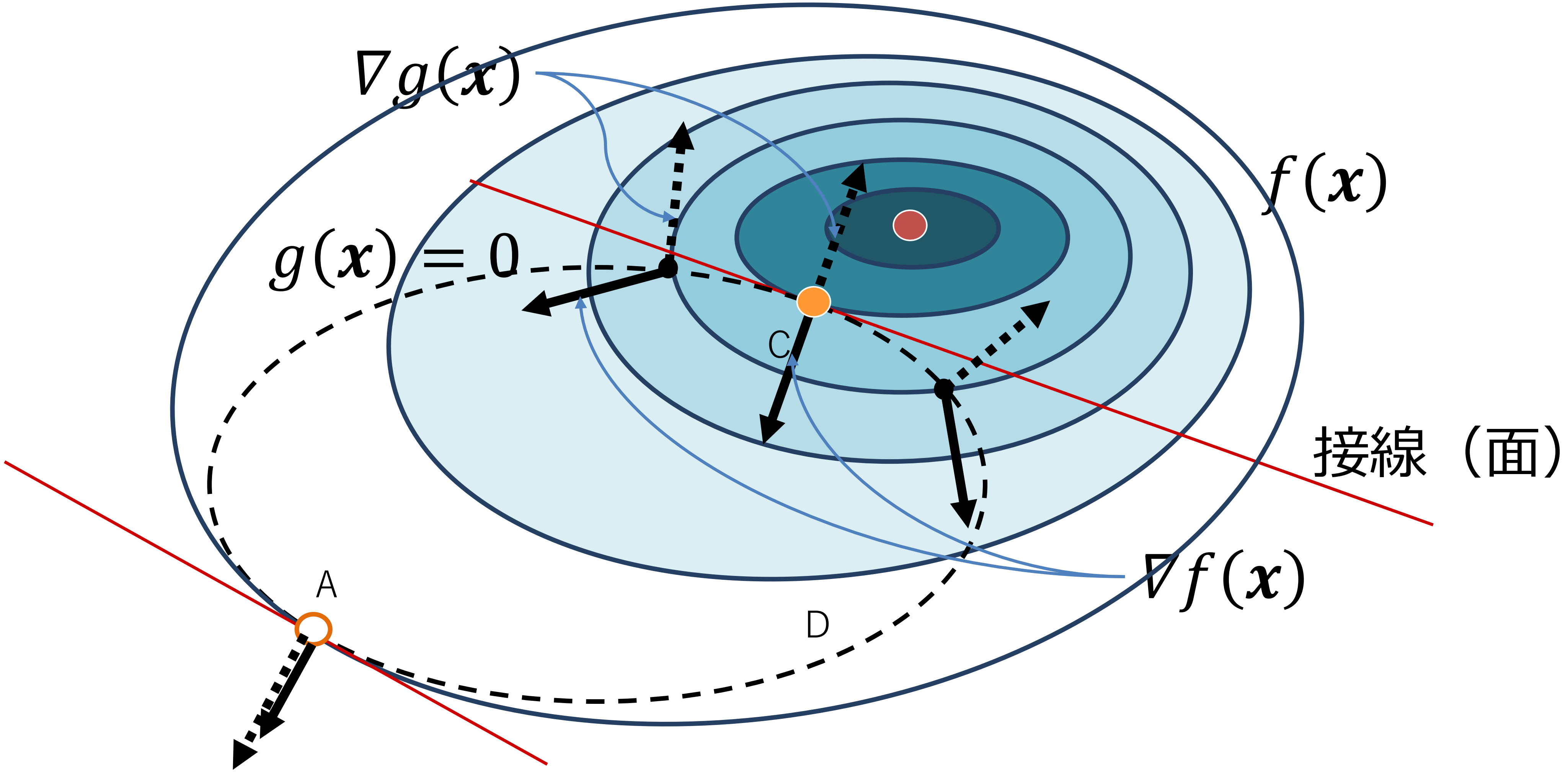
Lagrangeの未定乗数法

$$\Rightarrow \quad \nabla f(\boldsymbol{x}) + \lambda \nabla g(\boldsymbol{x}) = 0, \qquad g(\boldsymbol{x}) = 0$$

$$\nabla L(\boldsymbol{x},\lambda) = 0 \quad \rightarrow \quad \nabla f(\boldsymbol{x}) + \lambda \nabla g(\boldsymbol{x}) = 0$$

$$\frac{\partial}{\partial \lambda} L(\boldsymbol{x},\lambda) = 0 \quad \rightarrow \quad g(\boldsymbol{x}) = 0$$

等式制約の最適化問題



Note: $\nabla f(\boldsymbol{x}) + \lambda \nabla g(\boldsymbol{x}) = 0$, は必要条件にすぎない
λの正負も無関係

例題 1

問題：　周囲の長さが X の長方形の中で，面積の一番大きなものの辺の長さを求めよ。

解答：

各辺を a, b とすれば,

問題は,

$$\max ab$$
$$\text{s.t. } 2(a+b) = X$$

となる。

$$L(a,b,\lambda) = ab - \lambda(2a + 2b - X)$$

$$\frac{\partial}{\partial a} L(a,b,\lambda) = b - 2\lambda = 0$$

$$\frac{\partial}{\partial b} L(a,b,\lambda) = a - 2\lambda = 0$$

$$\frac{\partial}{\partial \lambda} L(a,b,\lambda) = -2a - 2b + X = 0$$

$$a = b = 2\lambda$$

$$-4\lambda - 4\lambda + X = 0, \quad \lambda = X/8$$

$$a = b = 2\lambda = X/4$$

例題2：一般化逆行列

- 不定方程式：

$$A\boldsymbol{x} = \boldsymbol{b} \qquad A: M \times N; M < N \tag{1}$$

- 解の中で，ノルム最小のものを求める

$$\min \boldsymbol{x}^T \boldsymbol{x}$$
$$\text{s.t. } A\boldsymbol{x} = \boldsymbol{b}$$

解答：

*M*本の方程式で拘束されているので，

それぞれに対し，ラグランジュ乗数を置く。

$$\boldsymbol{\lambda} = (\lambda_1 \lambda_2 \cdots \lambda_M)^{\boldsymbol{T}}$$

ラグランジュ関数は,

$$E = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} - \boldsymbol{\lambda}^T(A\boldsymbol{x} - \boldsymbol{b})$$

$\boldsymbol{x}$で偏微分して零とおくと,

$$\frac{\partial}{\partial \boldsymbol{x}}E = \frac{\partial}{\partial \boldsymbol{x}}\left(\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} - \boldsymbol{\lambda}^T(A\boldsymbol{x} - \boldsymbol{b})\right)$$

$$= \boldsymbol{x} - A^T\boldsymbol{\lambda} = 0 \qquad (2)$$

両辺にAを掛けて, (1)を用いて $\boldsymbol{x}$ を消去する。

$$A\boldsymbol{x} - AA^T\boldsymbol{\lambda} = \boldsymbol{b} - AA^T\boldsymbol{\lambda} = 0$$

$$\boldsymbol{\lambda} = (AA^T)^{-1}\boldsymbol{b}$$

(2)に代入する。

$$\boldsymbol{x} = A^T\boldsymbol{\lambda} = A^T(AA^T)^{-1}\boldsymbol{b} \qquad (3)$$

例題 3

条件,

$$g(\boldsymbol{x}) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T \boldsymbol{x} + 2 = 0 \qquad (1)$$

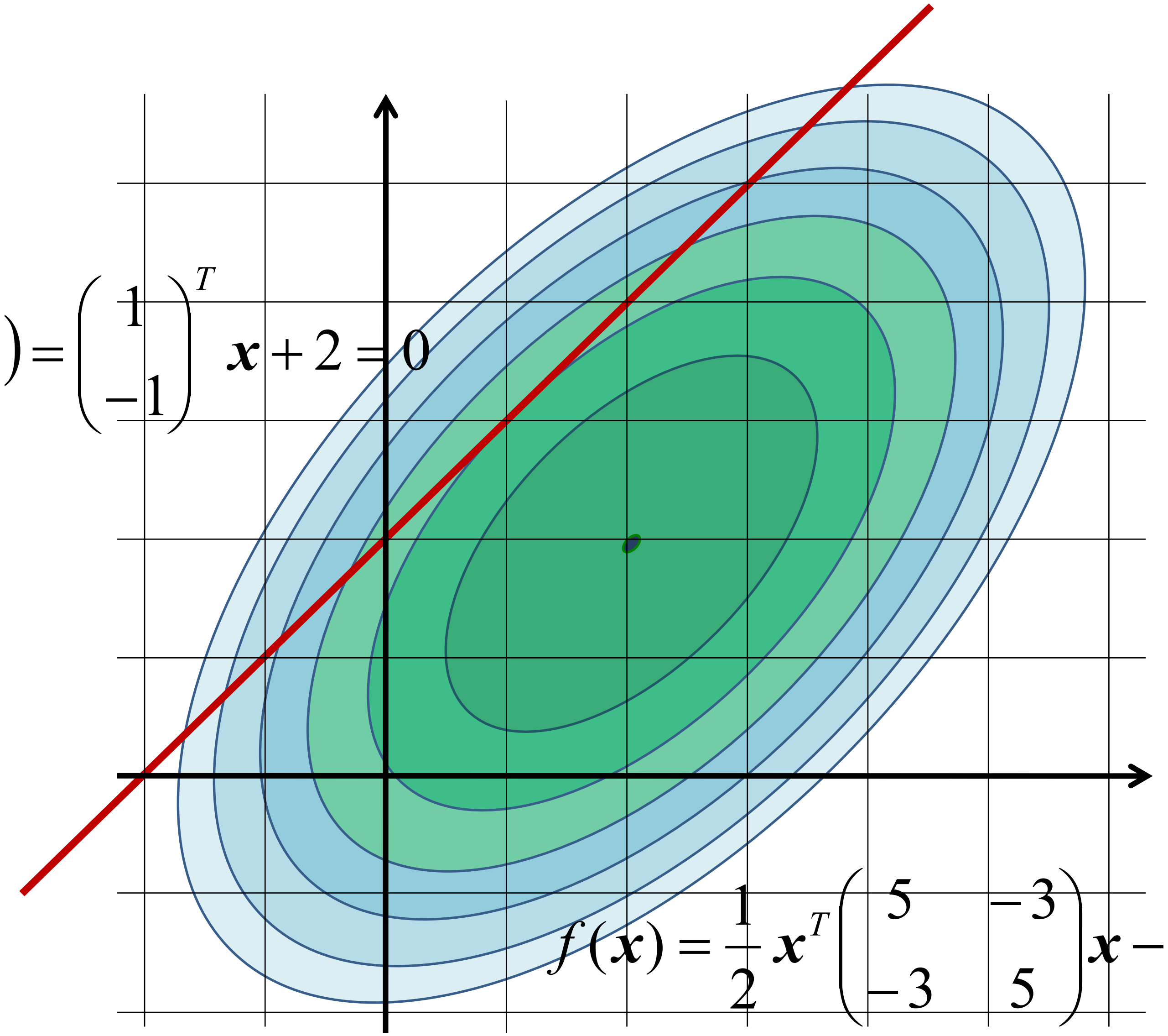
のもとで

関数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix} \boldsymbol{x} - 4\begin{pmatrix} 1 \\ 1 \end{pmatrix}^T \boldsymbol{x} \qquad (2)$$

の　最小値を与えるxを求めよ。

$$g(\boldsymbol{x}) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T \boldsymbol{x} + 2 = 0$$



$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix} \boldsymbol{x} - 4 \begin{pmatrix} 1 \\ 1 \end{pmatrix}^T \boldsymbol{x} = c$$

$$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^T\begin{pmatrix}5 & -3\\ -3 & 5\end{pmatrix}\boldsymbol{x}-4\begin{pmatrix}1 & 1\end{pmatrix}\boldsymbol{x} \qquad (1)$$

$$g(x)=\begin{pmatrix}1 & -1\end{pmatrix}\boldsymbol{x}+2=0 \qquad (2)$$

評価関数を次のように置く

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

$$=\frac{1}{2}\boldsymbol{x}^T\begin{pmatrix}5 & -3\\ -3 & 5\end{pmatrix}\boldsymbol{x}-4\begin{pmatrix}1\\ 1\end{pmatrix}^T\boldsymbol{x}+\lambda\left(\begin{pmatrix}1\\ -1\end{pmatrix}^T\boldsymbol{x}+2\right) \qquad (3)$$

xとλで偏微分して0と置く

$$\frac{\partial}{\partial \boldsymbol{x}}L(\boldsymbol{x},\lambda)=\begin{pmatrix}5 & -3\\ -3 & 5\end{pmatrix}\boldsymbol{x}-4\begin{pmatrix}1\\ 1\end{pmatrix}+\lambda\begin{pmatrix}1\\ -1\end{pmatrix}=0 \qquad (4)$$

$$\frac{\partial}{\partial \lambda}L(\boldsymbol{x},\lambda)=\begin{pmatrix}1\\ -1\end{pmatrix}^T\boldsymbol{x}+2=0 \qquad (5)$$

(4)より，

$$\boldsymbol{x}=\begin{pmatrix}5 & -3\\ -3 & 5\end{pmatrix}^{-1}\begin{pmatrix}4-\lambda\\ 4+\lambda\end{pmatrix}=\frac{1}{16}\begin{pmatrix}5 & 3\\ 3 & 5\end{pmatrix}\begin{pmatrix}4-\lambda\\ 4+\lambda\end{pmatrix}=\frac{1}{8}\begin{pmatrix}16-\lambda\\ 16+\lambda\end{pmatrix} \quad (6)$$

これを(5)に代入する。

$$\frac{1}{8}\begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}16-\lambda\\ 16+\lambda\end{pmatrix}+2=-\frac{1}{4}\lambda+2=0$$

$$\lambda=8 \quad (7)$$

これを(6)に代入して，$\boldsymbol{x}$を求め，さらに$f(\boldsymbol{x})$を求める。

$$\boldsymbol{x}=\frac{1}{8}\begin{pmatrix}16-\lambda\\ 16+\lambda\end{pmatrix}=\frac{1}{8}\begin{pmatrix}8\\ 24\end{pmatrix}=\begin{pmatrix}1\\ 3\end{pmatrix} \quad (8)$$

$$f\begin{pmatrix}1\\ 3\end{pmatrix}=\frac{1}{2}\begin{pmatrix}1 & 3\end{pmatrix}\begin{pmatrix}5 & -3\\ -3 & 5\end{pmatrix}\begin{pmatrix}1\\ 3\end{pmatrix}-4\begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}1\\ 3\end{pmatrix}=0 \quad (9)$$

例題 4

n個の事象の表すエントロピーは，

$$H = -\sum_{i=1}^{n} p_i \log p_i\,, \quad \sum_{i=1}^{n} p_i = 1$$

で定義される。

このとき， Hの最大値を与えるp_iを求めよ。

$$H = -\sum_{i=1}^{n} p_i \log p_i \,, \qquad \sum_{i=1}^{n} p_i = 1$$

Lagrange の未定乗数λを用いて、

$$L = -\sum_{i=1}^{n} p_i \log p_i - \lambda \left(\sum_{i=1}^{n} p_i - 1 \right)$$

Lをp_iで微分して0とすると、

$$\frac{\partial L}{\partial p_i} = -\log p_i - \log e - \lambda = 0$$

$$\log p_i = -\log e - \lambda = -\log e 2^{\lambda}$$

これより、

$$p_i = e^{-1}2^{-\lambda}$$

p_iはiに無関係で決まり，$\sum_{i=1}^{n} p_i = 1$であるから、

$$p_i = \frac{1}{n}, \qquad (i = 1,2,\cdots,n)$$

となり、これらの p_i が H の最大値を与える。

このとき H は、

$$H = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = \log n$$

となる。

・すべてが同じ確率で生じる。

⇒　どれが起こるか分からない。⇒　最も不確定。

§ まとめ

- $g(\boldsymbol{x})=0$ の制約のもとで，$f(\boldsymbol{x})$ の最大値（最小値）を求める際には,

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

とおいて，これを $\boldsymbol{x},\lambda$ で最大化（最小化）する。（Lagrangeの未定乗数法)

§ 演習問題

1.

条件

$$g(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} - 4\begin{pmatrix}0 & 1\end{pmatrix}\boldsymbol{x} + 7 = 0 \qquad (1)$$

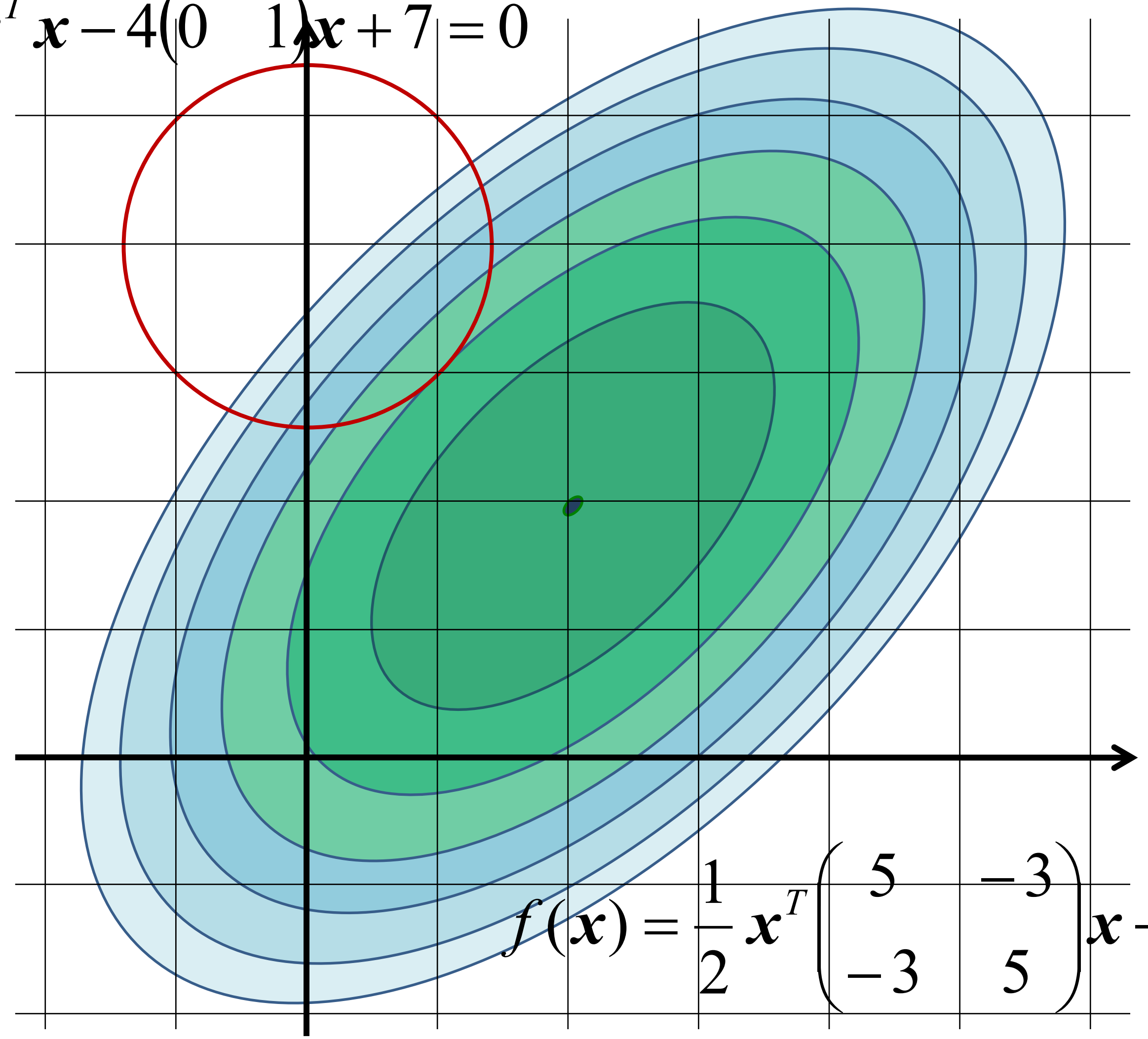
の下で，

関数，

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\begin{pmatrix}5 & -3 \\ -3 & 5\end{pmatrix}\boldsymbol{x} - 4\begin{pmatrix}1 \\ 1\end{pmatrix}^T\boldsymbol{x} \qquad (2)$$

の最小値を与える*x*を求めよ。

$$g(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} - 4\begin{pmatrix}0 & 1\end{pmatrix}\boldsymbol{x} + 7 = 0$$



$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\begin{pmatrix}5 & -3\\ -3 & 5\end{pmatrix}\boldsymbol{x} - 4\begin{pmatrix}1\\ 1\end{pmatrix}^T\boldsymbol{x} = c$$

演習問題

2.

条件，

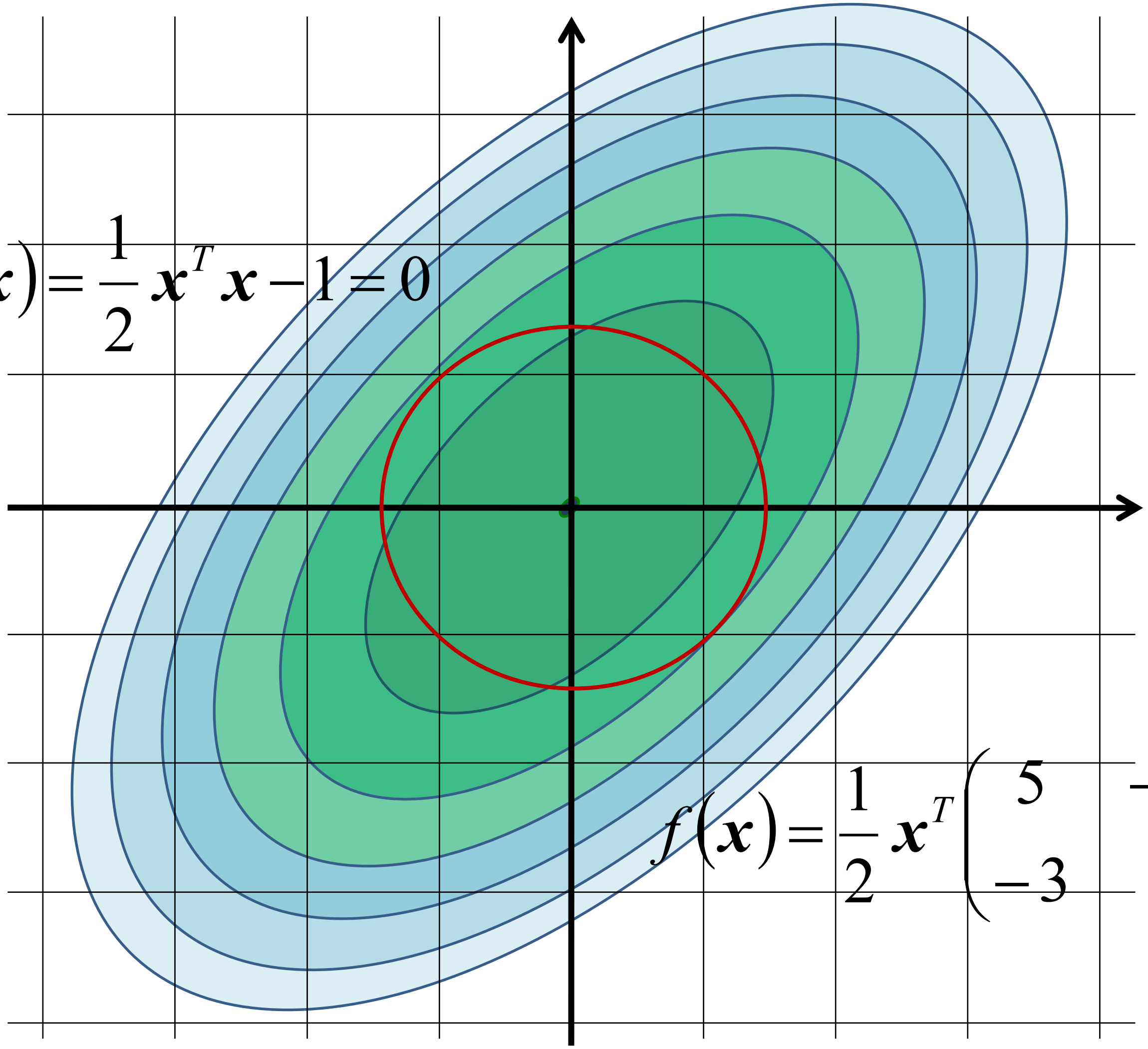
$$g(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} - 1 = 0 \qquad (1)$$

のもとで，関数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}\boldsymbol{x} \qquad (2)$$

の最小値を与える*x*を求めよ。

$$g(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} - 1 = 0$$



$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}\boldsymbol{x} = c$$

演習問題

3. 4つのデータ,

$$\boldsymbol{x}_1 = \begin{pmatrix} 5 \\ 0 \end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix} -3 \\ -4 \end{pmatrix}, \boldsymbol{x}_3 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \boldsymbol{x}_4 = \begin{pmatrix} -5 \\ 0 \end{pmatrix}$$

が与えられている。

1) このデータの共分散行列 C を求めよ。

2) ベクトル $\boldsymbol{w} = (w_1\ w_2)^T$を用いて，$z_i = \boldsymbol{x}_i{}^T\boldsymbol{w}$ と座標変換するとき，$\|\boldsymbol{w}\| = 1$ の条件の下で，z_iの分散を最も大きくするように, $\boldsymbol{w}$ を定めることにするとき，Cと $\boldsymbol{w}$ の関係を導け。

3) 2)の条件を満たす $\boldsymbol{w}$ を実際に求めよ。

4) データの散布図を描け。この散布図に重ねて$\boldsymbol{w}$ を描け。

演習問題

4. 次のような5つのデータが与えられたとする。

(-3,-2), (1,-1), (0,0), (3,2), (-1,1)

1) 平均値ベクトル，および共分散行列　C　を求めよ。

2)　Cの最大固有値と，その固有値に対する固有ベクトルを求めよ。

3)　データの散布図を描け。また，この散布図に重ねて2.2で求めた固有ベクトルを描け。

演習問題

5．クラスαのデータとして，4つのデータ，

$$\boldsymbol{x}_1^\alpha = \begin{pmatrix} -5 \\ 0 \end{pmatrix}, \boldsymbol{x}_2^\alpha = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \boldsymbol{x}_3^\alpha = \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \boldsymbol{x}_4^\alpha = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$$

が与えられ，クラスβのデータとして，4つのデータ，

$$\boldsymbol{x}_1^\beta = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \boldsymbol{x}_2^\beta = \begin{pmatrix} 2 \\ -3 \end{pmatrix}, \boldsymbol{x}_3^\beta = \begin{pmatrix} 5 \\ 0 \end{pmatrix}, \boldsymbol{x}_4^\beta = \begin{pmatrix} -1 \\ -4 \end{pmatrix}$$

が与えられている。

1)クラス毎に，平均ベクトル$\boldsymbol{\mu}_\alpha$，$\boldsymbol{\mu}_\beta$，共分散行列C_α, C_βを求めよ。

2) 級内分散行列 C_W を，２つのクラスの共分散行列の平均として定義するとき, C_W を求めよ。

3) 級間分散行列 C_B を，それぞれのクラスの平均ベクトルの共分散行列として定義するとき， C_B を求めよ。

4) ベクトル $\boldsymbol{w} = (w_1\ w_2)^T$を用いて，$z_i = \boldsymbol{x}_i{}^T\boldsymbol{w}$ と座標変換するとき，変換後の座標空間における級間分散 $\sigma_B{}^2$ と級内分散 $\sigma_W{}^2$ を，C_B，C_W，$\boldsymbol{w}$ の式で表せ。

5) 級間分散 $\sigma_B{}^2$ と級内分散 $\sigma_W{}^2$ の比 $\sigma_B{}^2/\sigma_W{}^2$ を最大化するように $\boldsymbol{w}$ を定めるとき，変換後の座標空間において，同じクラスのデータが纏まり，違うクラスのデータが離れる分布を作るため，座標空間は識別に適したものとなる。このような空間を作る，C_B，C_W，$\boldsymbol{w}$ の関係を導け。

6) 5)の条件を満たす$\boldsymbol{w}$を実際に求めよ。

7) データの散布図を描け。この散布図に重ねて$\boldsymbol{w}$を描け。