最適化理論

凸関数とその性質

凸計画問題の標準形

$f_0, f_1, \ldots, f_m$ は凸関数。

$x \in \mathbb{R}^n, a_i \in \mathbb{R}^n, b_i \in \mathbb{R}$.

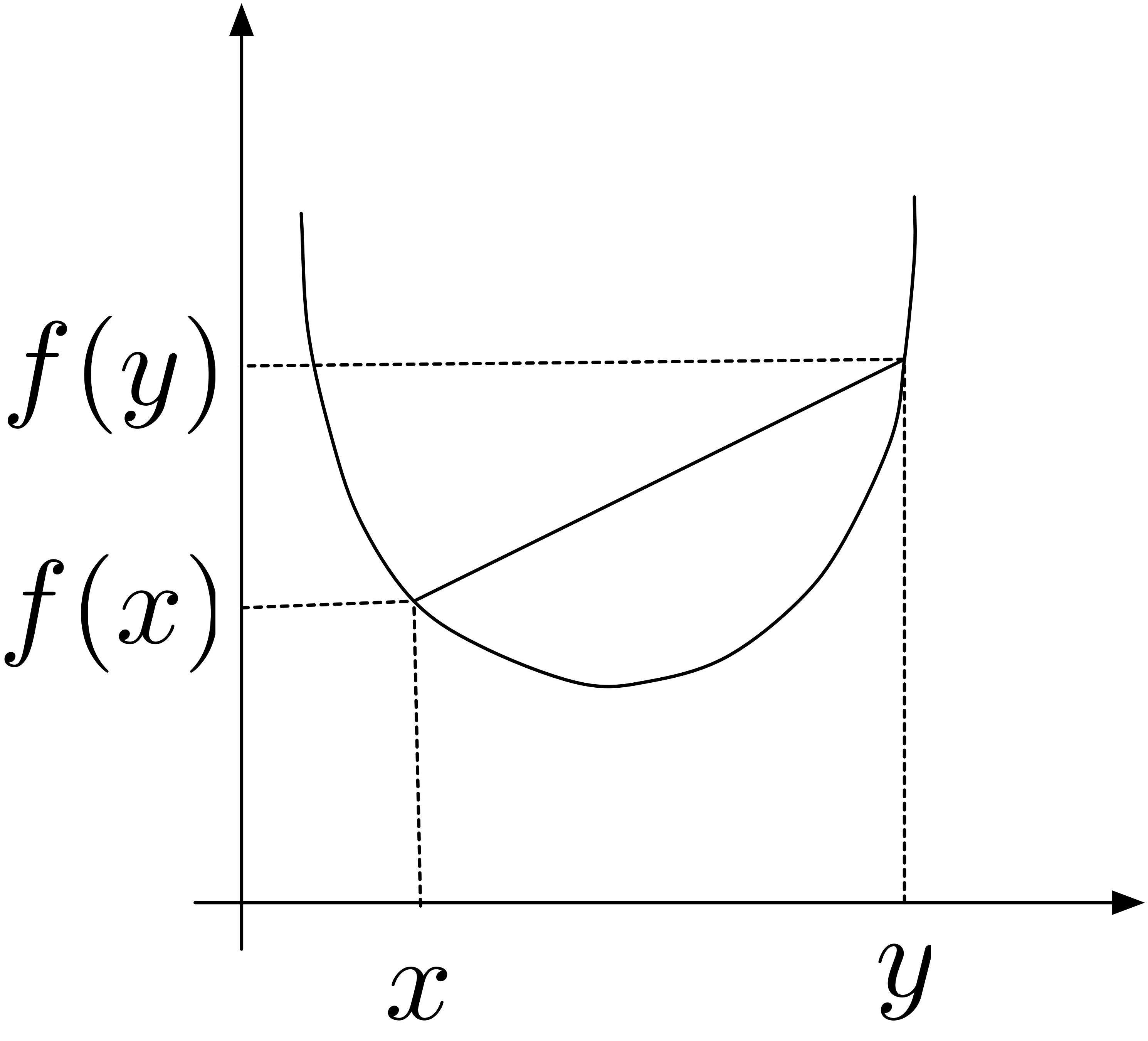
$$
\begin{aligned}
\text{minimize } & f_0(x) \\
\text{subject to } & f_i(x) \leq 0, \quad i = 1, 2, \ldots, m \\
& a_i^T x = b_i, \quad i = 1, 2, \ldots, p
\end{aligned}
$$

凸関数(convex function)

$f : \mathbb{R}^n \to \mathbb{R}$ が任意の $x, y \in \mathrm{dom} f$ について

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta) f(y)$$

を満たすとき、f を凸関数(convex function)と呼ぶ。ただし、θ は $0 \leq \theta \leq 1$ を満たす実数。



f(y)
f(x)
x
y

凸関数：補足

- f がconvex(＝凸) の場合、$-f$ はconcaveという。

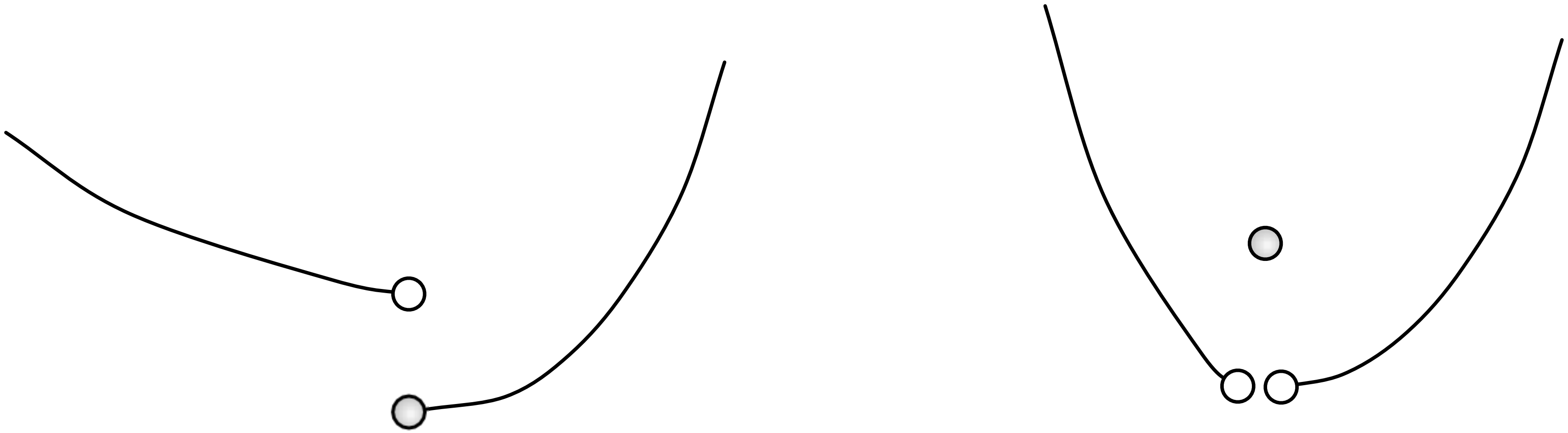
- 不等式

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$$

が成り立つとき、f は真に凸(strictly convex)である、という

凸関数の連続性

- 凸関数はかならず連続関数になる(自明なことではない)



非連続関数（非凸関数）

凸関数の例 (1次元の場合)

- アフィン関数: $ax + b (a, b \in \mathbb{R})$
- 指数関数: $e^{ax} (a \in \mathbb{R})$
- ベキ関数: $x^{\alpha} (x \in \mathbb{R}_{++}, \alpha \geq 1$ or $\alpha \leq 0)$
- 負対数関数: $-\ln(x) (x \in \mathbb{R}_{++})$

凸関数の例 (多次元の場合)

- アフィン関数:
  $ax^T + b (x \in \mathbb{R}^n, a \in \mathbb{R}^n, b \in \mathbb{R})$
- ノルム関数: $||x|| (x \in \mathbb{R}^n)$

ノルム関数

以下では、$x \in \mathbb{R}^n$ を仮定する。

ℓ_1-ノルム

$$||x||_1 = |x_1| + |x_2| + \cdots |x_n|$$

ℓ_2-ノルム

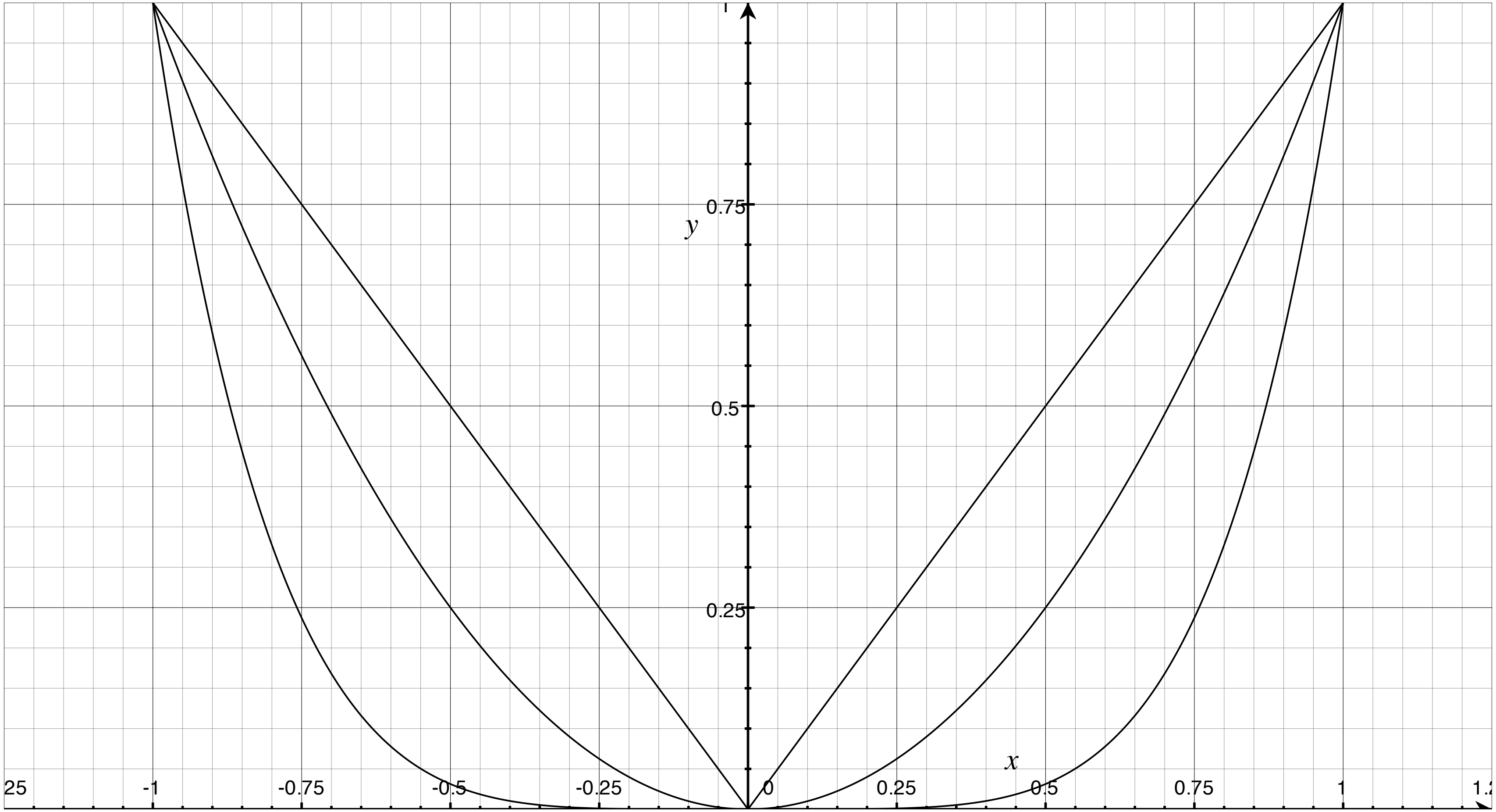
$$||x||_2 = \sqrt{|x_1|^2 + |x_2|^2 + \cdots |x_n|^2}$$

ℓ_∞-ノルム

$$||x||_\infty = \max\{|x_1|, |x_2|, \cdots |x_n|\}$$

ℓ_p-ノルム

$$||x||_p \equiv \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \quad p \geq 1$$



ノルム関数のプロット$(p = 1, 2, 5)$

凸関数の組み合わせ

- $\alpha, \beta \geq 0$ とし、$f(x), g(x)$ を凸関数 $(x \in \mathbb{R}^n)$ とすると
$$\alpha f(x) + \beta g(x)$$
  も凸関数となる(凸関数の非負係数和)。

- $f(x)$ が凸関数の場合、$f(Ax + b)$ も凸関数となる(アフィン関数との合成関数)。

その他の凸性を保つ操作

- $f(x, y)$ が凸関数で、C が凸集合の場合、

$$g(x) = \inf_{y \in C} f(x, y)$$

  も凸関数となる。

この性質からラグランジュ双対関数(後述)の concave 性が導かれる。

共役関数(ルジャンドル変換)

$f : \mathbb{R}^n \to \mathbb{R}$ について

$$f^*(y) = \sup_{x \in \mathrm{dom} f} (y^T x - f(x))$$

と定義される関数 f^* を f の共役(conjugate)と呼ぶ。共役関数は必ず凸関数となる。

- f が凸関数ならば、f の共役関数の共役はふたたび f に戻る。

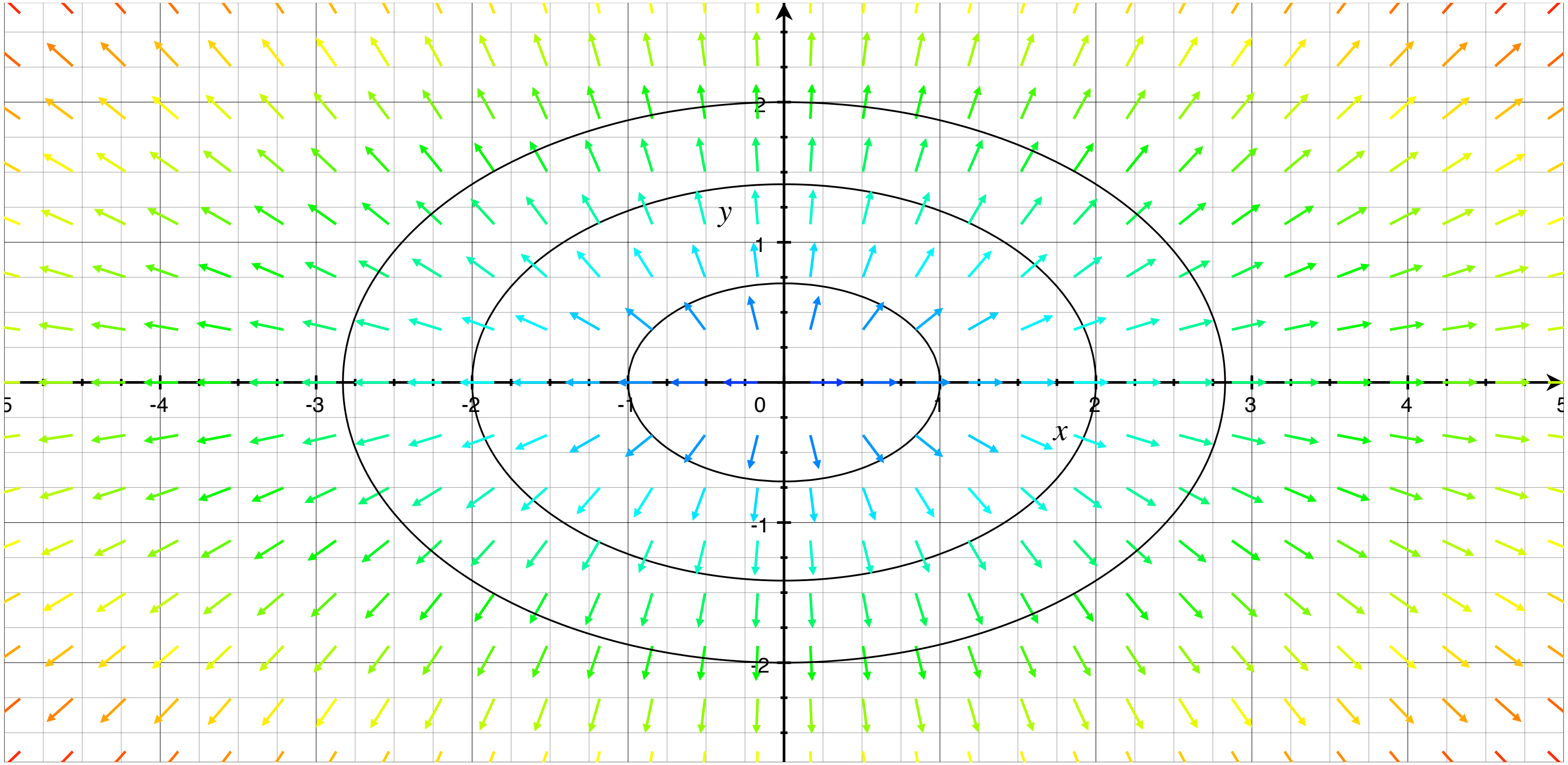
勾配ベクトル

f の勾配ベクトル(gradient):

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)^T$$

例：$f(x_1, x_2) = x^2 + 2y^2$ の場合、

$$\nabla f(x) = \begin{pmatrix} 2x \\ 4y \end{pmatrix}$$



関数$f(x_1, x_2) = x^2 + 2y^2$の勾配ベクトル場楕円は等高線$x^2 + 2y^2 = q(q = 1, 4, 8)$を表す。

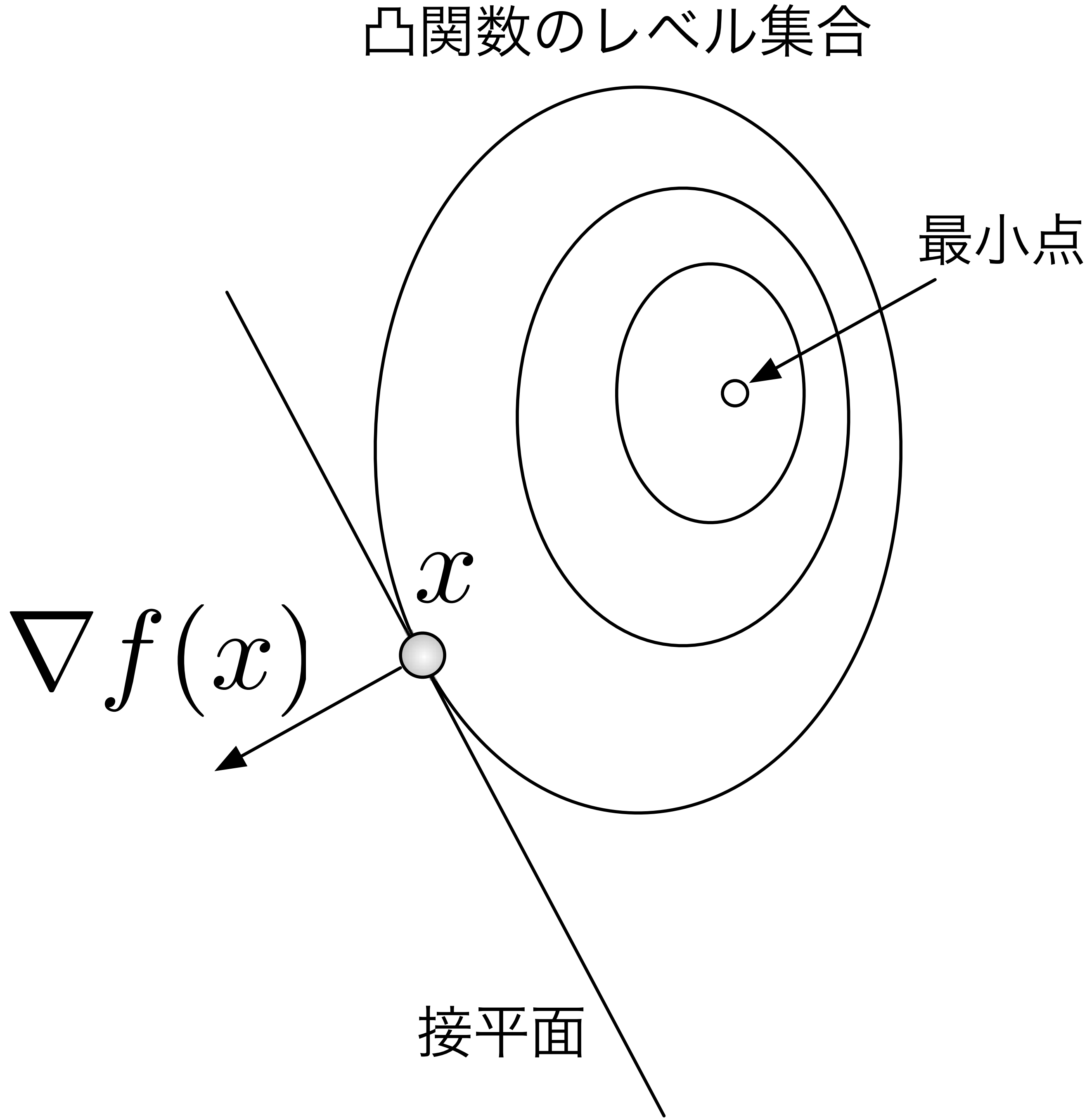
レベル集合と勾配ベクトル

$$S_\alpha = \{x \in \mathrm{dom} f : f(x) = \alpha\}$$

を f のレベル集合と呼ぶ。

- 勾配ベクトルは、レベル集合の接平面の法線ベクトルとなっている。

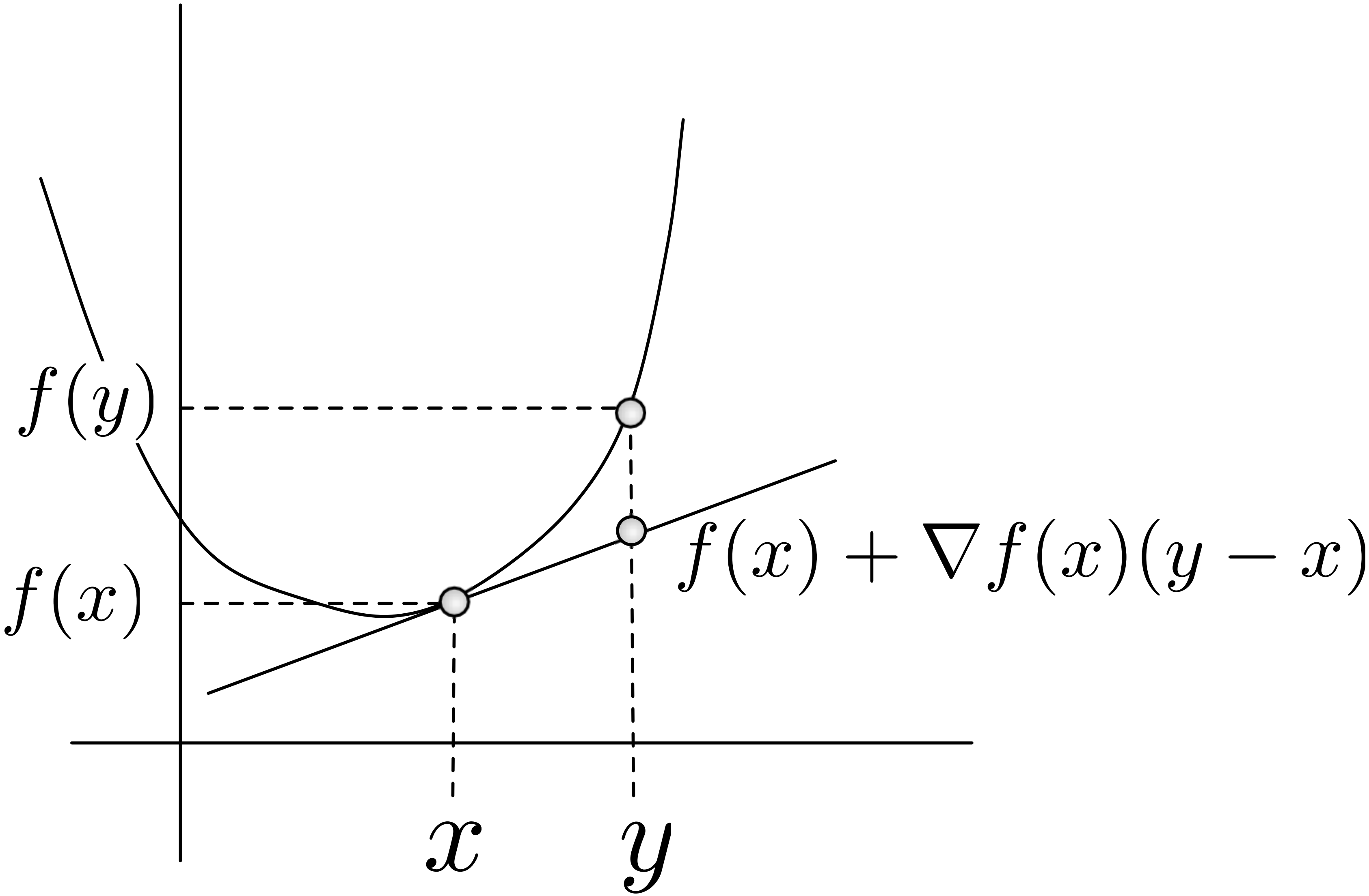
例：天気図の場合、気圧の等高線がレベル集合となる。


凸関数のレベル集合
最小点
x
$\nabla f(x)$
接平面

凸関数の1次条件

f が凸関数であることと次の関係は同値である。

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \quad \forall x, y \in \mathrm{dom}\, f$$


$f(y)$
$f(x)$
$f(x) + \nabla f(x)(y - x)$
x
y

2次の凸関数条件

$f : \mathbb{R}^n \to \mathbb{R}$, $\mathrm{dom}\, f$ が凸集合とする。

$$\nabla^2 f(x) \succeq 0, \quad x \in \mathrm{dom}\, f$$

と f が凸関数であることは同値である。
行列 A が $A \succeq 0$ であるとは、A が半正定値行列 (semi definite matrix) であることを意味する。

ヘッセ行列 (Hessian matrix)

$$\nabla^2 f(x) \equiv \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \vdots & \vdots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} f(x) \end{pmatrix}$$

クイズ

$$f(x) \equiv (x_1 - 6)^2 + 2x_2^2$$

について、ヘッセ行列を求めよ。

半正定値行列(semidefinite matrix)

$n \times n$ 行列 $A \in \mathbb{R}^{n\times n}$ について

$$\forall x \in \mathbb{R}^n, \quad x^T A x \geq 0$$

が成立するとき、$A \succeq 0$ と書き、A を半正定値行列と呼ぶ。半正定値行列の固有値は、すべて非負である。

2次関数の凸性

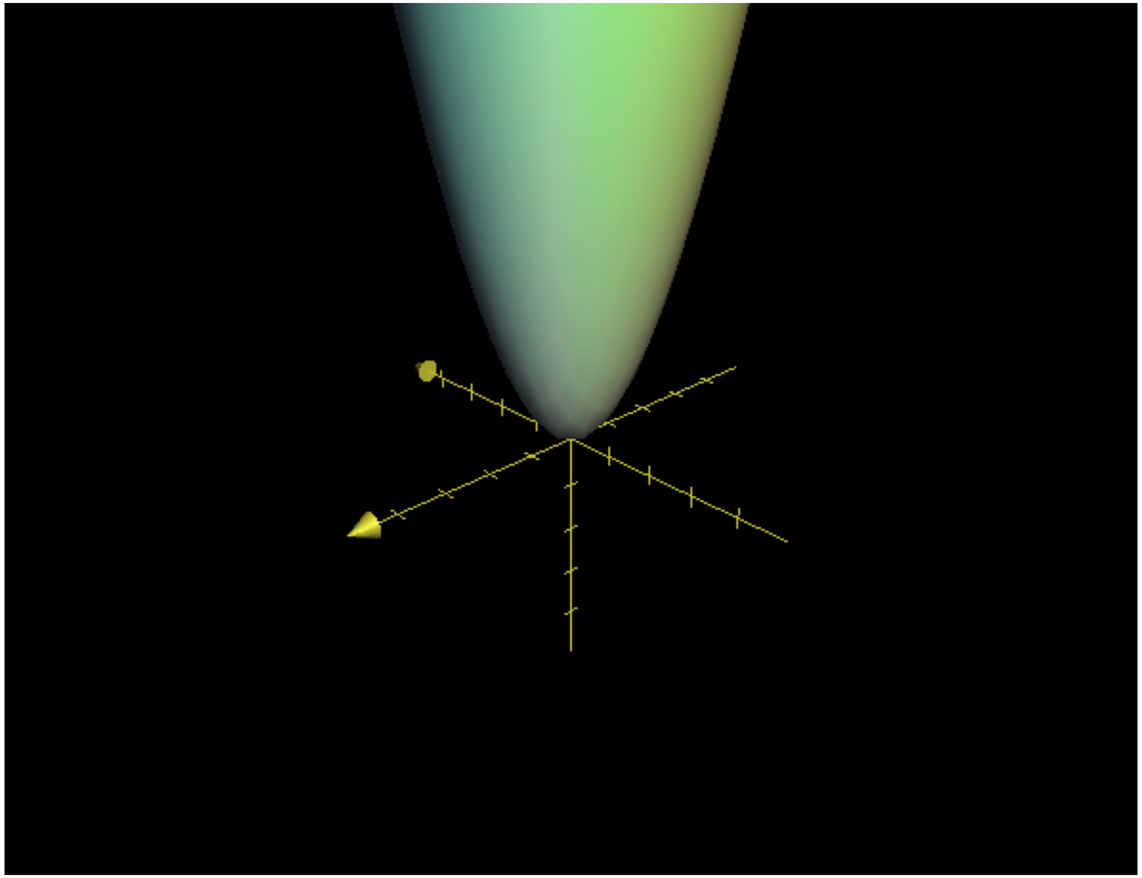
２次関数

$$f(x) = \frac{1}{2}x^T P x + q^T x + r$$

$(P \in \mathbb{S}^n, q, r \in \mathbb{R}^n)$ が凸関数であるための必要十分条件は

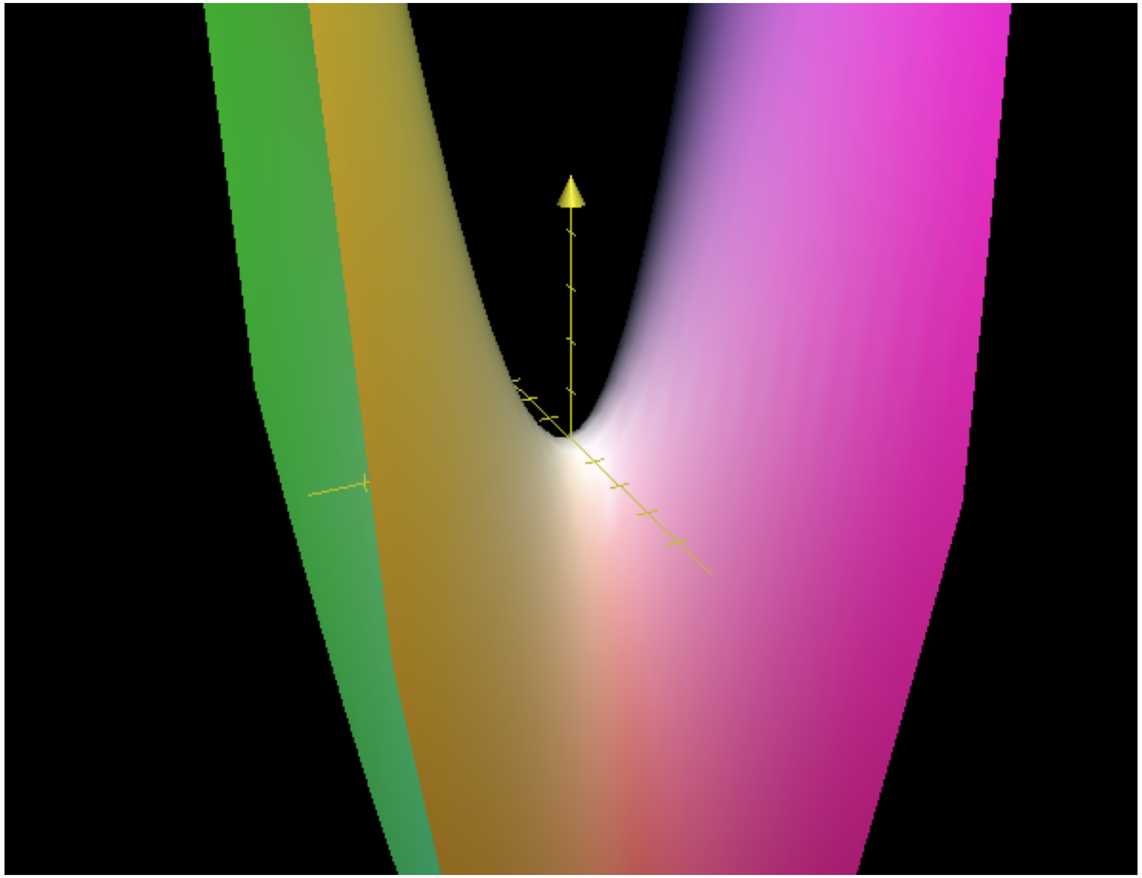
$$P \succeq 0$$

である。ここで、$\mathbb{S}^n$ は $n \times n$ 対称実行列の集合である。



2変数2次関数 $f(x_1, x_2) = x_1^2 + x_2^2$ (凸関数)

$$x_1^2 + x_2^2 = (x_1 \ x_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$



2変数2次関数 $f(x_1, x_2) = x_1^2 - x_2^2$ (凸ではない)

$$x_1^2 - x_2^2 = (x_1 \;\; x_2) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$