17. 最適化

- 制約なし最適化問題
  - $f(\boldsymbol{x})$ を最大（または最小）にする $\boldsymbol{x}$ を見つける

$f(\boldsymbol{x})$ の $\boldsymbol{x}_0$ の周りのテーラー展開，ただし $\boldsymbol{x} = \boldsymbol{x}_0 + \Delta\boldsymbol{x},\ \Delta\boldsymbol{x} = \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix}$

$$
\begin{aligned}
f(\boldsymbol{x}) = f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) = f(\boldsymbol{x}_0) &+ \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x_i}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} \cdot \Delta x_i \\
&+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} \cdot \Delta x_i \cdot \Delta x_j + \dots
\end{aligned}
$$

$$
= f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T \Delta\boldsymbol{x} + \frac{1}{2} \Delta\boldsymbol{x}^T H(\boldsymbol{x}_0) \Delta\boldsymbol{x} + \dots
$$

17. 最適化

$f(\boldsymbol{x})$ が$\boldsymbol{x}_0$ で極大値を取る条件

$\forall\Delta\boldsymbol{x}$ に対して $\underline{f(\boldsymbol{x}_0) \geq f(\boldsymbol{x}_0+\Delta\boldsymbol{x})}$，ただし $\|\Delta\boldsymbol{x}\|$ は十分に小さい

$$\Updownarrow$$

$$f(\boldsymbol{x}_0) \geq f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T \Delta\boldsymbol{x} + \frac{1}{2}\Delta\boldsymbol{x}^T H(\boldsymbol{x}_0)\Delta\boldsymbol{x} + \dots$$

$$0 \geq \nabla f(\boldsymbol{x}_0)^T \Delta\boldsymbol{x} + \frac{1}{2}\Delta\boldsymbol{x}^T H(\boldsymbol{x}_0)\Delta\boldsymbol{x} + \dots$$

次の条件は，$f(\boldsymbol{x})$ が$\boldsymbol{x}_0$において極大値を取るための十分条件である

$$\begin{cases} \nabla f(\boldsymbol{x}_0) = \boldsymbol{0}, \\ \forall\Delta\boldsymbol{x} \text{ に対して } \Delta\boldsymbol{x}^T H(\boldsymbol{x}_0)\Delta\boldsymbol{x} < 0 \text{，すなわち } H(\boldsymbol{x}_0) \text{ は負定値行列} \end{cases}$$

スカラー変数 x の場合は，対応する条件は以下の通り

$$\frac{d}{dx}f(x_0) = 0,\ \frac{d^2}{dx^2}f(x_0) < 0.$$

17. 最適化

前のスライドに書いた条件は，必要条件でもある

Δx が十分小さく，Δx の２次以上の高次の項が無視できるとき，
任意の小さい Δx に対して，次が成り立つことが必要である

$$f(\boldsymbol{x}_0) \geq f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T \Delta \boldsymbol{x}$$

これから $\nabla f(\boldsymbol{x}_0) = \boldsymbol{0}$ が成り立つことが必要である

さらに，３次以上の高次の項が無視できるとき，

$$\begin{aligned} f(\boldsymbol{x}_0) &\geq f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T \Delta \boldsymbol{x} + \frac{1}{2} \Delta \boldsymbol{x}^T H(\boldsymbol{x}_0) \Delta \boldsymbol{x} \\ &= f(\boldsymbol{x}_0) + \frac{1}{2} \Delta \boldsymbol{x}^T H(\boldsymbol{x}_0) \Delta \boldsymbol{x} \end{aligned}$$

$$0 \geq \Delta \boldsymbol{x}^T H(\boldsymbol{x}_0) \Delta \boldsymbol{x}$$

17. 最適化

- 等式制約つき最適化
  - 関数 $f(\boldsymbol{x})$ を最大（または最小）にする $\boldsymbol{x}$ を, 以下の条件下で求める

$$\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} h_1(\boldsymbol{x}) \\ \vdots \\ h_m(\boldsymbol{x}) \end{bmatrix} = \boldsymbol{0},$$

ここで $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ であり, $m < n$ である

17. 最適化

– $\boldsymbol{x}_0$ が解であるとし, 解 $\boldsymbol{x}_0$ が満足すべき条件を見つけよう

$\boldsymbol{x}_0$ は，もちろん，次式を満たす必要がある

$$\boldsymbol{h}(\boldsymbol{x}_0) = \boldsymbol{0}$$

さらに，$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) = \boldsymbol{0}$ を満たす十分小さい $\Delta\boldsymbol{x}$ に対して

$$f(\boldsymbol{x}_0) \geq f(\boldsymbol{x}) = f(\boldsymbol{x}_0 + \Delta\boldsymbol{x})$$

を満たす必要がある

$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x}_0 + \Delta\boldsymbol{x})$ は次のように近似することができる

$$\boldsymbol{h}(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) \simeq \boldsymbol{h}(\boldsymbol{x}_0) + \boldsymbol{\nabla h}(\boldsymbol{x}_0)^T \Delta\boldsymbol{x}$$

ここで $\boldsymbol{\nabla h}(\boldsymbol{x}_0) = \begin{bmatrix} \nabla h_1(\boldsymbol{x}_0) & \ldots & \nabla h_m(\boldsymbol{x}_0) \end{bmatrix}$ は$n \times m$ 行列である

$\boldsymbol{h}(\boldsymbol{x}_0) = \boldsymbol{0}$, $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$ であるので，$\boldsymbol{\nabla h}(\boldsymbol{x}_0)^T \Delta\boldsymbol{x} = \boldsymbol{0}$ ，すなわち

$$\Delta\boldsymbol{x} \in \mathcal{N}\left(\boldsymbol{\nabla h}(\boldsymbol{x}_0)^T\right)$$

がなりたつ

17. 最適化

$f(\boldsymbol{x}_0+\Delta\boldsymbol{x})$も

$$f(\boldsymbol{x}_0+\Delta\boldsymbol{x}) \simeq f(\boldsymbol{x}_0)+\nabla f(\boldsymbol{x}_0)^T\Delta\boldsymbol{x}$$

と近似でき，これから

$$f(\boldsymbol{x}_0) \geq f(\boldsymbol{x}_0+\Delta\boldsymbol{x}) \simeq f(\boldsymbol{x}_0)+\nabla f(\boldsymbol{x}_0)^T\Delta\boldsymbol{x}$$

の条件が導かれ，したがって，以下の条件が得られる

$$0 \geq \nabla f(\boldsymbol{x}_0)^T\Delta\boldsymbol{x} \text{ for } \forall\Delta\boldsymbol{x}\in\mathscr{N}(\boldsymbol{\nabla h}(\boldsymbol{x}_0)^T).$$

上の条件において，$0>\nabla f(\boldsymbol{x}_0)^T\Delta\boldsymbol{x}$ は除外される．理由は，もし $0>\nabla f(\boldsymbol{x}_0)^T\Delta\boldsymbol{x}$ が，ある $\Delta\boldsymbol{x}\in\mathscr{N}(\boldsymbol{\nabla h}(\boldsymbol{x}_0)^T)$ についてなりたつとすると，$-\Delta\boldsymbol{x}$ も$\mathscr{N}(\boldsymbol{\nabla h}(\boldsymbol{x}_0)^T)$に属する一方，$0<\nabla f(\boldsymbol{x}_0)^T(-\Delta\boldsymbol{x})$ となってしまうからである

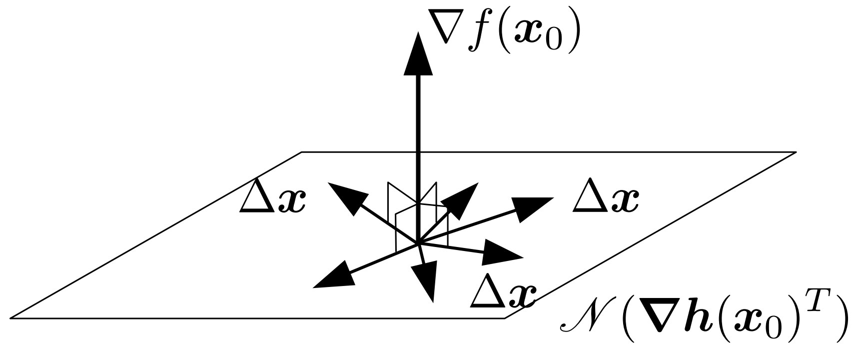
したがって，満たされるべき条件は次の通りとなる

$$\nabla f(\boldsymbol{x}_0)^T\Delta\boldsymbol{x}=0 \text{ for } \forall\Delta\boldsymbol{x}\in\mathscr{N}(\boldsymbol{\nabla h}(\boldsymbol{x}_0)^T).$$

この条件は，$f(\boldsymbol{x}_0)\leq f(\boldsymbol{x}_0+\Delta\boldsymbol{x})$ から議論を始めても同様に得られる．すなわち，これは必要条件である

17. 最適化

$\forall \Delta \boldsymbol{x} \in \mathcal{N}(\boldsymbol{\nabla h}(\boldsymbol{x}_0)^T)$ に対して $\nabla f(\boldsymbol{x}_0)^T \Delta \boldsymbol{x} = 0$ がなりたつという条件は，ベクトル $\nabla f(\boldsymbol{x}_0)$ が $\mathcal{N}(\boldsymbol{\nabla h}(\boldsymbol{x}_0)^T)$ に属する任意のベクトルと直交している，すなわち，$\nabla f(\boldsymbol{x}_0)$ は $\mathcal{N}(\boldsymbol{\nabla h}(\boldsymbol{x}_0)^T)$ に直交していることを意味する



17. 最適化

一般に

ベクトル $\boldsymbol{y}$ が $\mathscr{N}(A^T)$ に直交している $\iff$ $\boldsymbol{y} \in \mathscr{R}(A)$

$A \in \mathbb{R}^{m \times n}$

まず, $\mathscr{N}(A^T) \perp \mathscr{R}(A)$ がなりたつ.

$\forall \boldsymbol{x} \in \mathscr{N}(A^T)$ に対して

$$A^T \boldsymbol{x} = \boldsymbol{0},$$

$$\begin{bmatrix} {\boldsymbol{a}_1}^T \\ \vdots \\ {\boldsymbol{a}_n}^T \end{bmatrix} \boldsymbol{x} = \boldsymbol{0},$$

$${\boldsymbol{a}_i}^T \boldsymbol{x} = 0, i = 1, \ldots, n.$$

$\forall \boldsymbol{y} \in \mathscr{R}(A)$ に対して

$$\exists \boldsymbol{z}, \boldsymbol{y} = A\boldsymbol{z}$$

$$= \begin{bmatrix} \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$

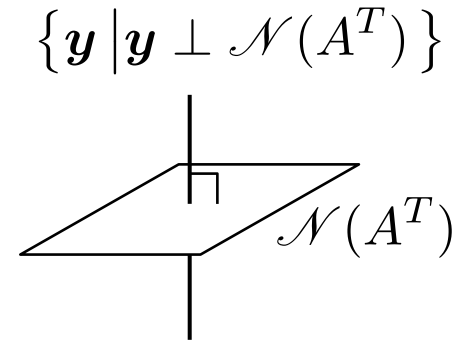
$$= \sum_{i=1}^{n} z_i \boldsymbol{a}_i$$

$$\boldsymbol{y}^T \boldsymbol{x} = \left(\sum_{i=1}^{n} z_i \boldsymbol{a}_i \right)^T \boldsymbol{x} = \sum_{i=1}^{n} z_i {\boldsymbol{a}_i}^T \boldsymbol{x} = 0$$

$\therefore \boldsymbol{x} \perp \boldsymbol{y}$

17. 最適化

次に $\dim\left\{\boldsymbol{y}\,\middle|\,\boldsymbol{y}\perp\mathscr{N}(A^T)\right\}$ を調べる



$$
\begin{aligned}
\dim\left\{\boldsymbol{y}\,\middle|\,\boldsymbol{y}\perp\mathscr{N}(A^T)\right\} &= m-\dim\mathscr{N}(A^T)\\
&= m-\bigl(m-\dim\mathscr{R}(A^T)\bigr)\\
&= \dim\mathscr{R}(A^T)\\
&= \mathrm{rank}A^T = \mathrm{rank}A\\
&= \dim\mathscr{R}(A)
\end{aligned}
$$

よって，$\mathscr{N}(A^T)\perp\mathscr{R}(A)$ がなりたち，また，$\mathscr{N}(A^T)$ に直交しているすべてのベクトルから構成される線形空間の次元は$\mathscr{R}(A)$ の次元に等しい．したがって，$\left\{\boldsymbol{y}\,\middle|\,\boldsymbol{y}\perp\mathscr{N}(A^T)\right\}=\mathscr{R}(A)$ がなりたつ．

以上から $\nabla f(\boldsymbol{x}_0)\in\mathscr{R}(\boldsymbol{\nabla h}(\boldsymbol{x}_0))$.

17. 最適化

条件 $\nabla f(\boldsymbol{x}_0) \in \mathscr{R}(\boldsymbol{\nabla h}(\boldsymbol{x}_0))$ は，$\nabla f(\boldsymbol{x}_0) = \boldsymbol{\nabla h}(\boldsymbol{x}_0)(-\boldsymbol{v})$ がなりたつようなベクトル $-\boldsymbol{v}$ が存在することを言っている

したがって，以下が得られる

$$\nabla f(\boldsymbol{x}_0) + \boldsymbol{\nabla h}(\boldsymbol{x}_0)\boldsymbol{v} = \boldsymbol{0},$$

$$\nabla f(\boldsymbol{x}_0) + \begin{bmatrix} \nabla h_1(\boldsymbol{x}_0) & \dots & \nabla h_m(\boldsymbol{x}_0) \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} = \boldsymbol{0},$$

$$\nabla f(\boldsymbol{x}_0) + \sum_{i=1}^{m} v_i \nabla h_i(\boldsymbol{x}_0) = \boldsymbol{0}.$$

これは，条件 $\boldsymbol{h}(\boldsymbol{x}_0) = \boldsymbol{0}$ と合わせて，$\boldsymbol{x}_0$が等式制約条件つき最適化問題の解となる必要条件を構成する

v_i はラグランジュ乗数と呼ばれる

17. 最適化

– 例

制約条件

$$h(\boldsymbol{x}) = {x_1}^2 + {x_2}^2 - 1 = 0.$$

の下で以下を最大化する

$$f(\boldsymbol{x}) = x_1 + x_2$$

必要条件は以下の通り

$$\nabla f(\boldsymbol{x}) + v\nabla h(\boldsymbol{x}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + v \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 2vx_1 + 1 \\ 2vx_2 + 1 \end{bmatrix} = \mathbf{0},$$

$$h(\boldsymbol{x}) = {x_1}^2 + {x_2}^2 - 1 = 0.$$

これから

$$x_1 = x_2 = -\frac{1}{2v} \text{ and } {x_1}^2 + {x_2}^2 - 1 = 2\left(\frac{1}{2v}\right)^2 - 1 = 0$$

が得られ, $v = \pm\frac{\sqrt{2}}{2}$ $x_1 = x_2 = \pm\frac{\sqrt{2}}{2}$ が導かれる

17. 最適化

解は二つある

$$\begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \quad \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}.$$