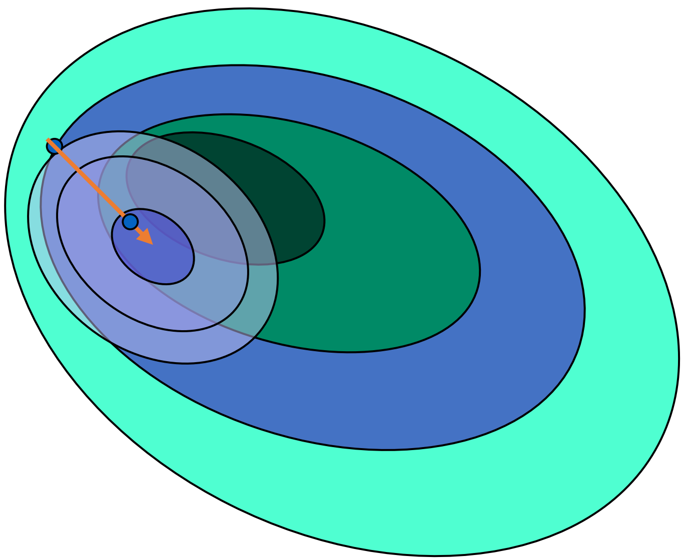
制約なし問題の解法①

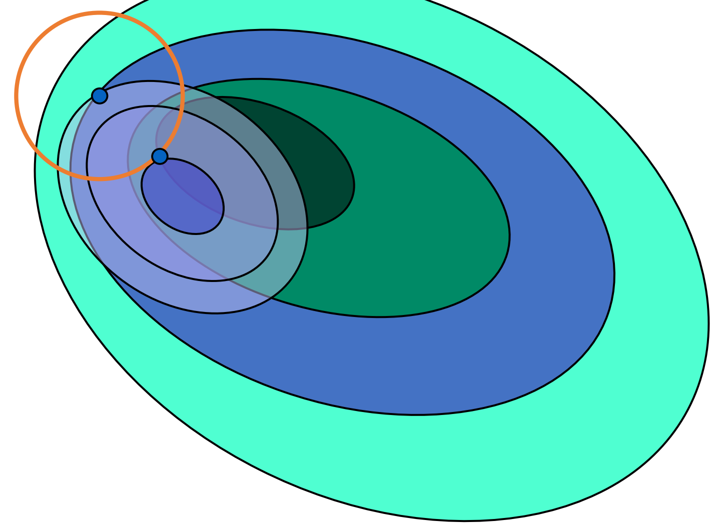
山下信雄

大域的収束性を持たせる技術

- 直線探索法と信頼領域法 -



直線探索法

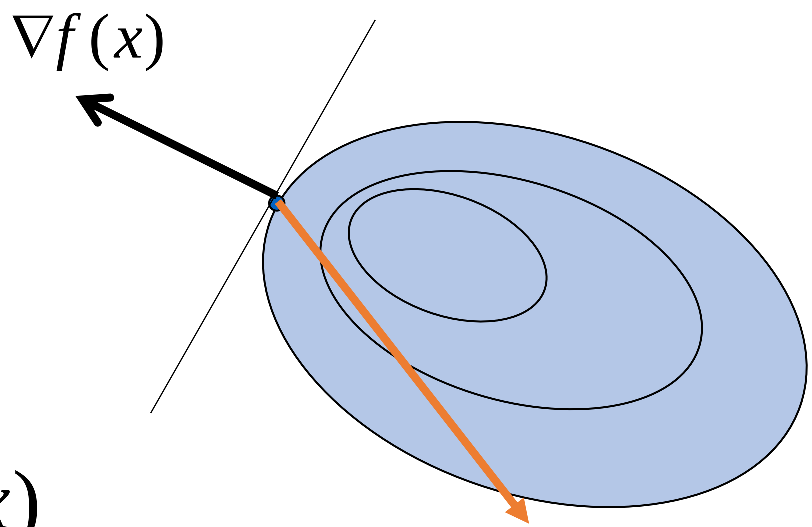


信頼領域法

直線探索法

降下方向：

$$\nabla f(x)^{\top} d < 0$$



ステップ幅：

$$f(x+td) < f(x)$$

を満たすようにステップ幅 t を選ぶ

テーラー展開：

$$\Rightarrow \quad f(x+td) = f(x) + t\nabla f(x)^{\top} d + o(t)$$

$$\frac{f(x+td) - f(x)}{t} = \nabla f(x)^{\top} d + \frac{o(t)}{t} < 0$$

直線探索法
- ステップ幅の決め方 -

厳密に解く：

$$\boxed{\begin{array}{l} \min\ f(x+td) \\ \text{s.t.}\ \ t \in R \end{array}} \quad \Rightarrow \quad \nabla f(x+td)^\top d = 0$$

[解法]　黄金分割法，補間法

近似的に解く：

[アルミホのルール] $\rho \in (0,1)$

$$f(x+td) \le f(x) + \rho t \nabla f(x)^\top d$$

[ウルフのルール]　アルミのルールと

$$|\nabla f(x+td)^\top d| \le -\gamma \nabla f(x)^\top d$$

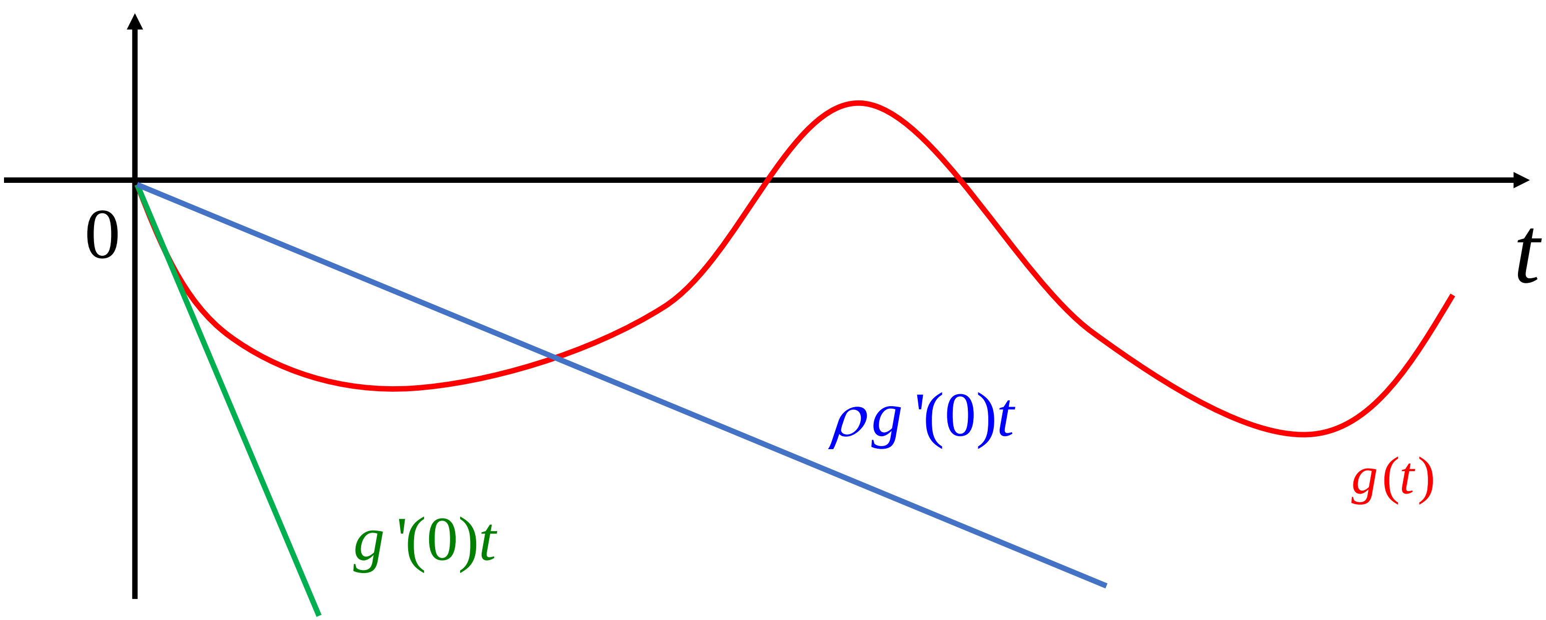
アルミホのルール

$$f(x+td) \le f(x) + \rho t \nabla f(x)^\top d$$

$g(t) = f(x+td) - f(x)$ とする.

$$g(0) = 0,\ g'(t) = \nabla f(x+td)^\top d,\ g'(0) = \nabla f(x)^\top d$$

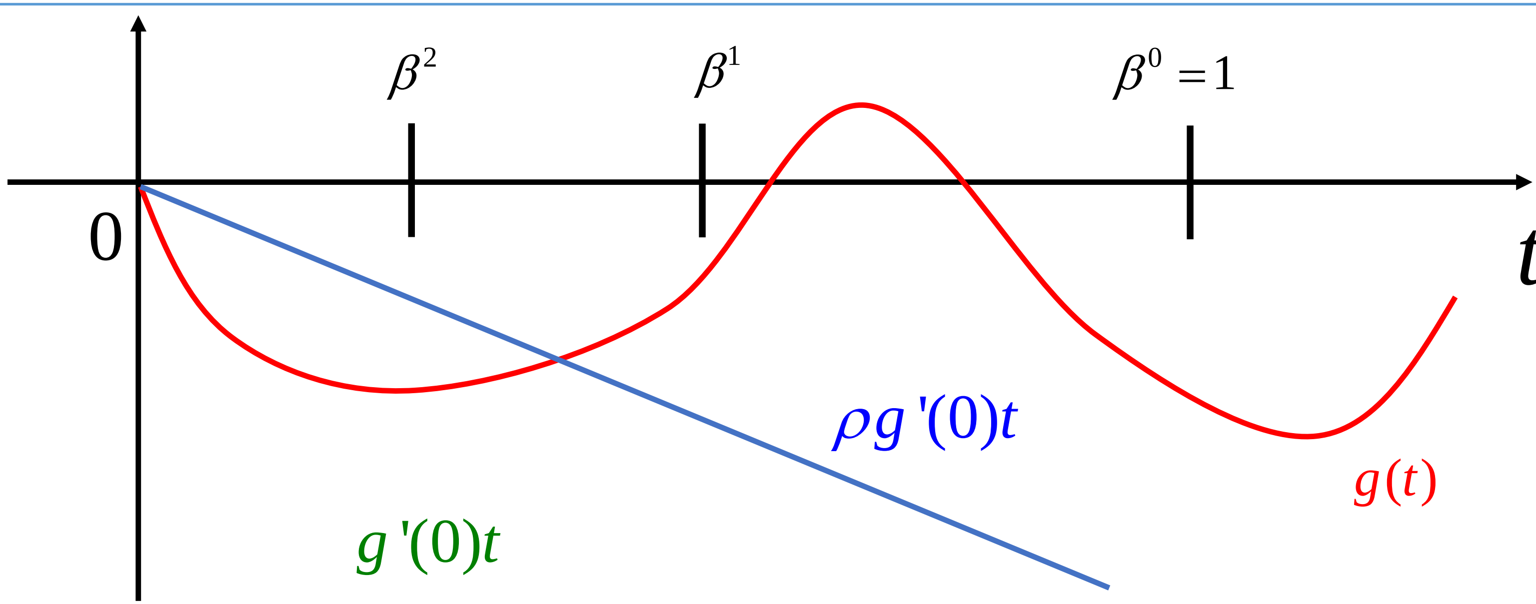
アルミホのルール： $g(t) \le \rho g'(0) t < 0$



アルミホのルールを満たすステップ幅の求め方

$\beta \in (0,1)$ とする.
次の不等式をみたす最小の非負の整数 ℓ を ℓ_k とし,
$t_k = \beta^{\ell_k}$ とする.

$$g(\beta^{\ell}) \le \rho g'(0)\beta^{\ell}$$



制約なし最小化問題の解法
- 直線探索法 -

ステップ１：　降下方向 d^k を求める.

ステップ２：　ステップ幅 t_k を定める.

ステップ３：　$x^{k+1} = x^k + t_k d^k$ として，ステップ１へ

適当な仮定のもとで，停留点($\nabla f(x^*) = 0$)
に 大域的収束する.

大域的収束性

定理　次の不等式を満たす $c_1, c_2 > 0$ が存在するとする.

$$\nabla f(x^k)^\top d^k \le -c_1 \left\| \nabla f(x^k) \right\|^2$$

$$\left\| d^k \right\| \le c_2 \left\| \nabla f(x^k) \right\|$$

$\rho, \beta \in (0,1)$ としたアルミホのルールでステップ幅を求めるとする.
このとき, $\left\{ x^k \right\}$ の任意の集積点は停留点である.

注意：1点に収束するわけではない.
大域的最適解に収束するわけではない.

部分列で考える

いま $\{x^k\}$ の集積点のひとつを $\bar{x}$ とする.

このとき，部分列 $\{x^k\}_K$ が存在し,

$$\lim_{\substack{k\to\infty \\ k\in K}} x^k = \bar{x}$$

証明①：ステップ幅の存在

任意の k に対してアルミホのルールを満たす t_k が存在する.
(背理法) もし存在しないとすると,

$$f(x^k + td^k) - f(x^k) > \rho t \nabla f(x^k)^\top d^k \quad \forall t \in (0,1]$$

となるが,

$$\frac{f(x^k + td^k) - f(x^k)}{t} > \rho \nabla f(x^k)^\top d^k \quad \forall t \in (0,1]$$

より， $t \to 0$ とすると， $\nabla f(x^k)^\top d^k \geq \rho \nabla f(x^k)^\top d^k$
$\rho < 1$ より， $\nabla f(x^k)^\top d^k \geq 0$ 　となり
降下方向であることに矛盾.

証明②：　$t_k \nabla f(x^k)^\top d^k \to 0$

$$f(x^{k+1}) - f(x^k) \le \rho t_k \nabla f(x^k)^\top d^k$$

であるから，この不等式を足すと，

$$\sum_{i=0}^{k} \left\{ f(x^{i+1}) - f(x^i) \right\} \le \rho \sum_{i=0}^{k} t_k \nabla f(x^k)^\top d^k$$

なるから，

$$f(x^*) - f(x^0) \le f(x^{k+1}) - f(x^0) \le \rho \sum_{i=0}^{k} t_k \nabla f(x^k)^\top d^k$$

ここで，　$t_k \nabla f(x^k)^\top d^k < 0$　であることから，

$$t_k \nabla f(x^k)^\top d^k \to 0$$

証明③： t_k が 0 に収束しないとき

$K_1 \subseteq K, \left\{ \nabla f(x^k)^\top d^k \right\}_{K_1} \to 0$ となる部分列が存在する．
このとき，定理の仮定：

$$c_1 \left\| \nabla f(x^k) \right\|^2 \leq -\nabla f(x^k)^\top d^k$$

より，その部分列上では

$$\nabla f(x^k) \to 0$$

よって，$\nabla f(\overline{x}) = 0$

証明④： $t_k \to 0$ のとき

十分大きい k では， $t_k < 1$ となる．よって，

$$f\left(x^k + \frac{t_k}{\beta} d^k\right) - f(x^k) > \rho \frac{t_k}{\beta} \nabla f(x^k)^\top d^k$$

平均値の定理より

$$f\left(x^k + \frac{t_k}{\beta} d^k\right) - f(x^k) = \frac{t_k}{\beta} \nabla f\left(x^k + \theta \frac{t_k}{\beta} d^k\right)^\top d^k$$

となる $\theta \in [0,1]$ が存在する．よって，

$$\nabla f\left(x^k + \theta \frac{t_k}{\beta} d^k\right)^\top d^k > \rho \nabla f(x^k)^\top d^k$$

ここで部分列上では $\left\{\nabla f(x^k)\right\}_K$ が有界となるから，
定理の仮定より $\left\{d^k\right\}_K$ も有界となる．

よって，$\left\{d^k\right\}_K$ は集積点 $\bar{d}$ をもつ．
つまり，さらなる部分列 $K_2 \subseteq K$ が存在して，

$$\lim_{\substack{k\to\infty \\ k\in K_2}} d^k = \bar{d}$$

前のスライドの最後の不等式に対して $k \in K_2, k \to \infty$ とすると，

$$\nabla f(\bar{x})^\top \bar{d} \geq \rho \nabla f(\bar{x})^\top \bar{d}$$

つまり， $\nabla f(\bar{x})^\top \bar{d} \geq 0$ を得る．定理の仮定から，

$$\lim_{\substack{k\to\infty \\ k\in K_2}} \left\|\nabla f(x^k)\right\|^2 \leq -\frac{1}{c_1} \lim_{\substack{k\to\infty \\ k\in K_2}} \nabla f(x^k)^\top d^k = -\frac{1}{c_1} \nabla f(\bar{x})^\top \bar{d} \leq 0$$