

$$f(x), \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

↑ 変数 x .

Optimizationの基本問題

$f(x)$ が最小となる点 x^* を求めたい.

↓ $f(x)$ が極小となる点 x^* を求めたい.

$$\frac{\partial f}{\partial x_1} = \dots = \frac{\partial f}{\partial x_n} = 0 \text{ を解く. } n \text{ が大になると大変.}$$

$$a, h \in \mathbb{R}^n, \quad |h| \text{ は小.}$$

$$f(a+h) = f(a) + (\nabla f)(a) \cdot h + O(|h|^2) \quad \text{①}$$

$$(n=2 \quad \nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right), \quad \frac{\partial f}{\partial x_1}(a) \cdot h_1 + \frac{\partial f}{\partial x_2}(a) \cdot h_2)$$

$$\textcircled{2} \quad h = -(\nabla f)(a)^T \cdot \varepsilon \quad \text{①に代入}$$

$$f(a+h) = f(a) - \frac{1}{2} |(\nabla f)(a)|^2 \cdot \varepsilon + O(\varepsilon^2)$$

$(\nabla f)(a) \neq 0$ かつ, $\exists$ 十分に小さい s ,
 $f(a+h) < f(a)$.

$h = d \varepsilon$ $d \in \mathbb{R}^n$ 対し.

$(\nabla f)(a) \cdot d < 0$ 同様に.

$f(a+h) < f(a)$.

減少方向, descent direction.

プロキシミティの基本型

$x_k \in \mathbb{R}^n$ 対し x_{k+1}

漸化式

$$x_{k+1} = x_k + d_k \varepsilon_k$$

と書き換える.

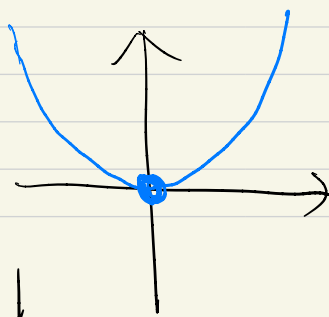
$$x_k \rightarrow x^*$$

と収束する.

例. $n=1$. $f(x) = x^2$.

$$\nabla f = 2x.$$

$$d_k = -2x_k. \quad \varepsilon_k = 0.1.$$



A

Step 幅.

減少方向.

$$x_{k+1} = x_k - 2x_k \times 0.1$$

$$= (1 - 2 \times 0.1) x_k$$

$$= 0.8 x_k$$

$$x_0 = 1$$

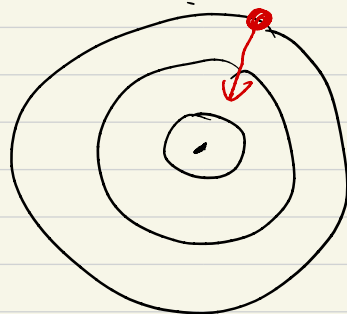
$$x_k = (0.8)^k \longrightarrow 0 \quad (k \rightarrow \infty)$$

例1. $f(x_1, x_2) = x_1^2 + x_2^2$

等高線

$$\nabla f = (2x_1, 2x_2)$$

$$-\nabla f = (-2x_1, -2x_2)$$



問題1. 減少方向 d_k をどう決めよう?

問題2. Step length ϵ_k をどう決めよう?

例. 2次式

$g \in \mathbb{R}^n$ 定. H 対称行列

$x \in \mathbb{R}^n$ 変数.

$$f(x) = g^T x + \frac{1}{2} x^T H x \quad \text{を考へよう.}$$

(a) 2)

$$\nabla f(\mathbf{r}) = \mathbf{g} + \mathbf{H} \mathbf{r}.$$

$$\textcircled{c1} \quad \frac{1}{2} \mathbf{r}^T \mathbf{H} \mathbf{r} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{r}_i \mathbf{H}_{ij} \mathbf{r}_j$$

$$\frac{\partial}{\partial \mathbf{r}} (\quad) = \frac{1}{2} \sum_{j=1}^n \mathbf{H}_{ij} \mathbf{r}_j + \frac{1}{2} \sum_{i=1}^n \mathbf{r}_i \mathbf{H}_{ij}$$

$$\left(\begin{array}{c} \mathbf{H}_{ij} \\ = \mathbf{H}_{ji} \end{array} \right) \Rightarrow \sum_{j=1}^n \mathbf{H}_{ij} \mathbf{r}_j \quad //$$

$$\nabla f(\mathbf{r}) = 0 \text{ となる条件は、 } \mathbf{r} = -\mathbf{H}^{-1} \mathbf{g}.$$

一般の $f(\mathbf{x})$.

$$\mathbf{g}_k = (\nabla f)(\mathbf{x}_k)^T \quad \text{Hessian 行列.}$$

$$\mathbf{H}_k = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right) (\mathbf{x}_k)$$

$\mathbf{r} \in \mathbb{R}^n$, $\mathbf{g}_k$ と $\mathbf{H}_k$ Taylor 展開.

$$f(\mathbf{x}_k + \mathbf{r}) = f(\mathbf{x}_k) + \mathbf{g}_k^T \mathbf{r} + \frac{1}{2} \mathbf{r}^T \mathbf{H}_k \mathbf{r} + O(|\mathbf{r}|^3)$$

Alg (Newton) 法.

$$\mathbf{d}_k = -\mathbf{H}_k^{-1} \mathbf{g}_k \quad \text{etc.}$$

漸化式 (A) を計算.

lem 6.1. H_k が 正定値 なら、 $-H_k^{-1} g_k$ は減少方向。
 → 上記の固有値が正.

① $g_k^T (-H_k^{-1} g_k) = -g_k^T H_k^{-1} g_k < 0$
 H_k^{-1} も正定値、なので ↑ は任意の g_k に対して、正. //

Newton法のバリエーションが BFGS法.

n が大. Hessian を計算しない.

Hessian に近いものを作る。

$$n \Rightarrow \phi(x+h) - \phi(x) \doteq \phi'(x) \cdot h$$

$$\phi'(x+h) - \phi'(x) \doteq \phi''(x) \cdot h. \quad \text{の } n \text{ 変数での類い.}$$

$$\underbrace{g_{k+1}}_{(\nabla f)(x_{k+1})} - \underbrace{g_k}_{(\nabla f)(x_k)} \doteq H_k \cdot (x_{k+1} - x_k)$$

$$(\nabla f)(x_{k+1}) - (\nabla f)(x_k)$$

① 練習 //

$$H_k^{-1} (g_{k+1} - g_k) \doteq (x_{k+1} - x_k)$$

$$B_R (g_{k+1} - g_k) = (x_{k+1} - x_k)$$

と仮定 B_R を作り, $-B_R g_k \in d_k$ として, 連立方程式を作るのが, BFGS alg

A

の方法.

$\swarrow$ $n \times n$ 対称行列.

lem 6.2. $x_{i+1} - x_i = B_i (g_{i+1} - g_i)$ とある.

$$B_{i+1} = B_i + \frac{(\overset{\text{ベクトル}}{x_{i+1} - x_i})(\overset{\text{ベクトル}}{x_{i+1} - x_i})^T}{(x_{i+1} - x_i)^T (g_{i+1} - g_i)}$$

$$= \frac{(B_i (\overset{\text{ベクトル}}{g_{i+1} - g_i})) (\overset{\text{ベクトル}}{B_i (g_{i+1} - g_i)})^T}{\boxed{(g_{i+1} - g_i)^T B_i (g_{i+1} - g_i)}}$$

と示めると.

$x_{i+1} - x_i = B_{i+1} (g_{i+1} - g_i)$ が成立.

① 注

ベクトル

行列

行列

rank = 1.

②

$\swarrow$

右から, $(g_{i+1} - g_i)$ をかける.

$$\begin{aligned}
 B_{i+1}(g_{i+1} - g_i) &= B_i(g_{i+1} - g_i) \\
 + \frac{(x_{i+1} - x_i) ((x_{i+1} - x_i)^T (g_{i+1} - g_i))}{(x_{i+1} - x_i)^T (g_{i+1} - g_i)} \\
 - \frac{(B_i(g_{i+1} - g_i)) ((g_{i+1} - g_i)^T B_i(g_{i+1} - g_i))}{\cancel{\text{略}}}
 \end{aligned}$$

$$\begin{aligned}
 &= B_i(g_{i+1} - g_i) + (x_{i+1} - x_i) \rightarrow B_i(g_{i+1} - g_i) \\
 &= (x_{i+1} - x_i) //
 \end{aligned}$$

lem 6.3 同いC
(BFGS)

$$B_{i+1} = B_i + \text{lem 6.2}$$

行3!

$$+ (g_{i+1} - g_i)^T B_i (g_{i+1} - g_i) U \cdot U^T$$

$$U = \frac{x_{i+1} - x_i}{(x_{i+1} - x_i)^T (g_{i+1} - g_i)} - \frac{B_i (g_{i+1} - g_i)}{(g_{i+1} - g_i)^T B_i (g_{i+1} - g_i)}$$

も.

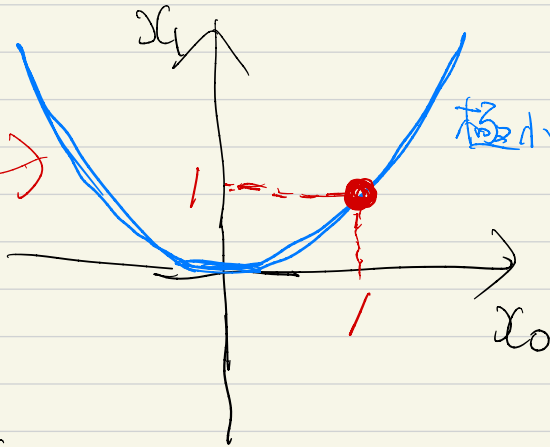
$$x_{i+1} - x_i = B_{i+1} (g_{i+1} - g_i) \text{ をみたす.}$$

証. 同い計算, 略.

$d_k = -B_k g_k$ とおきながら BFGS 法、

$n=2$ の banana fit. (Rosenbrock 関数)

$$f(x_0, x_1) = 100(x_1 - x_0^2)^2 + (1 - x_0)^2$$



bfgs0.py.

Conjugate gradient method.

$$f(x) = \frac{1}{2} x^T H x - b^T x \quad \text{に對して説明.}$$

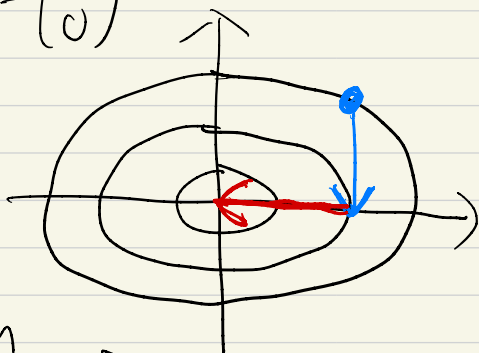
$x \in \mathbb{R}^n$ 変数

H $n \times n$ 正定値対称行列.

$b \in \mathbb{R}^n$ 変数.

$$\nabla f = Hx - b.$$

例.
 $n=2$. $H = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ $b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$



Def. $P_0, P_1, \dots, P_{n-1} \in \mathbb{R}^n$ が

H に n 回 conjugate して $P_i^T H P_j = 0$
 $(i \neq j)$.

(H に n 回 直交)

step length
 減少方向

Th6.1. $x_{k+1} = x_k + \alpha_k P_k$, x_0 よりスタート.

$$\alpha_k = - \frac{(Hx_k - b)^T P_k}{\|P_k\|_H^2}$$

$$\|P_k\|_H^2 \stackrel{\text{def}}{=} P_k^T H P_k$$

$\therefore n$ 回, $x_n = x^*$ となる.

ポイント α_k は何?

$$\phi(t) = f(x_k + t p_k) \quad t \in \mathbb{R}$$

↑
tの2次関数

$$\phi(t) = \frac{1}{2} (x_k + t p_k)^T H (x_k + t p_k) - b^T (x_k + t p_k)$$

$$= \frac{1}{2} \|p_k\|_H^2 t^2 + \left(\frac{1}{2} p_k^T H x_k + \frac{1}{2} x_k^T H p_k - b^T p_k \right) t + \left(\frac{1}{2} \right)$$

$$= \frac{1}{2} \|p_k\|_H^2 \underbrace{t^2}_{x^2} + \underbrace{(x_k^T H p_k - b^T p_k)}_{(x_k^T H - b^T) p_k} t + \left(\frac{1}{2} \right)$$

$$t = \frac{-(x_k^T H - b^T) p_k}{\|p_k\|_H^2} \text{ で最小. } \rightarrow (\alpha_k) //$$

Th 6.1 の証明: $p_0, \dots, p_{n-1}$ は $\mathbb{R}^n$ の基底なので、

$$x_* - x_0 = \sum_{i=0}^{n-1} \alpha_i p_i \quad \text{と可なり.}$$

一方、漸化式より

$$x_1 - x_0 = \alpha_0 p_0$$

$$x_2 - x_0 = x_2 - x_1 + x_1 - x_0$$

$$= \alpha_1 P_1 + \alpha_0 P_0$$

$$\vdots$$

$$x_l - x_0 = \alpha_0 P_0 + \alpha_1 P_1 + \dots + \alpha_{l-1} P_{l-1}$$

$$\boxed{\alpha_i = \alpha_i} \text{ を示す.}$$

②

$P_R^T H$ を ① ② の左よりかいた?

conjugate の def より

$$P_R^T H (x_* - x_0) = \alpha_R \|P_R\|_H^2$$

$l > R$ のとき

$$P_R^T H (x_l - x_0) = \alpha_R \|P_R\|_H^2$$

$l \leq R$ のとき

$$P_R^T H (x_l - x_0) = 0$$

$$\forall l: P_R^T H (x_l - x_0) = 0$$

$$P_R^T H (x_* - x_R + x_R - x_0) = \alpha_R \|P_R\|_H^2$$

$$P_R^T H (x_* - x_R)$$

x_* は $H x_* - b = 0$ をみたす. ($\nabla f(x_*) = 0$ より)

$$P_R^T b - P_R^T H x_R$$

$$= \underline{P_R^T (b - H x_R)}$$

$$\begin{aligned} \therefore a_R &= \frac{P_R^T (b - H x_R)}{\|P_R\|_H^2} \\ &= \frac{(b - H x_R)^T P_R}{\|P_R\|_H^2} = \alpha_R // \end{aligned}$$

— 一般の $f(x)$ に適用.

$$\begin{cases} P_R = -V_R + \beta_R P_{R-1} \\ V_R = H_R x_R - b_R \end{cases}$$

β_R は P_R と P_{R-1} が H_R について直交する数

H_R は f の x_R における Hessian

b_R は ∇f

これらは略.

① Lagrange 乗数法

← 目的関数

$f(x)$ の極小値を求めよ。

(x は $g(x)=0$ をみたす $\mathbb{R}^n$ の元)

↑ 制約条件

$t \in \mathbb{R}$, $x(t)$ が $g(x)=0$ 上にある曲線

$$\frac{d}{dt} g(x(t)) = \dot{x}^T \nabla g \quad \dot{x} = \frac{dx}{dt}$$

$$\frac{d}{dt} f(x(t)) = \dot{x}^T \nabla f$$

$x(0) = x^*$ とする。この点で $\frac{d}{dt} g(x(t)) = 0$
 $\frac{d}{dt} f(x(t)) = 0$

うっなギ'0>

② $\dot{x}(0)$ は、 $n-1$ 次元分自由に選べる。

$(\nabla g)(x^*)$ と $(\nabla f)(x^*)$ は同じ

1次元空間上にあるベクトル

$$\exists \lambda_0, \lambda_1$$

$$\lambda_0 (\nabla f)(x^*) + \lambda_1 (\nabla g)(x^*) = 0$$

$$\nabla g \neq 0 \text{ 且 } g \geq 0$$

$$\exists \lambda \quad \nabla f(x^*) + \lambda \nabla g(x^*) = 0$$

$$L = f(x) + \lambda g(x)$$

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \lambda \frac{\partial g}{\partial x_i}$$

$$\frac{\partial L}{\partial \lambda} = g$$

$$\nabla = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}, \frac{\partial}{\partial \lambda} \right)$$

$$(\nabla L)(x^*) = 0$$

令 $x^* \neq 0$
 $U = \mathbb{R}^n$

① kernel 法

④ asir

pari (roots, x^3-1);

$x^3-1=0$ の根.

④ Macaulay2 online search. M2 (Macaulay2)

google colab python.

have brew