3.3.2 2次関数補間法

＜基本的な考え方＞

- 最小化したい関数 $\phi(\alpha)$ が 最小解の近傍では２次関数による近似が良い と仮定する．
- 関数 $\phi(\alpha)$ の最小解が区間 $[\alpha_1, \alpha_3]$ に入っているとして，$\alpha_2 \in [\alpha_1, \alpha_3]$ とする．
- これらの３点で２次関数 $q(\alpha) = a\alpha^2 + b\alpha + c$ と $\phi(\alpha)$ が一致するとする．
- $q(\alpha)$ の最小解 $\bar{\alpha}$ は以下のようにして求められる．

$$\bar{\alpha} = -\frac{b}{2a} = \frac{\alpha_1 + \alpha_2}{2} + \frac{(\phi(\alpha_1) - \phi(\alpha_2))(\alpha_2 - \alpha_3)(\alpha_3 - \alpha_1)}{2((\alpha_2 - \alpha_3)\phi(\alpha_1) + (\alpha_3 - \alpha_1)\phi(\alpha_2) + (\alpha_1 - \alpha_2)\phi(\alpha_3))} \tag{3.2}$$

＜２次関数補間法：アルゴリズム＞

Step 0: 問題 (3.1) の近似解が存在すると思われる区間 $[\alpha_1, \alpha_3] \ni \alpha_2$ と “解の精度” $\epsilon > 0$ を決める． $\phi(\alpha_1), \phi(\alpha_2), \phi(\alpha_3)$ を計算する．

Step 1: 式 (3.2) より，$\bar{\alpha}$ を求める．

Step 2: もし $(\bar{\alpha} - \alpha_1)(\bar{\alpha} - \alpha_3) \geq 0$ ならば，Step 3へ，そうでなれば Step 4へ．

Step 3: $\alpha_1, \alpha_2, \alpha_3, \bar{\alpha}$ から $\alpha_1 < \alpha_2 < \alpha_3$ を再定義し，$\phi(\alpha_1), \phi(\alpha_2), \phi(\alpha_3)$ を計算する．Step 1へ．

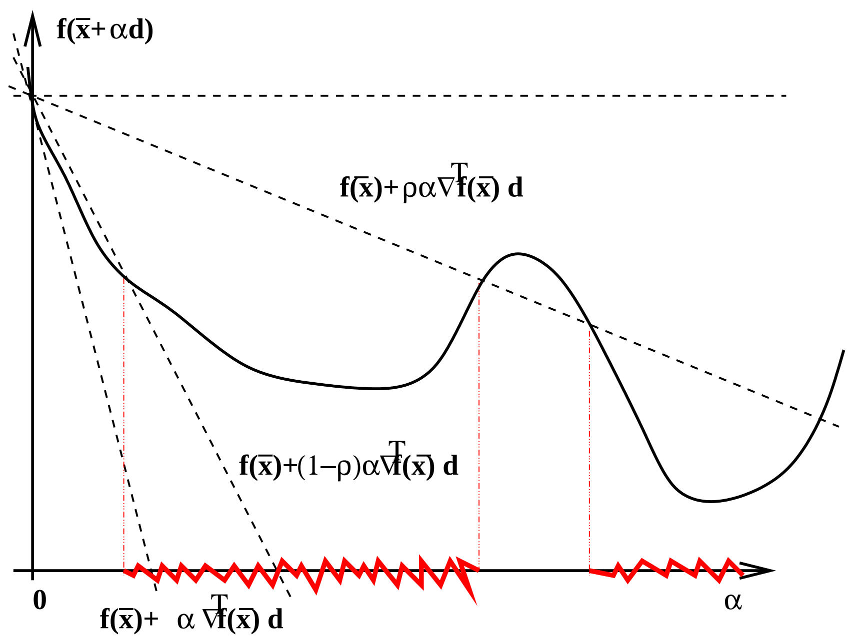
Step 4: もし $|\bar{\alpha} - \alpha_2| < \epsilon$ ならば終了．そうでなければ Step 3 へ．

3.3.3 Armijo-Goldstein ルール

- 直線探索法のなかでは次の Wolfe 条件とともにもっともポピュラーな方法である．
- 最小化する関数が微分可能である必要がある．

パラメータ $\rho \in (0, \frac{1}{2})$ をもちいて、次の不等式を満すステップ・サイズ α を求める．

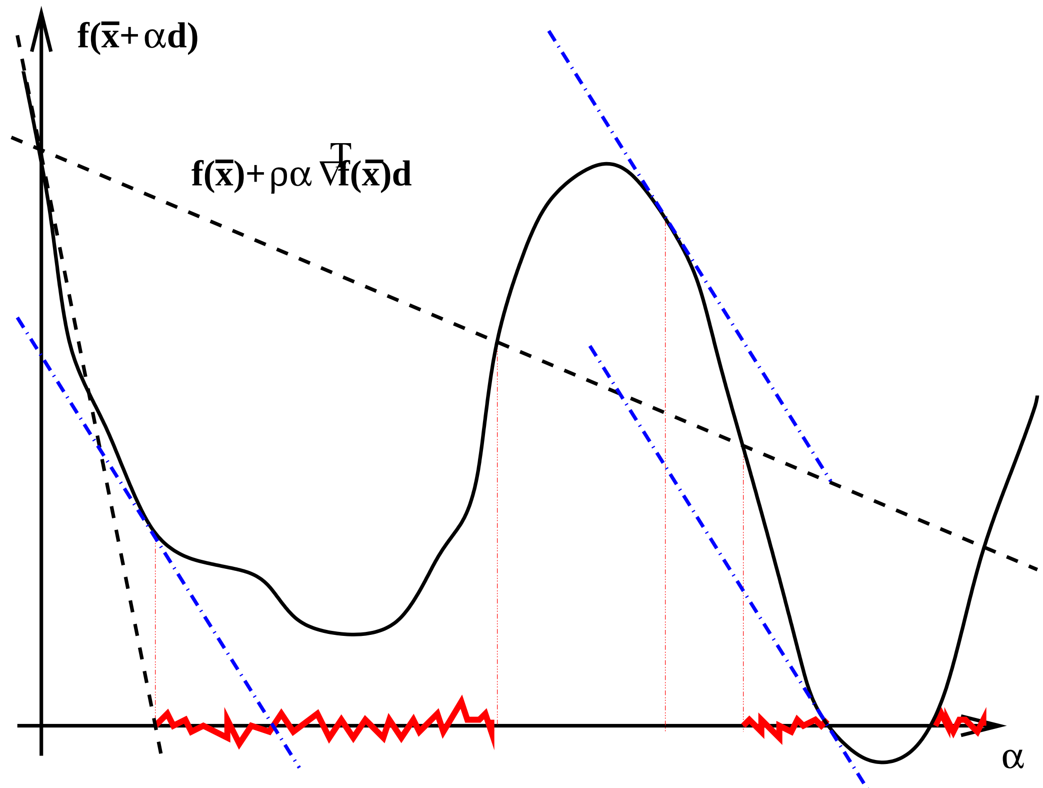
$$f(\bar{x}) + (1-\rho)\alpha\nabla f(\bar{x})^T d \le f(\bar{x} + \alpha d) \le f(\bar{x}) + \rho\alpha\nabla f(\bar{x})^T d$$



3.3.4 Wolfe条件

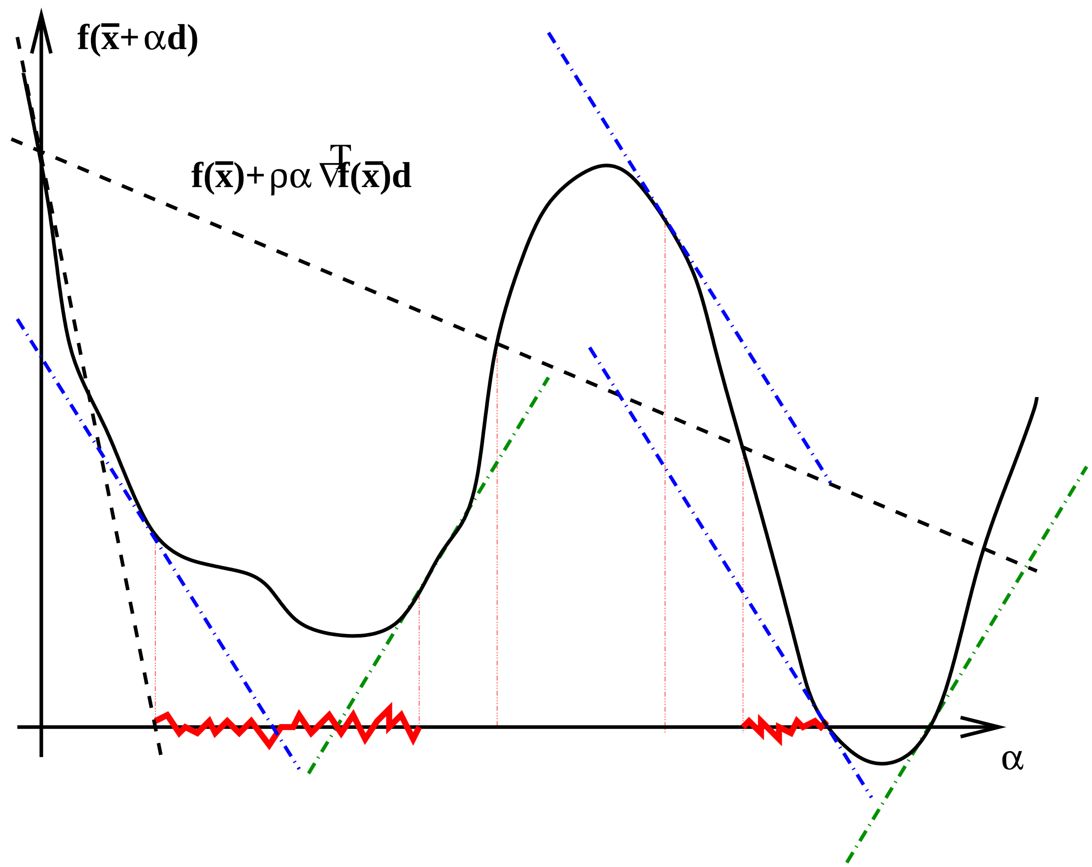
● パラメータ $0 < \rho_1 < \rho_2 < 1$ をもちいて、次の不等式を満す α が Wolfe条件 を満すステップ・サイズと呼ばれている.

$$\begin{aligned} f(\bar{x} + \alpha d) &\le f(\bar{x}) + \rho_1 \alpha \nabla f(\bar{x})^T d \\ \nabla f(\bar{x} + \alpha d)^T d &\ge \rho_2 \nabla f(\bar{x})^T d \end{aligned}$$



• パラメータ $0 < \rho_1 < \rho_2 < 1$ をもちいて、次の不等式を満す α が strong Wolfe条件を満すステップ・サイズと呼ばれている．

$$
\begin{aligned}
f(\bar{x}+\alpha d) &\leq f(\bar{x}) + \rho_1 \alpha \nabla f(\bar{x})^T d \\
|\nabla f(\bar{x}+\alpha d)^T d| &\leq \rho_2 |\nabla f(\bar{x})^T d|
\end{aligned}
$$



3.4　直交探索法

＜基本的な考え方＞

- お互いに直交する方向ベクトルに従って直線探索を行う．

＜直交探索法：アルゴリズム＞

Step 0: d^i $(i = 1, 2, \ldots, n)$ を直交する探索方向，x^0 を初期点，$\epsilon > 0$ を停止基準，$k := 0$ を反復回数とする．

Step 1: $i := 1$ として，$x_i^k := x^0$ とする．

Step 2: $\alpha \in \mathbb{R}$ に対する以下の直線探索を近似的に解く．

$$\{\text{最小化：}\ f(x_i^k + \alpha d^i);\ \text{条件：}\ \alpha \in \mathbb{R}$$

この解を α_i^k として，$x_{i+1}^k := x_i^k + \alpha_i^k d^i$ とおく．

Step 3: $i < n$ であれば，$i := i + 1$ として，step 2 に戻る．

Step 4: $x^{k+1} := x_n^k$ とおき，終了条件 $\|x^{k+1} - x^k\| < \epsilon$ を満たしているならば終了．

Step 5: $k := k + 1$ として step 1 に戻る．

3.5 パターン探索法 [KOWALIK1968]

<基本的な考え方>

- 直交探索を使って，現在の点の近くで目的関数が減少する "谷の方向" を求める．
- 上記の "谷の方向" を使って直線探索を行う．

<パターン探索法：アルゴリズム>

Step 0: d^i $(i = 1, 2, \ldots, n)$ を直交する探索方向，x^0 を初期点，$\epsilon > 0$ を停止基準，$k := 0$ を反復回数とする．

Step 1: d^i $(i = 1, 2, \ldots, n)$ を使って，直交探索を 1 反復行い，x_T^k を求める．

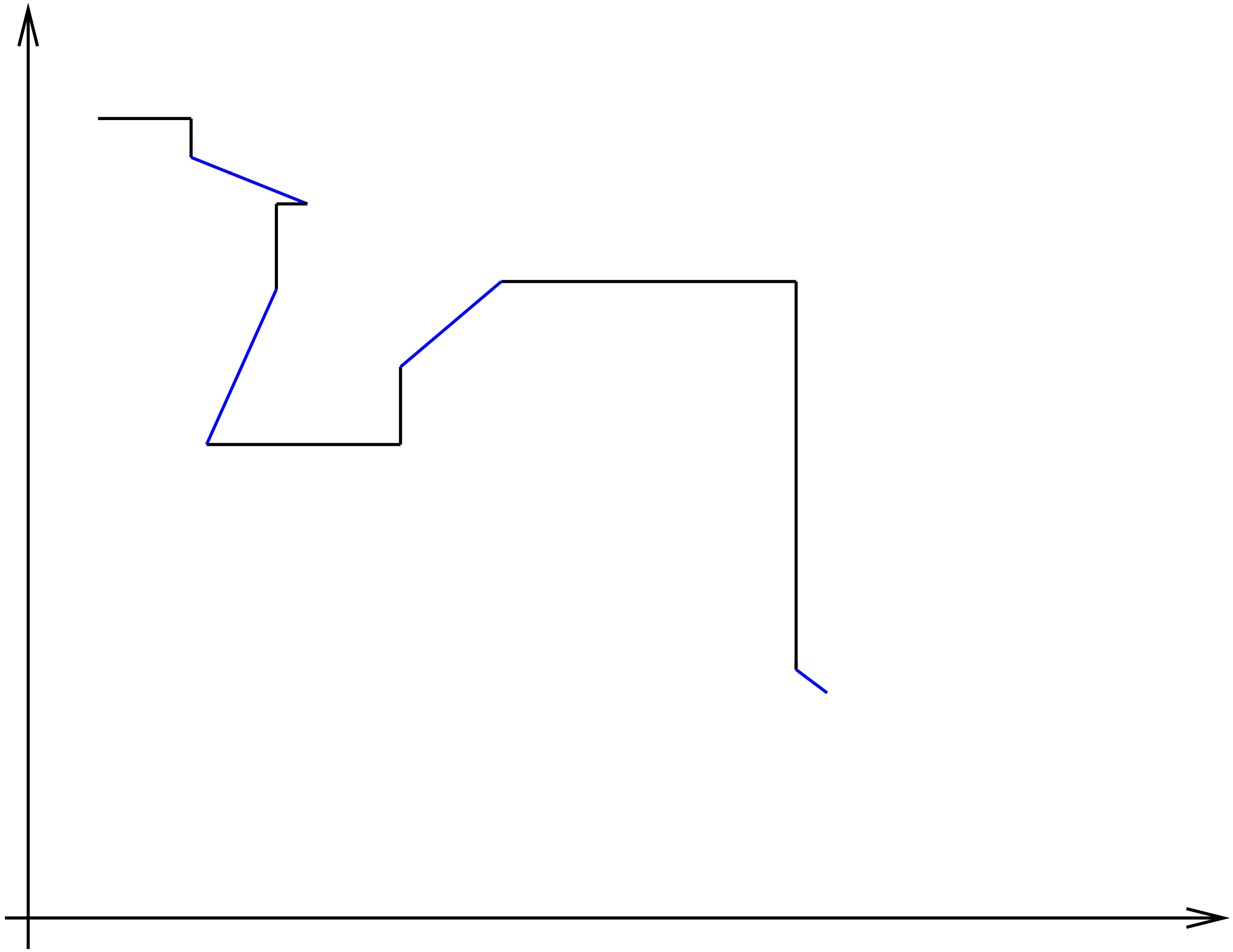
Step 2: $p^k = x_T^k - x^k$ を探索方向として，$\alpha \in \mathbb{R}$ に対する以下の直線探索を近似的に解く．

$$\{\text{最小化：}\ f(x_T^k + \alpha p^k);\ \text{条件：}\ \alpha \in \mathbb{R}$$

この解を α^k として，$x^{k+1} := x_T^k + \alpha^k p^k$ とおく．

Step 3: 終了条件 $\|x^{k+1} - x^k\| < \epsilon$ を満たしているならば終了．

Step 5: $k := k + 1$ として step 1 に戻る．



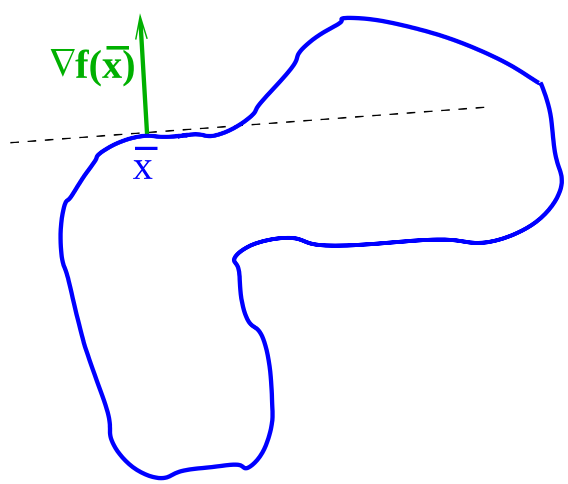
3.6　最急降下法

- 最小化したい関数 $f(x)$ が 連続微分可能だと仮定する．
- $\bar{x} \in \mathbb{R}^n$ において $f(x)$ を線形近似すると $f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x})$ となる．

以下の最小化問題を解くと

$$\begin{cases} \text{最小化：} & f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}) \\ \text{条件：} & \|\bar{x} - x\| = 1 \end{cases}$$

$d = x - \bar{x} = -\nabla f(\bar{x})/\|\nabla f(\bar{x})\|$ となり，勾配ベクトルの逆方向が関数を局所的に減少させる方向であることが分かる．



f の等高線 $\{x \in \mathbb{R} : \ f(x) = f(\bar{x})\}$　　$\bar{x}$ にての勾配ベクトル $\nabla f(\bar{x})$

<最急降下法：アルゴリズム>

Step 0: x^0 を初期点，$\epsilon > 0$ を停止基準，$k := 0$ を反復回数とする．

Step 1: $d^k := -\nabla f(x^k)$ を計算．

Step 2: $\alpha \geq 0$ に対する以下の直線探索を近似的に解く．

$$\{\text{最小化：}\ f(x^k + \alpha d^k);\ \text{条件：}\ \alpha \geq 0$$

この解を α^k として，$x^{k+1} := x^k + \alpha^k d^k$ とおく．

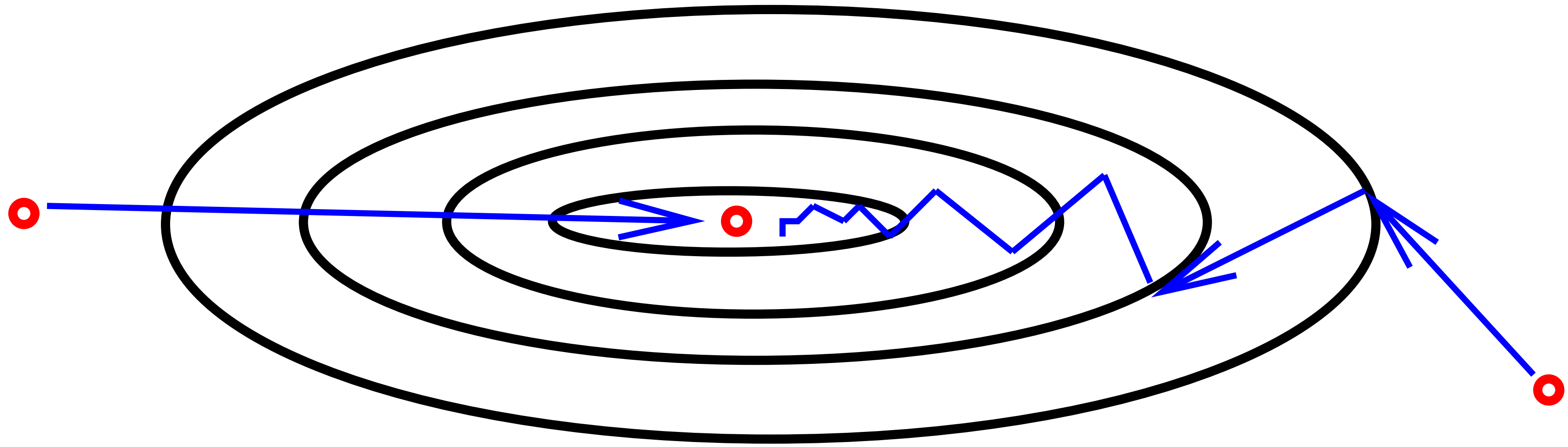
Step 3: $\|x^{k+1} - x^k\| < \epsilon$ ならば終了．そうでなければ $k := k + 1$ として Step 1 へ．

<利点>

- 比較的簡単で理解しやすい方法である．

<欠点>

- 微分を必要とする（数値微分で置き換えることもできる）．
- ２次微分を利用する方法と比べて収束が遅い．



定理 3 [矢部 2006]

$f(x)$ が下に有界で，初期点 x^0 を含む集合 $\{x \in \mathbb{R}^n |\ f(x) \leq f(\bar{x})\}$ を含む集合 U で連続微分可能で，$\nabla f(x)$ が U で Lipschitz 連続であると仮定する．その時，Wolfe 条件を満たす直線探索を用いた最急降下法で生成される点列 $\{x^k\}_{k=0}^{\infty}$ は

$$\lim_{k \to \infty} \|\nabla f(x^k)\| = 0$$

を満足する．

定理4 [SUN2010, YUAN2010]

次の2次強凸関数の最小化を考える．

$$f(x) = \frac{1}{2}x^T H x + h^T x$$

ただし，H は実対称正定値行列とし，条件数 $\kappa = \frac{\lambda_1(H)}{\lambda_2(H)}$ とする．厳密直線探索をもちいた最急降下法では

$$\frac{f(x^{k+1}) - f(x^*)}{f(x^k) - f(x^*)} \le \frac{(\kappa-1)^2}{(\kappa+1)^2}, \quad \frac{\|x^{k+1} - x^*\|_2}{\|x^k - x^*\|_2} \le \frac{\kappa - 1}{\kappa + 1/\sqrt{2\kappa}}$$

が成り立つ．ただし，x^* は $f(x)$ の最小解 $x^* = -H^{-1}h$ であるとする．

3.7 Nesterovの1次法

- Y. Nesterovによって1983年 [NESTEROV1983, NESTEROV2004] に提案されているがごく最近になって注目を浴びるようになった．
- 背景には Compressive sensing など大規模最適化問題を解く必要性などがある．
- 最も基本的なアルゴリズムでは関数に幾つかの条件が課せられる．

- 関数 $f(x)$ の1次微分が Lipschitz 連続である．

$$\exists L > 0, \quad \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n$$

- 関数 $f(x)$ が強凸関数である．

$$\exists \mu > 0, \quad f(x) + \nabla f(x)^T (y - x) + \frac{\mu\|y - x\|^2}{2} \leq f(y), \quad \forall x, y \in \mathbb{R}^n$$

＜Nesterov の1次法：アルゴリズム＞

Step 0: x^0 を初期点，$\gamma_0 \geq \mu$，$v^0 := x^0$，$\epsilon > 0$ を停止基準，$k := 0$ を反復回数とする．

Step 1: 次の2次方程式 $L\alpha_k^2 = (1 - \alpha_k)\gamma_k + \alpha_k\mu$ の根で $[\sqrt{\mu/L}, 1)$ の区間に入っている α_k を求めよ．

Step 2: $\gamma_{k+1} := (1 - \alpha_k)\gamma_k + \alpha_k\mu$，$y^k := \frac{\alpha_k\gamma_k v^k + \gamma_{k+1} x^k}{\gamma_k + \alpha_k\mu}$ を求めよ．

Step 3: $f(y^k)$ と $\nabla f(x^k)$ を計算せよ．

Step 4: 直線探索を用いて $f(x^{k+1}) \leq f(y^k) - \frac{\|\nabla f(x^k)\|^2}{2L}$ を満たす x^{k+1} を近似的に求めよ．

Step 5: $v^{k+1} := \frac{(1-\alpha_k)\gamma_k v^k + \alpha_k\mu y^k - \alpha_k\nabla f(y^k)}{\gamma_{k+1}}$ を求め，$k := k + 1$ として Step 1 へ．

定理 5 [NESTEROV2004]

関数 $f(x)$ の１次微分が Lipschitz 連続であり，強凸関数であるとする．前記のアルゴリズムにて $\gamma_0 := L$ を用いると

$$\frac{f(x^k) - f(x^*)}{\|x_0 - x^*\|_2^2} \le L \min\left\{ \left(1 - \sqrt{\frac{\mu}{L}}\right)^k, \frac{4}{(k+2)^2} \right\}$$

が成り立つ．ただし，x^* は $f(x)$ の最小解である．

- 前述のアルゴリズムで直線探索を行わないで，ステップ・サイズを

$$x^{k+1} := y^k - \frac{1}{L}\nabla f(y^k)$$

のように定数にとれば，アルゴリズムは少し簡略化される．

＜Nesterov の１次法 — 定数ステップ・サイズ：アルゴリズム＞

Step 0: x^0 を初期点，$1 > \alpha_0 \ge \frac{-\mu+\sqrt{\mu^2+\mu L}}{L}$ を初期ステップ・サイズ，$y^0 := x^0$，$\epsilon > 0$ を停止基準，$k := 0$ を反復回数とする．

Step 1 $f(y^k)$ と $\nabla f(y^k)$ を計算せよ．

Step 2: $x^{k+1} := y^k - \frac{\nabla f(y^k)}{L}$ を求めよ．

Step 3: 次の 2 次方程式 $\alpha_k^2 = (1-\alpha_{k+1})\alpha_k^2 + \mu\alpha_{k+1}/L$ の根で $(0,1)$ の区間に入っている α_{k+1} を求めよ．

Step 4: $\beta_k := \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}$ を求めよ．

Step 5: $y^{k+1} := x^{k+1} + \beta_k(x^{k+1} - x^k)$ を求め，$k := k+1$ として Step 1 へ．

- 一見，最小化する関数がかなり限定的に見えるが変数が比較的 “シンプルな” 凸集合上での制約付き最小化問題にも拡張できる．
- さらに関数が微分可能でない場合にも拡張できる．
- 最適化問題が min-max 型の場合にも拡張可能である．
- 収束の鍵を握るのは関数のヘッセ行列のノルムの大きさにある．
- 一方，Lipschitz 定数 L や上記のヘッセ行列ノルムの下界 μ の正しい見積りがあらかじめ要求される．

3.8 ニュートン法

- 最小化したい関数 $f(x)$ が 2 回連続微分可能だと仮定する．

- $\bar{x} \in \mathbb{R}^n$ において $f(x)$ を 2 次近似すると

$$f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \nabla^2 f(\bar{x})(x - \bar{x})$$

となる．

以下の最小化問題を解くと

$$\begin{cases} \text{最小化：} & f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \nabla^2 f(\bar{x})(x - \bar{x}) \\ \text{条件：} & x \in \mathbb{R} \end{cases}$$

$$d = x - \bar{x} = -\left[\nabla^2 f(\bar{x})\right]^{-1} \nabla f(\bar{x})$$

となる．

＜ニュートン法：アルゴリズム＞

Step 0: x^0 を初期点，$\epsilon > 0$ を停止基準，$k := 0$ を反復回数とする．

Step 1: $d^k := -\left[\nabla^2 f(x^k)\right]^{-1} \nabla f(x^k)$ を計算．

Step 2: $x^{k+1} := x^k + d^k$ とおく．

Step 3: $\|x^{k+1} - x^k\| < \epsilon$ ならば終了．そうでなければ $k := k + 1$ として Step 1 へ．

<利点>

- 収束がとても速い（２次収束）.
- 大規模であってもデータの疎性がある程度活用できる.

<欠点>

- 関数の２回微分を必要とする.
- ヘッセ行列の逆行列が存在しないことがあるので，その場合，アルゴリズムが破綻してしまう.

定理６[矢部2006]

関数 $f(x)$ の最小解 x^* の開凸近傍 U で２回連続微分可能とし，$\nabla^2 f(x^*)$ は正定値行列であるとする．また，ヘッセ行列 $\nabla^2 f(x)$ が U において Lipschitz 連続であるとする．このとき，初期点 x^0 を x^* の十分近くに選べば，ニュートン法で生成される点列 $\{x^k\}_{k=0}^{\infty}$ は x^* に収束し，$\|x^{k+1} - x^*\| \leq \nu \|x^k - x^*\|^2$ が成り立つ．ただし，ν は正の定数である.