3.2 シンプレックス法 [KOWALIK1968]

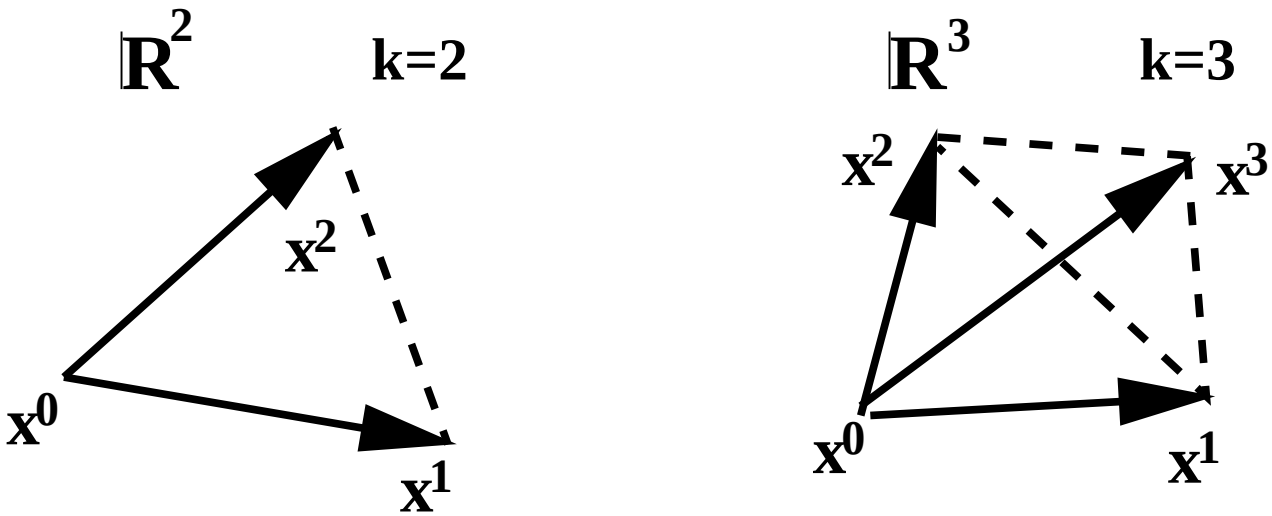
- 線形計画問題に対するシンプレックス法（単体法）とは異なる.
- 関数値だけが分かれば実装できる.

＜＜基本的な考え方＞＞

- ２次元 $(n = 2)$ では三角形，３次元 $(n = 3)$ では四面体を巧妙に反転，拡大，縮小させて１反復で次元に比例した回数の関数評価で最小化を行う.
- 最小点から遠いところでは加速するようになっている.
- 最小点から近いところでは減速し，高精度な近似解を求めるようになっている.

定義１

$(k \leq n)$ $x^0, x^1, \ldots, x^k \in \mathbb{R}^n$ がアフィン独立であるとは $x^1 - x^0, x^2 - x^0, \ldots, x^k - x^0 \in \mathbb{R}^n$ が線形独立である時をいう．



定義２

S が $k+1$ $(\leq n+1)$ のアフィン独立な頂点の多面体であるとき，S を $k-$シンプレックスであるという．

<<シンプレックス法>>

- $x^0 \in \mathbb{R}^n$を１つの頂点として持つ$n-$シンプレックスを用いて実数値関数$f : \mathbb{R}^n \to \mathbb{R}$を最小化する．

<<初期シンプレックス>>

$$S^0 = (x^0, x^1, \ldots, x^n) = x^0 E + \begin{pmatrix} 0 & \delta & \sigma & \cdots & \sigma \\ 0 & \sigma & \delta & \cdots & \sigma \\ 0 & \sigma & \sigma & \cdots & \sigma \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma & \sigma & \cdots & \delta \end{pmatrix} \in \mathbb{R}^{n \times (n+1)}$$

ただし，

$$\delta = \lambda \frac{\sqrt{n+1} + n - 1}{\sqrt{2}n}, \quad \sigma = \lambda \frac{\sqrt{n+1} - 1}{\sqrt{2}n},$$

$\lambda > 0$は<u>スケーリング・ファクター</u>である．Eは全て１要素の行列．

例：

$n = 2$のとき，$\delta \approx 0.97$, $\sigma \approx 0.26$

$n = 3$のとき，$\delta \approx 0.94$, $\sigma \approx 0.24$

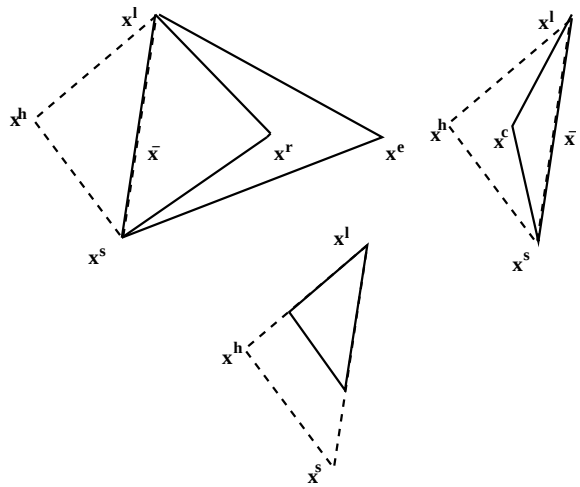
<<シンプレックス法の基本操作>>

各反復にて次の頂点を求める.

$x^h = \mathrm{argmax}_i f(x^i)$	最大の関数値を実現する頂点
$x^s = \mathrm{argmax}_{i \neq h} f(x^i)$	２番目に最大の関数値を実現する頂点
$x^\ell = \mathrm{argmin}_i f(x^i)$	最小の関数値を実現する頂点
$\bar{x} = \frac{\sum_{i \neq h} x^i}{n}$	最大の関数値を実現する頂点を除いた重心

次の操作が定義されている.

鏡映	(reflection)	$x^r = \bar{x} + \alpha(\bar{x} - x^h), \ \ \alpha > 0$
拡張	(expansion)	$x^e = \bar{x} + \gamma(x^r - \bar{x}), \ \ \gamma > 0$
収縮	(contraction)	$x^c = \bar{x} + \beta(x^h - \bar{x}), \ \ \beta \in (0, 1)$
縮小	(reduction)	全ての頂点を x^ℓ の方向に縮小する



$x^h = \mathrm{argmax}_i f(x^i)$ 最大の関数値を実現する頂点
$x^s = \mathrm{argmax}_{i \neq h} f(x^i)$ ２番目に最大の関数値を実現する頂点
$x^\ell = \mathrm{argmin}_i f(x^i)$ 最小の関数値を実現する頂点
$\bar{x} = \frac{\sum_{i \neq h} x^i}{n}$ 最大の関数値を実現する頂点を除いた重心

<<シンプレックス法：アルゴリズム>>

Step 0: シンプレックス S^0 にて，$x^h, x^s, x^\ell, \bar{x}$ を求める．最小解の候補：x^ℓ，最小値の候補：$f(x^\ell)$，$k := 0$ とおく．

Step 1: x^r を求めて，$f(x^r)$ を計算する．

Step 2: $f(x^\ell) \leq f(x^r) \leq f(x^s)$ ならば，x^h を x^r で置き換え，S^{k+1} を更新し，$k := k+1$ として Step 1 にもどる．

Step 3: $f(x^r) < f(x^\ell)$ ならば，$f(x^e)$ を計算する．

Step 3-a: $f(x^e) < f(x^\ell)$ ならば，x^h を x^e で置き換え，S^{k+1} を更新し，$k := k+1$ として Step 1 にもどる．

Step 3-b: そうでなければ，x^h を x^r で置き換え，S^{k+1} を更新し，$k := k+1$ として Step 1 にもどる．

Step 4: 収縮を行うため x^c を求め，$f(x^c)$ を計算する．$(f(x^s) <)\ f(x^r) < f(x^h)$ ならば，x^h を x^r で置き換える．

Step 5: $f(x^c) < f(x^h)$ ならば，x^h を x^c で置き換え，S^{k+1} を更新し，$k := k+1$ として Step 1 にもどる．

Step 6: 縮小を行って，S^{k+1} を更新し，$k := k+1$ として Step 1 にもどる．

$$\text{終了条件:} \quad \frac{1}{n+1}\sum_{i=0}^{n}(f(x^i) - \bar{f})^{\frac{1}{2}} < \varepsilon$$

ただし，$\bar{f} = \dfrac{1}{n+1}\displaystyle\sum_{i=0}^{n} f(x^i)$ とする．

3.3 直線探索法

＜基本的な考え方＞

- 一般的によく使われる手法．
- 出発点 $\bar{x} \in \mathbb{R}^n$ と関数を局所的に減少させる探索方向 $d \in \mathbb{R}^n$ $(d \neq 0)$ が与えられているとする．つまり，十分小さな $\bar{\epsilon} > 0$ が存在し，

$$f(\bar{x} + \bar{\epsilon} d) < f(\bar{x})$$

が成り立っていると仮定する．

- 次の１次元半直線の最適化問題を近似的に解きたい．

$$\begin{cases} \text{最小化：} & \phi(\alpha) \equiv f(\bar{x} + \alpha d) \\ \text{条件：} & \alpha > 0 \end{cases} \tag{3.1}$$

3.3.1 黄金分割法 (golden section method)

- 以下の２次方程式の正の根を考える．

$$\tau^2 + \tau - 1 = 0, \quad \tau = \frac{-1 \pm \sqrt{5}}{2}, \quad \tau_+ = \frac{-1 + \sqrt{5}}{2} \approx 0.618$$

＜黄金分割法：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 $[a, b]$ と “解の精度” $\epsilon > 0$ を決める．

Step 1: $k := 1$

$$\begin{aligned} \alpha_1 &= a + 0.382(b - a) \\ \beta_1 &= a + 0.618(b - a) \end{aligned}$$

を求め，$\phi(\alpha_1)$，$\phi(\beta_1)$ を計算する．$a_1 := a, \quad b_1 := b$.

Step 2: もし $\phi(\alpha_k) > \phi(\beta_k)$ ならば Step 3へ．もし $\phi(\alpha_k) \leq \phi(\beta_k)$ ならば Step 4へ．

Step 3-a: もし $b_k - \alpha_k < \epsilon$ ならば終了し，β_k を近似解として出力．

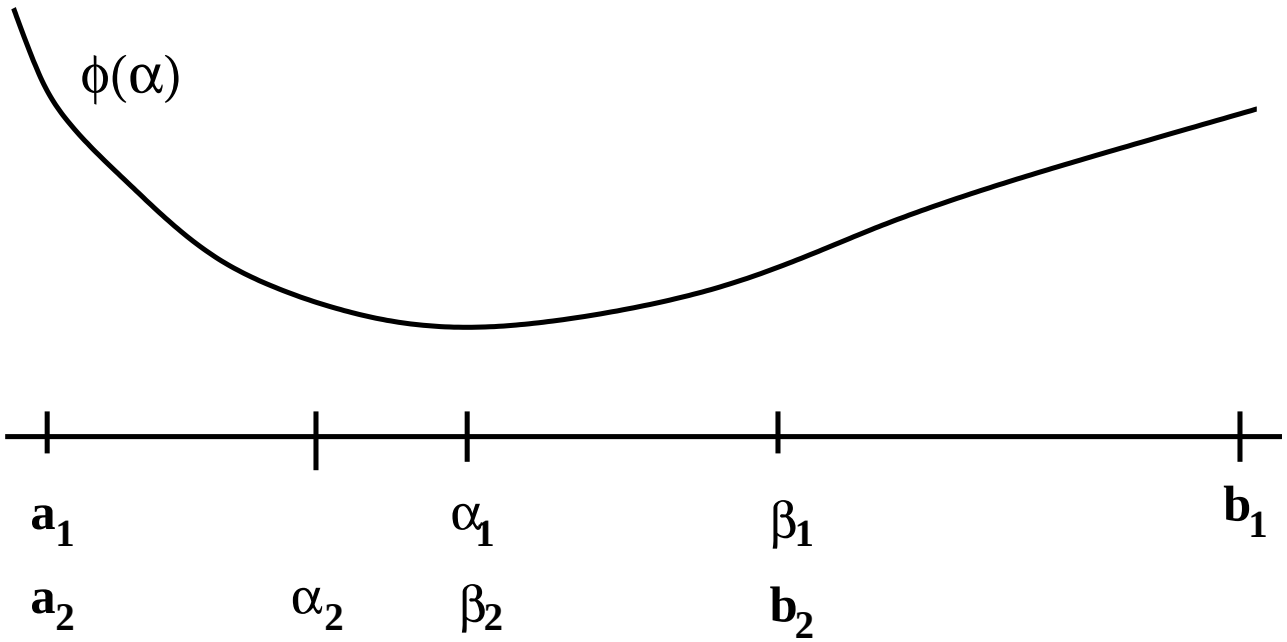
Step 3-b: そうでなければ，

$a_{k+1} := \alpha_k$, $b_{k+1} := b_k$, $\alpha_{k+1} := \beta_k$, $\phi(\alpha_{k+1}) := \phi(\beta_k)$, $\beta_{k+1} := \alpha_{k+1} + 0.618(b_{k+1} - a_{k+1})$, $\phi(\beta_{k+1})$ を計算し，$k := k + 1$，Step 2へ．

Step 4-a: もし $\beta_k - a_k < \epsilon$ ならば終了し，α_k を近似解として出力．

Step 4-b: そうでなければ，

$a_{k+1} := a_k$, $b_{k+1} := \beta_k$, $\beta_{k+1} := \alpha_k$, $\phi(\beta_{k+1}) := \phi(\alpha_k)$, $\alpha_{k+1} := \alpha_{k+1} + 0.382(b_{k+1} - a_{k+1})$, $\phi(\alpha_{k+1})$ を計算し，$k := k + 1$，Step 2へ．



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 <u>最小解の近傍では２次関数による近似が良い</u> 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，そうでなければ 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．

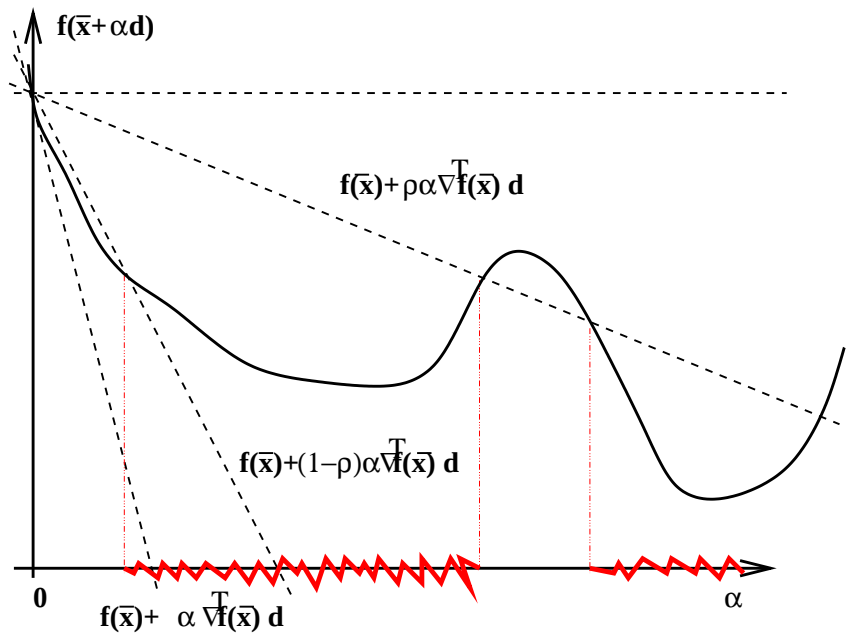
Step 4: もし $|\bar{\alpha} - \alpha_2| < \epsilon$ ならば終了．そうでなければ Step 3へ．

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.

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

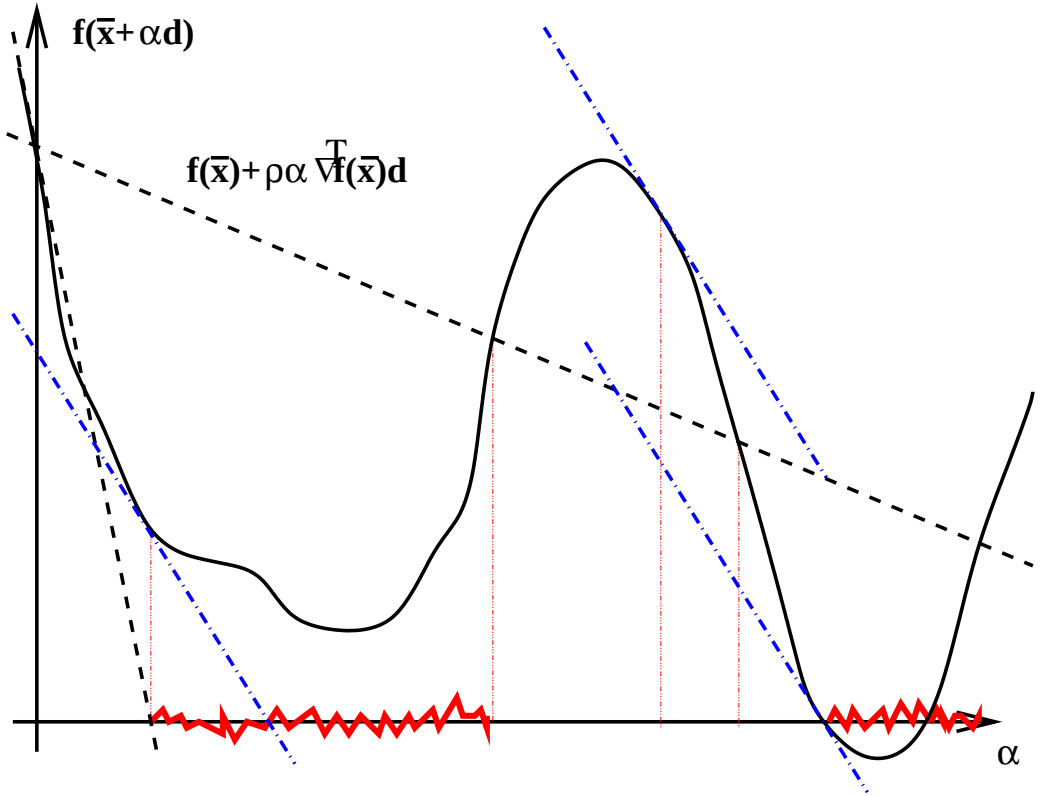


3.3.4 Wolfe条件

● 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.

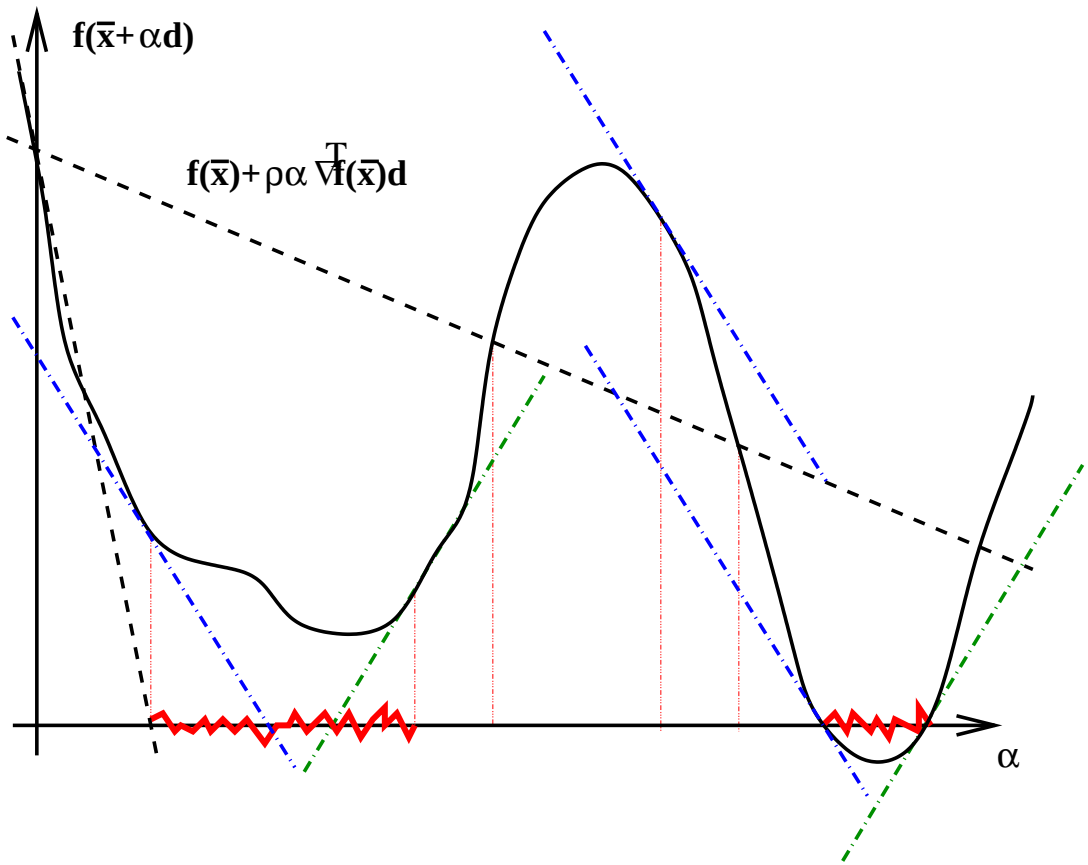
$$f(\bar{x} + \alpha d) \ \leq \ f(\bar{x}) + \rho_1\alpha\nabla f(\bar{x})^T d$$

$$\nabla f(\bar{x}+\alpha d)^T d \;\geq\; \rho_2 \nabla f(\bar{x})^T d$$



● 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 <u>直線探索を近似的に解く</u>．

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

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.

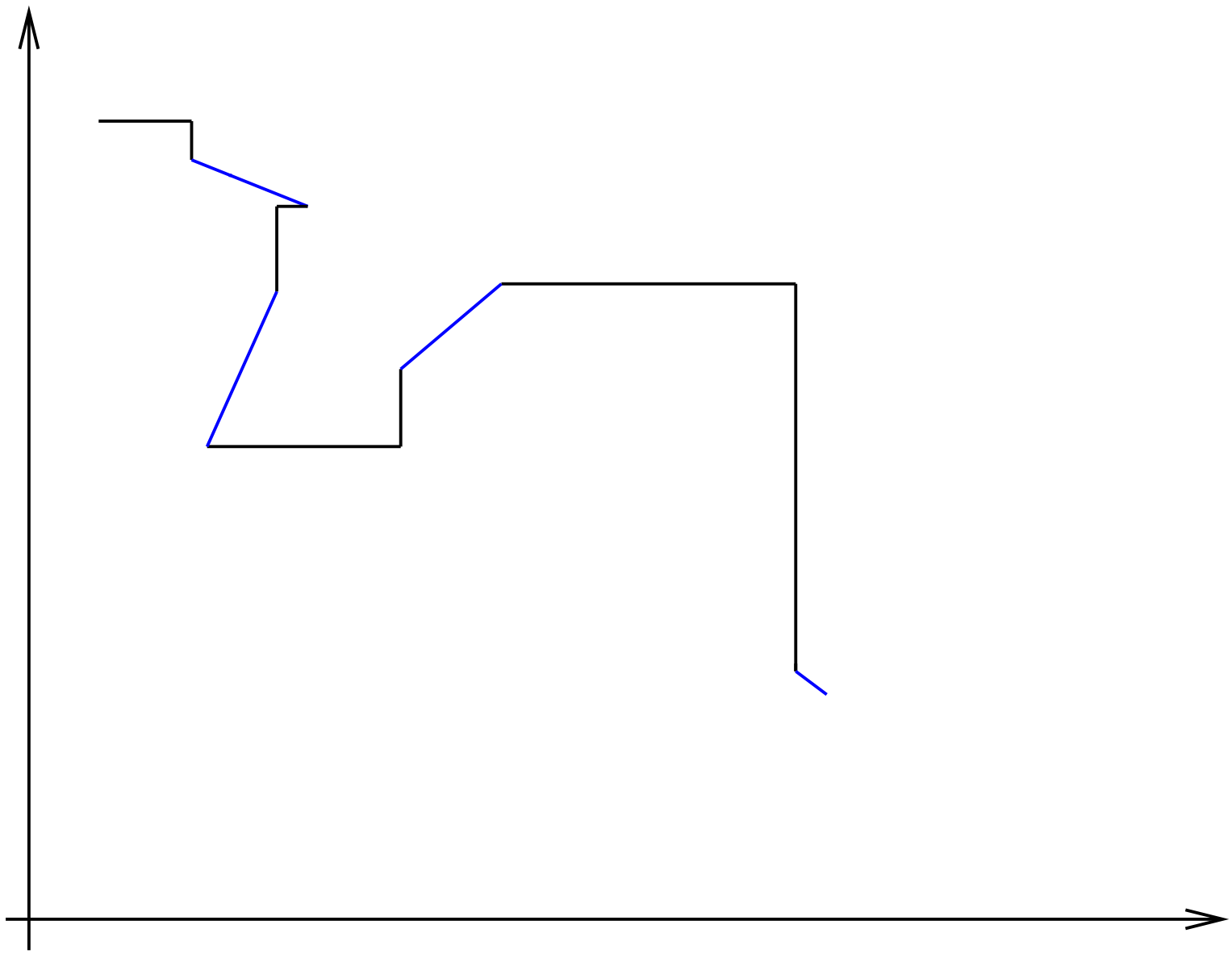
Step 2: $p^k = x_T^k - x^k$ を探索方向として，$\alpha \in \mathbb{R}$ に対する以下の直線探索を近似的に解く.

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

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.



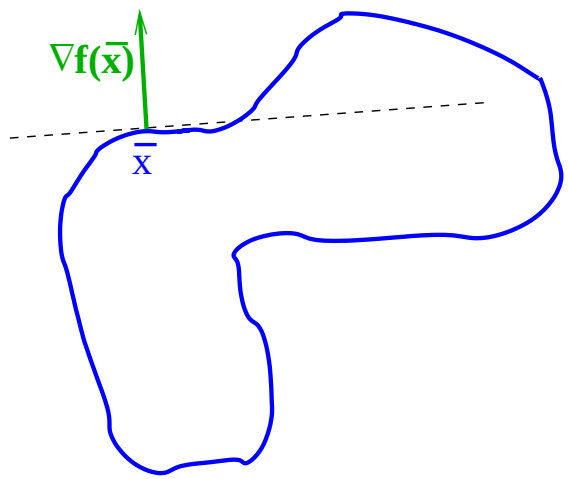
3.6 最急降下法

- 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．

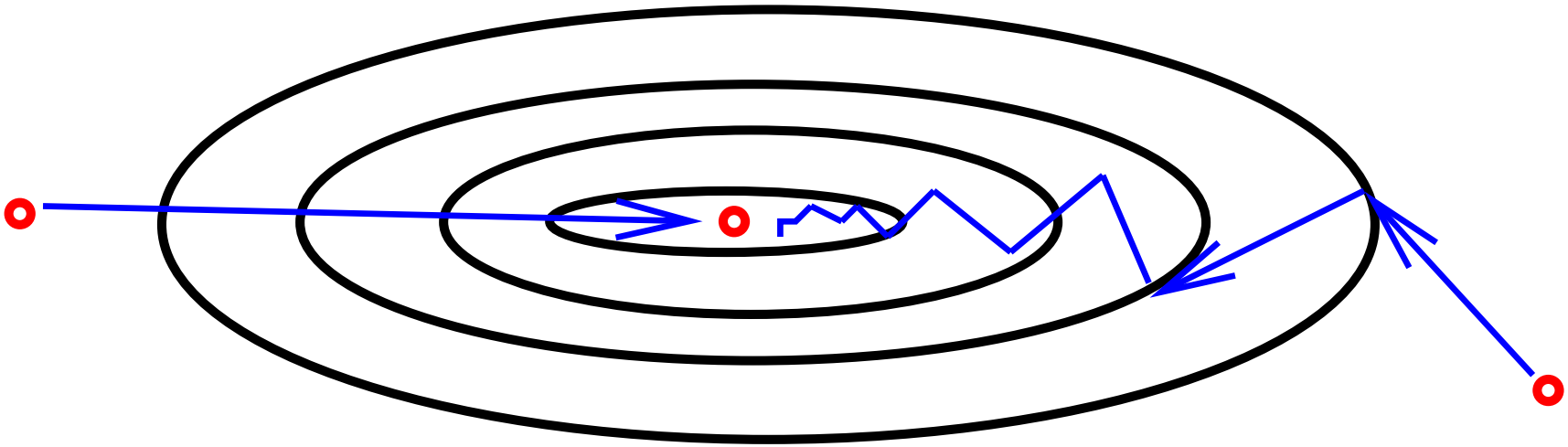
Step 3: $\|x^{k+1} - x^k\| < \epsilon$ ならば終了．そうでなければ $k := k + 1$ として Step 1 へ．

＜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]

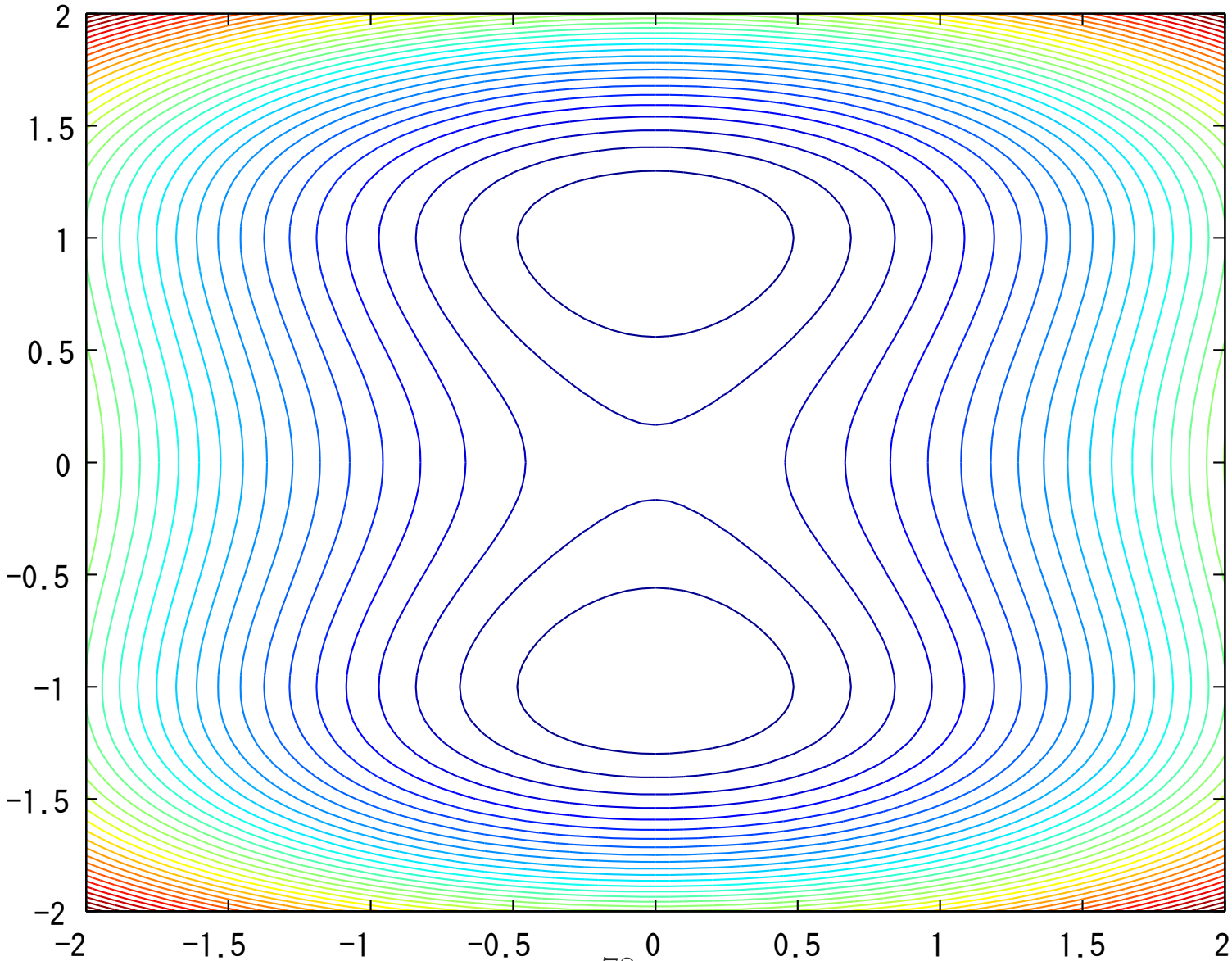
$f(x)$ が下に有界で，初期点 x^0 を含む集合 $\{x \in \mathbb{R}^n \mid f(x) \leq f(\bar{x})\}$ を含む集合 U で連続微分可能で，$\nabla f(x)$ が U で Lipschitz 連続であると仮定する．その時，Wolfe 条件を満たす直線探索を用いた最急降下法で生成される点列 $\{x^k\}_{k=0}^{\infty}$ は

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

を満足する．

例題：

関数 $f(x, y) = \frac{1}{2}x^2 + \frac{1}{4}y^4 - \frac{1}{2}y^2$ に対しては $(0, -1)$ と $(0, 1)$ が局所的最低解であるが，$(0, 0)$ は鞍点である．


2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2

定理4 [SUN2010, YUAN2010]

次の狭義2次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

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

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 加速勾配法（1次法）

- Yu. 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．

- その後，他の研究者からも同じような計算量がかかるアルゴリズムが多数提案されている．

- 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 $f(x)$ は凸関数，もしくは強凸関数である.

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

＜加速勾配法 [NESTEROV2004]：アルゴリズム＞

Step 0: x^0 を初期点，$L \geq \gamma_0 \geq \mu$，$v^0 := x^0$，$\epsilon > 0$ 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

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

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．

＜加速勾配法 [NESTEROV2004] — 定数ステップ・サイズ：アルゴリズム＞

Step 0: x^0 を初期点，$\mu \leq \frac{\alpha_0(\alpha_0 L-\mu)}{1-\alpha_0} \leq L$ を満たす $\alpha_0 \in (0,1)$ 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．

さらに比較的最近の論文 Yu. Nesterov, “Smooth minimization of non-smooth functions,” *Mathematical Programming* 103 (2005), pp. 127–152 ではさらなる改良が行われている．

＜加速勾配法 — ２００５年版：アルゴリズム＞

Step 0: y^0 を初期点，$\epsilon > 0$ 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: $x^k := y^k - \frac{\nabla f(y^k)}{L}$ を求めよ．

Step 3: $z^k := y^0 - \frac{1}{L}\sum_{i=0}^{k} \nabla f(y_i)$ を求めよ．

Step 4: $y^{k+1} := \frac{2}{k+3}z^k + \frac{k+1}{k+3}x^k$ を求め，$k := k+1$ として Step 1 へ．

定理 6 (Yu. Nesterov (2005))

関数 $f(x)$ の 1 次微分が Lipschitz 連続であり，凸関数であるとする．前記のアルゴリズムを用いると

$$\frac{f(x^k) - f(x^*)}{\|x_0 - x^*\|_2^2} \leq L\frac{2}{(k+1)(k+2)}$$

が成り立つ．ただし，x^* は $f(x)$ の最小解である．

- 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の最小固有値の下界に相当する 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 7 [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．

3.9 準ニュートン法

＜基本的な考え方＞

- ニュートン法の利点である速い収束を保ちながら，探索方法 $d = -[\nabla^2 f(\bar{x})]^{-1} \nabla f(\bar{x})$ の計算をさぼりたい．
- つまり $H \approx [\nabla^2 f(\bar{x})]^{-1}$ を近似的に求めたい．
- 数値近似にはいく通りの方法がある．
- $s^k = \nabla f(x^{k+1}) - \nabla f(x^k)$，$y^k = x^{k+1} - x^k$ とする．

対称ランク 1 公式

$$\boldsymbol{H}_{k+1} := \boldsymbol{H}_k + \frac{(\boldsymbol{y}^k - \boldsymbol{H}_k \boldsymbol{s}^k)(\boldsymbol{y}^k - \boldsymbol{H}_k \boldsymbol{s}^k)^T}{(\boldsymbol{y}^k - \boldsymbol{H}_k \boldsymbol{s}^k)^T \boldsymbol{s}^k}$$

DFP (Davidon-Fletcher-Powell) 公式

$$\boldsymbol{H}_{k+1} := \boldsymbol{H}_k + \frac{\boldsymbol{y}^k (\boldsymbol{y}^k)^T}{(\boldsymbol{s}^k)^T \boldsymbol{y}^k} - \frac{\boldsymbol{H}_k \boldsymbol{s}^k (\boldsymbol{s}^k)^T \boldsymbol{H}_k}{(\boldsymbol{s}^k)^T \boldsymbol{H}_k \boldsymbol{s}^k}$$

BFGS (Broyden-Fletcher-Goldfarb-Shanno) 公式

$$\boldsymbol{H}_{k+1} := \left(\boldsymbol{I} - \frac{\boldsymbol{y}^k (\boldsymbol{s}^k)^T}{(\boldsymbol{y}^k)^T \boldsymbol{s}^k} \right) \boldsymbol{H}_k \left(\boldsymbol{I} - \frac{\boldsymbol{s}^k (\boldsymbol{y}^k)^T}{(\boldsymbol{y}^k)^T \boldsymbol{s}^k} \right) + \frac{\boldsymbol{y}^k (\boldsymbol{y}^k)^T}{(\boldsymbol{y}^k)^T \boldsymbol{s}^k} \tag{3.3}$$

- 現時点で最も計算効率が良いとされるのが BFGS 公式である.

＜準ニュートン法：アルゴリズム＞

Step 0: x^0を初期点，$H_0 := I$，$\epsilon > 0$を停止基準，$k := 0$を反復回数とする．

Step 1: $d^k := -H_k \nabla f(x^k)$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．$y^k, s^k$を求める．

Step 4: H_{k+1}をいずれかの公式から計算し，$k := k + 1$として，Step 1へ．

＜利点＞

- ヘッセ行列やその逆行列を必要としない．
- ある条件のもとで超１次収束する．
- 狭義２次凸関数に対しては有限回の反復で収束する．

＜欠点＞

- ニュートン法よりも数値的な安定性に欠ける．

3.10 記憶制限準ニュートン法

＜基本的な考え方＞

- 大規模な問題に対して準ニュートン法を用いると，ヘッセ行列の（密な）逆行列を近似した $\boldsymbol{H}_k$ を保持しなければならないので不利になる．
- $\boldsymbol{H}_k$ を保持する代わりにベクトルの内積で（各反復において）$\widetilde{\boldsymbol{H}_k}$ を計算する．
- $\boldsymbol{s}^k = \nabla \boldsymbol{f}(\boldsymbol{x}^{k+1}) - \nabla \boldsymbol{f}(\boldsymbol{x}^k)$，$\boldsymbol{y}^k = \boldsymbol{x}^{k+1} - \boldsymbol{x}^k$ とする．