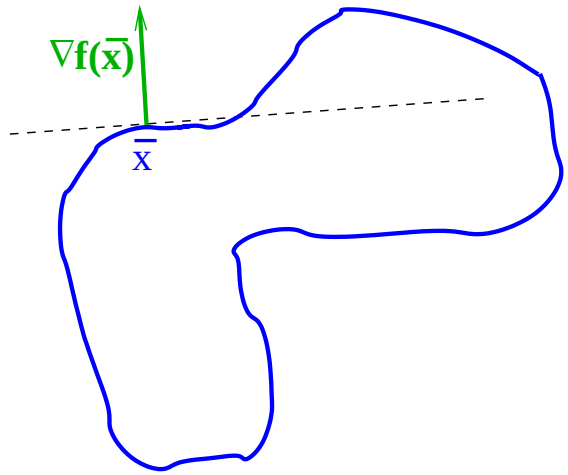
3.6 最急降下法

- 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．

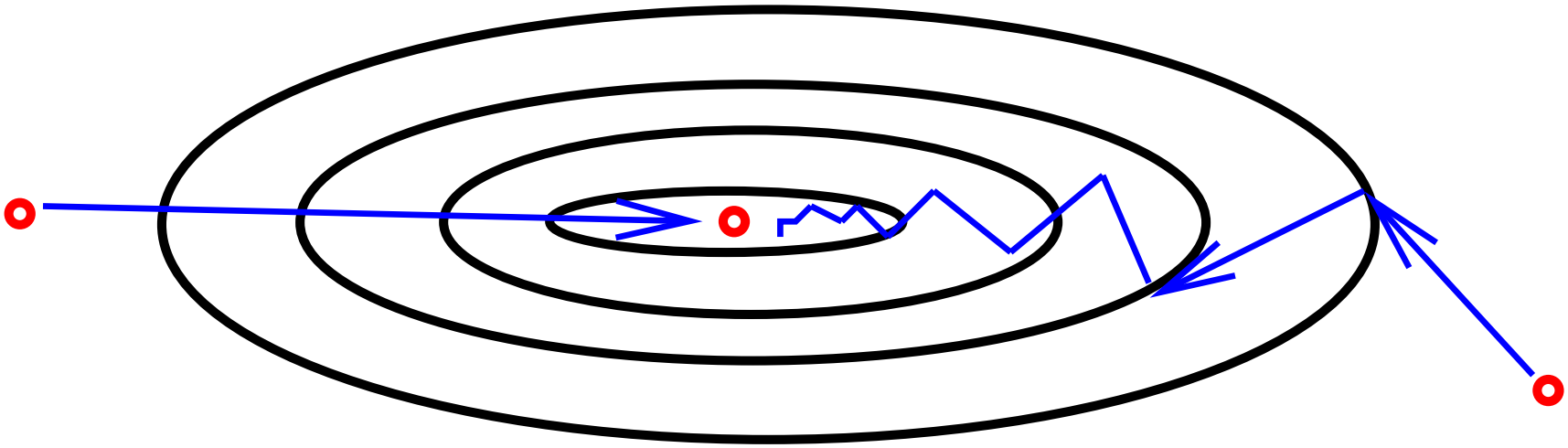
Step 3: $\|x^{k+1} - x^k\| < \epsilon$ ならば終了．そうでなければ $k := k + 1$ として Step 1 へ．

＜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．

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

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

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 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

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

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．

＜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，$1 > \alpha_0 \geq \frac{-\mu+\sqrt{\mu^2+\mu L}}{L}$ 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.

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

$$H_{k+1} := H_k + \frac{(y^k - H_k s^k)(y^k - H_k s^k)^T}{(y^k - H_k s^k)^T s^k}$$

DFP (Davidon-Fletcher-Powell) 公式

$$H_{k+1} := H_k + \frac{y^k (y^k)^T}{(s^k)^T y^k} - \frac{H_k s^k (s^k)^T H_k}{(s^k)^T H_k s^k}$$

BFGS (Broyden-Fletcher-Goldfarb-Shanno) 公式

$$H_{k+1} := \left(I - \frac{y^k (s^k)^T}{(y^k)^T s^k}\right) H_k \left(I - \frac{s^k (y^k)^T}{(y^k)^T s^k}\right) + \frac{y^k (y^k)^T}{(y^k)^T s^k} \tag{3.3}$$

- 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．
- 狭義凸２次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．

$$V_k = I - \frac{s^k (y^k)^T}{(y^k)^T s^k}$$

とおくと BFGS 公式 (3.3) は

$$H_k = V_{k-1}^T H_{k-1} V_{k-1} + \frac{y^{k-1}(y^{k-1})^T}{(y^{k-1})^T s^{k-1}}$$

と書き換えることができる．さらにこれを逐次的に代入していくと次の式が得られる．

$$\boxed{\begin{aligned} H_k \;=\; & (V_0 \cdots V_{k-1})^T H_0 (V_0 \cdots V_{k-1}) + (V_1 \cdots V_{k-1})^T \frac{y^0 (y^0)^T}{(y^0)^T s^0} (V_1 \cdots V_{k-1}) \\ & + \cdots + V_{k-1}^T \frac{y^{k-2}(y^{k-2})^T}{(y^{k-2})^T s^{k-2}} V_{k-1} + \frac{y^{k-1}(y^{k-1})^T}{(y^{k-1})^T s^{k-1}} \end{aligned}}$$

反復が進むに連れ，最適解に近づくことを考えると，初期の反復から得れた情報をある程度無視してもよい気がする．それで、最新の q 反復の情報 $V_{k-1}, \cdots, V_{k-q}$ だけを考えると，上記の式は次のようになる．

$$
\begin{aligned}
\widetilde{\boldsymbol{H}_k} \;=\; & (\boldsymbol{V}_{k-q}\cdots\boldsymbol{V}_{k-1})^{\boldsymbol{T}}\boldsymbol{H}_k^0(\boldsymbol{V}_{k-q}\cdots\boldsymbol{V}_{k-1}) \\
& +(\boldsymbol{V}_{k-q+1}\cdots\boldsymbol{V}_{k-1})^{\boldsymbol{T}}\frac{\boldsymbol{y}^{k-q}(\boldsymbol{y}^{k-q})^{\boldsymbol{T}}}{(\boldsymbol{s}^{k-q})^{\boldsymbol{T}}\boldsymbol{y}^{k-q}}(\boldsymbol{V}_{k-q+1}\cdots\boldsymbol{V}_{k-1}) \\
& +\cdots+\boldsymbol{V}_{k-1}^{\boldsymbol{T}}\frac{\boldsymbol{y}^{k-2}(\boldsymbol{y}^{k-2})^{\boldsymbol{T}}}{(\boldsymbol{y}^{k-2})^{\boldsymbol{T}}\boldsymbol{s}^{k-2}}\boldsymbol{V}_{k-1}+\frac{\boldsymbol{y}^{k-1}(\boldsymbol{y}^{k-1})^{\boldsymbol{T}}}{(\boldsymbol{y}^{k-1})^{\boldsymbol{T}}\boldsymbol{s}^{k-1}}
\end{aligned}
$$

ただし，$\boldsymbol{H}_k^0$は疎性を持った行列とする．

さらに，実質的には

$$\boxed{\boldsymbol{H}_k\boldsymbol{\nabla}\boldsymbol{f}(\boldsymbol{x}^k)} \text{ もしくは } \boxed{\widetilde{\boldsymbol{H}_k}\boldsymbol{\nabla}\boldsymbol{f}(\boldsymbol{x}^k)}$$

が必要なので$\boldsymbol{H}_k$もしくは$\widetilde{\boldsymbol{H}_k}$において右からベクトル$\boldsymbol{\nabla}\boldsymbol{f}(\boldsymbol{x}^k)$をかけるとほとんど全ての演算がベクトルの内積で出来ることがわかる．

＜記憶制限準ニュートン法 (L-BFGS)：アルゴリズム＞

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

Step 0': H_k^0 を与える (例えば疎行列 $((y^{k-1})^T s^{k-1})/((s^{k-1})^T s^{k-1})I$ とする)．

Step 1: $k \leq q$ の場合は $d^k := -H_k \nabla f(x^k)$，そうでなければ $d^k := -\widetilde{H_k} \nabla f(x^k)$ を計算する．

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

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

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: $k := k + 1$ として，Step 0' へ．

- 特に，$q = k$ の場合，記憶制限準ニュートン法は準ニュートン法になる．
- $q = 1$ の場合，次に紹介する共役勾配法になる．

＜利点＞

- ヘッセ行列やその逆行列を必要としなし，保持する必要もない．
- 大規模な問題にも適用できる．

＜欠点＞

- ニュートン法よりも数値的な安定性に欠ける．特にヘッセ行列の固有値が幅広く分布されている場合など．

3.11 共役勾配法

＜基本的な考え方＞

- ニュートン法や準ニュートン法では陽にヘッセ行列を保持する必要があるため，大規模な問題が解けない．
- ヘッセ行列やその逆行列を保持しないで行列ベクトル乗算のみで済ませる．

＜共役勾配法：アルゴリズム＞

Step 0: x^0を初期点，$f(x^0)$，$\nabla f(x^0)$ を計算し，$p^0 := -\nabla f(x^0)$，$\epsilon > 0$を停止基準，$k := 0$を反復回数とする．

Step 1: $\alpha \geq 0$に対する以下の直線探索を近似的に解く．

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

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

Step 2: $\|x^{k+1} - x^k\| < \epsilon$ならば終了．そうでなければ$k := k + 1$とする．

Step 3: $f(x^k)$，$\nabla f(x^k)$を計算する．

Step 4: β_kを次の公式から求める．

Step 4: $p^k := -\nabla f(x^k) + \beta_k p^{k-1}$としてStep 1へ．

Hestenes-Stiefel公式 (1952)

$$\beta_k := \frac{\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1}))}{(p^{k-1})^T(\nabla f(x^k) - \nabla f(x^{k-1}))}$$

Fletcher-Reeves 公式 (1964)

$$\beta_k := \frac{\|\nabla f(x^k)\|^2}{\|\nabla f(x^{k-1})\|^2}$$

Polak-Ribière 公式

$$\beta_k := \frac{\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1}))}{\|\nabla f(x^{k-1})\|^2}$$

Polak-Ribière plus 公式

$$\beta_k := \max\left\{0, \frac{\nabla f(x^k)^T(\nabla f(x^k) - \nabla f(x^{k-1}))}{\|\nabla f(x^{k-1})\|^2}\right\}$$

Dai-Yuan 公式 (1999)

$$\beta_k := \frac{\|\nabla f(x^k)^T\|^2}{(p^{k-1})^T(\nabla f(x^k) - \nabla f(x^{k-1}))}$$

＜利点＞

- ヘッセ行列を必要としないし，その保持の必要もない．

- 狭義凸２次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．

3.12 演習問題

問題１

２次関数補間法にて，2点$\alpha_1, \alpha_2 \in \mathbb{R}$が与えられていて，関数$\phi(\alpha)$を２次関数$q(\alpha) = a\alpha^2 + b\alpha + c$で補間するとする．この時，$q(\alpha_1) = \phi(\alpha_1)$, $q(\alpha_2) = \phi(\alpha_2)$, $q'(\alpha_1) = \phi'(\alpha_1)$となるように，関数$q(\alpha)$の最小解$\bar{\alpha}$を求めよ．

問題２

２次関数補間法にて，2点$\alpha_1, \alpha_2 \in \mathbb{R}$が与えられていて，関数$\phi(\alpha)$を２次関数$q(\alpha) = a\alpha^2 + b\alpha + c$で補間するとする．この時，$q(\alpha_1) = \phi(\alpha_1)$, $q'(\alpha_1) = \phi'(\alpha_1)$, $q'(\alpha_2) = \phi'(\alpha_2)$となるように，関数$q(\alpha)$の最小解$\bar{\alpha}$を求めよ．

問題３

定理４の結果から$x^0 \in \mathbb{R}^n$を初期点とした厳密直線探索を用いた最急降下法が$f(x)$に適用された時，$\|x^k - x^*\| < \epsilon$を得るには反復回数kが$\mathcal{O}(\ln(1/\epsilon))$必要であることを示せ．

問題４

定理６の仮定のもと，$x^0 \in \mathbb{R}^n$ を初期点としてニュートン法を適用して求まる点列にて，$\|x^k - x^*\| < \epsilon$ を得るには反復回数 k が $\mathcal{O}(\ln\ln(1/\epsilon))$ 必要であることを示せ．

問題５

準ニュートン法において，対称ランク１公式，DFP公式，BFGS公式がそれぞれ $H_{k+1}s^k = y^k$ を満たすことを示せ．

問題６

$u, v \in \mathbb{R}^n$，正則行列 $M \in \mathbb{R}^{n\times n}$ に対して，$1 + v^T M^{-1} u \neq 0$ ならば，

$$(M + uv^T)^{-1} = M^{-1} - \frac{M^{-1}uv^TM^{-1}}{1 + v^TM^{-1}u}$$

（Sherman-Morrison公式）が成り立つ．この公式を使い，対称ランク１公式， DFP公式，BFGS公式から H_{k+1} の逆行列をそれぞれ求めよ．

問題７

$f(x) = x^TQx/2 + q^Tx$ として定義された狭義凸２次関数を考える．ただし，Q は正定値対称行列とする．$f(x)$ に共役勾配法を適用した時，Fletcher-Reeves公式，Polak-Ribière公式， Hestenes-Stiefel公式，Dai-Yuan公式から求まる係数 β_k が一致することを示せ．

問題８

再度，共役勾配法を狭義凸２次関数 $f(x) = x^TQx/2 + q^Tx$ に対して適用することを考える．ただし，Q は正定値対称行列とする．この時，共役勾配法で生成される p^k $(k = 0, 1, \ldots)$ が $(p^i)^TQp^j = 0$ $(i, j = 0, 1, \ldots)$ を満すことを示せ．

Chapter 4

非線形計画問題, Nonlinear Programming, NLP

4.1 定式化，用語

(制約付き) 非線形計画問題

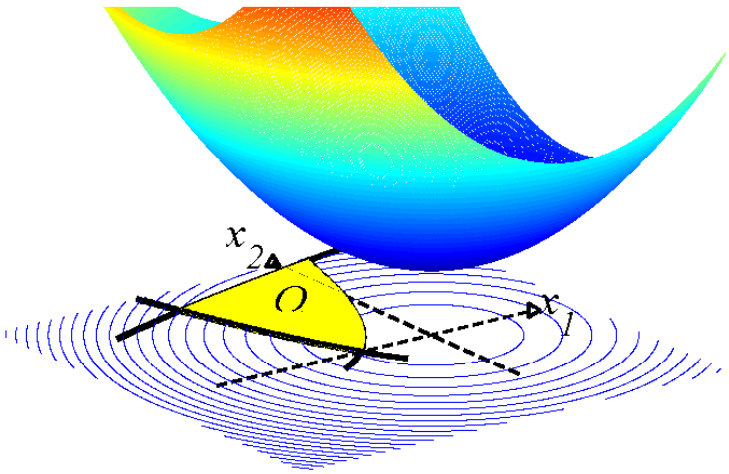
$$\boxed{\begin{cases} \text{最小化：} & f(x) \\ \text{条件：} & g_i(x) \leq 0, \quad i = 1, 2, \ldots, \ell \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, m \end{cases}}$$

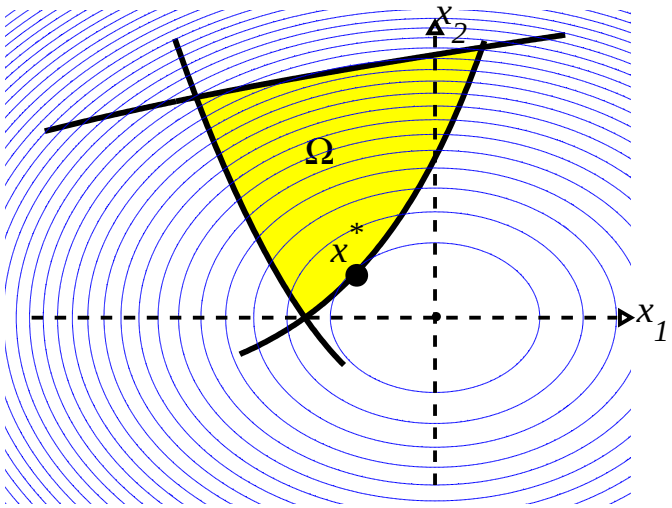
ただし，$f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $h_j : \mathbb{R}^n \longrightarrow \mathbb{R}$

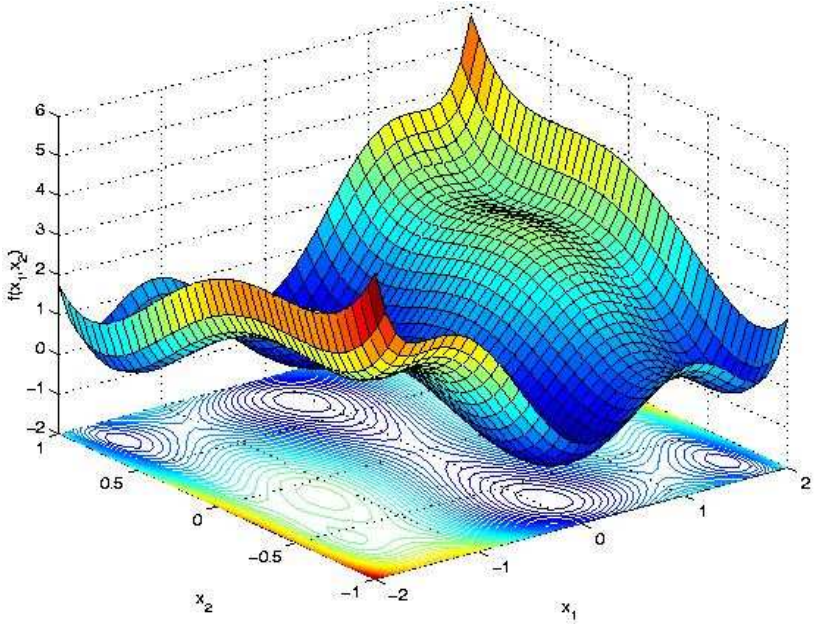
(制約付き) 非線形計画問題

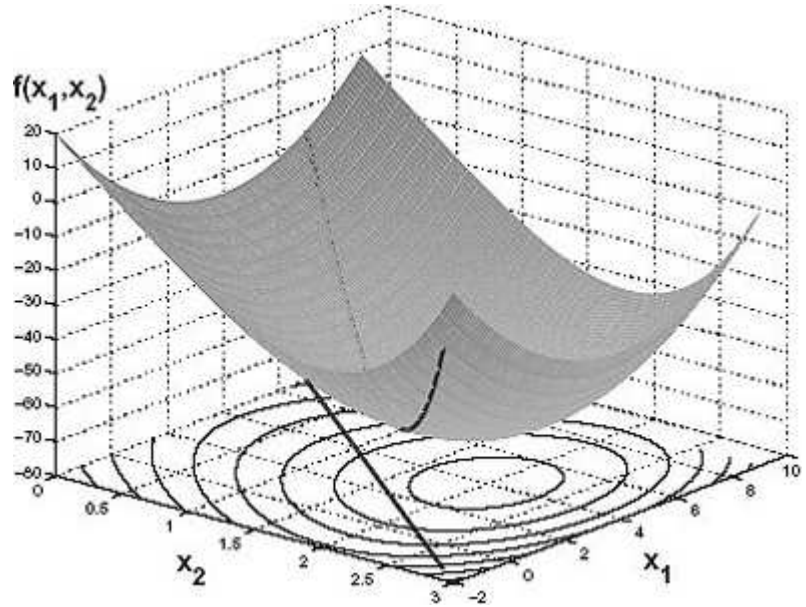
$$\begin{cases} \text{最小化：} & f(x) \\ \text{条件：} & g_i(x) \leq 0, \quad i = 1, 2, \ldots, \ell \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, m \end{cases}$$

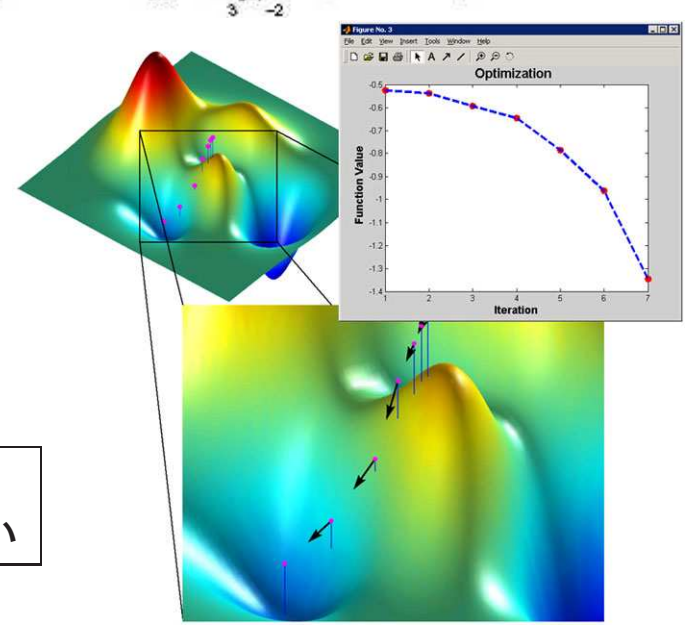
ただし， $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $h_j : \mathbb{R}^n \longrightarrow \mathbb{R}$











- 大域的最適解と大域的最適値 ⇒ 比較的難しい
- 局所的最適解と局所的最適値 ⇒ 変数が多いと局所的な情報しかない

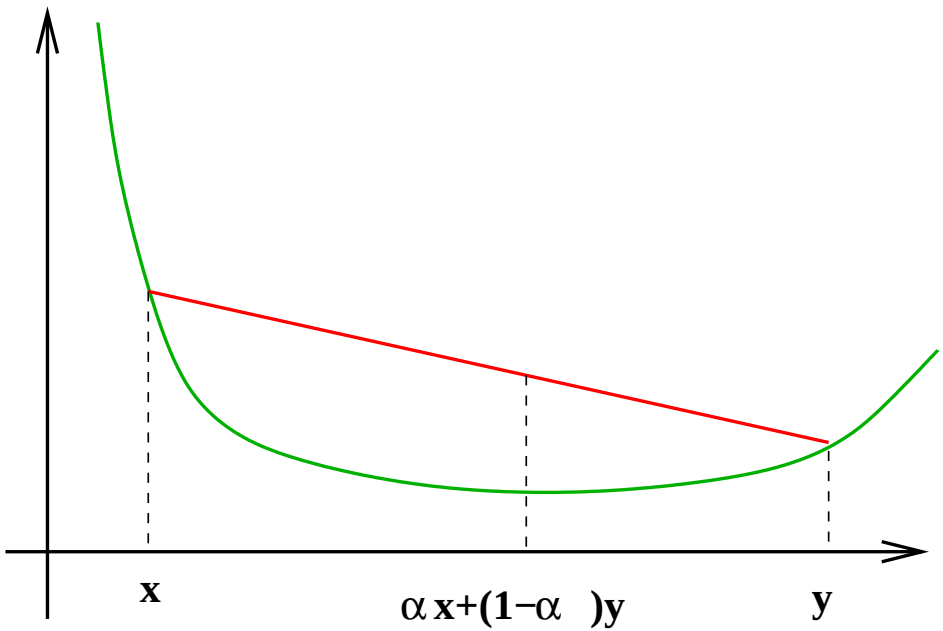
4.2 凸関数

定義 1

関数 $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ は以下の条件を満たしていれば凸関数と呼ばれる.

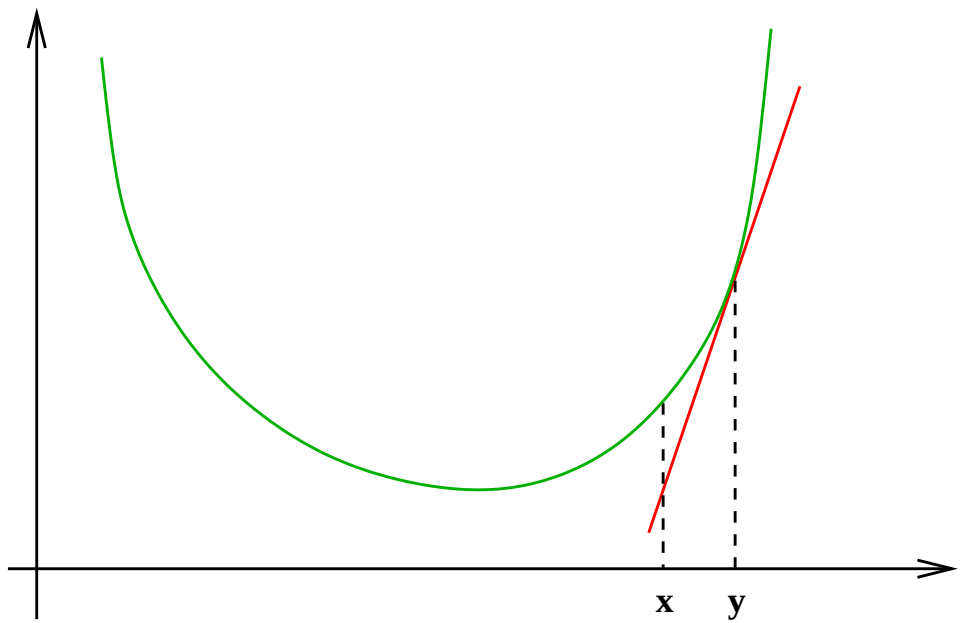
$$\forall x, y \in \mathbb{R}^n, \quad \forall \alpha \in [0, 1]$$

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$



命題２

定義域$\mathbb{R}^n$で定義された微分可能な関数fが凸関数である必要十分条件は

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \langle \nabla f(\boldsymbol{y}), (\boldsymbol{x} - \boldsymbol{y}) \rangle$$