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

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

- 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: $\boldsymbol{x}^0$ を初期点，$\boldsymbol{H}_0 := \boldsymbol{I}$，$\epsilon > 0$ を停止基準，$k := 0$ を反復回数とする．

Step 1: $\boldsymbol{d}^k := -\boldsymbol{H}_k \nabla f(\boldsymbol{x}^k)$ 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．

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

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

Step 3: $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\| < \epsilon$ ならば終了．$\boldsymbol{y}^k, \boldsymbol{s}^k$ を求める．

Step 4: $\boldsymbol{H}_{k+1}$ 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 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．

Fletcher-Reeves 公式

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

Polak-Ribière公式

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

Hestenes-Stiefel公式

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

Dai-Yuan公式

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

＜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．

3.11 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 $H_{k+1}s^k =$

$\boldsymbol{y}^k$を満たすことを示せ．

問題 6

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

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

（**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 $\boldsymbol{H}_{k+1}$ の逆行列をそれぞれ求めよ．

問題 7

$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}/2 + \boldsymbol{q}^T \boldsymbol{x}$ として定義された凸 2 次関数を考える．ただし，$\boldsymbol{Q}$ は正定値対称行列とする．$\boldsymbol{f}(\boldsymbol{x})$ 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 $\boldsymbol{\beta}_k$ が一致することを示せ．

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

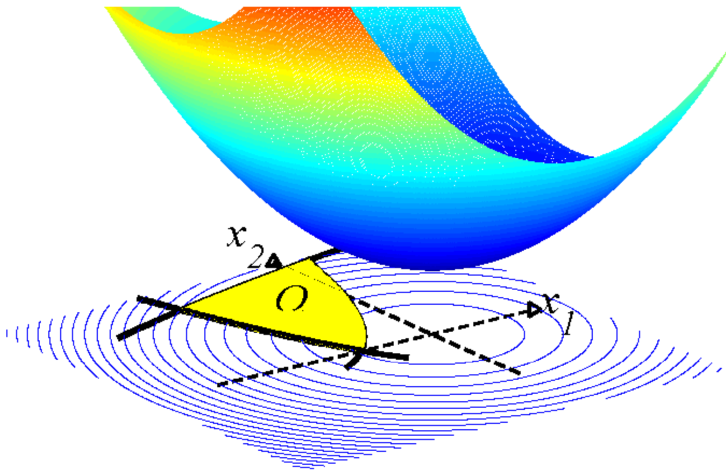
$$\boxed{\begin{cases} \text{最小化：} & f(x) \\ \text{条件：} & g_i(x) \le 0, \quad i = 1, 2, \ldots, \ell \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, m \end{cases}}$$

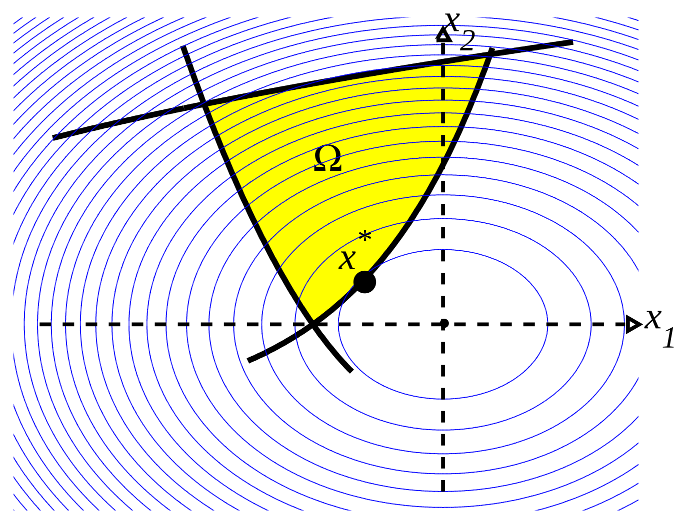
ただし，$f : \mathbb{R}^n \longrightarrow \mathbb{R}, \; g_i : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad h_j : \mathbb{R}^n \longrightarrow \mathbb{R}$

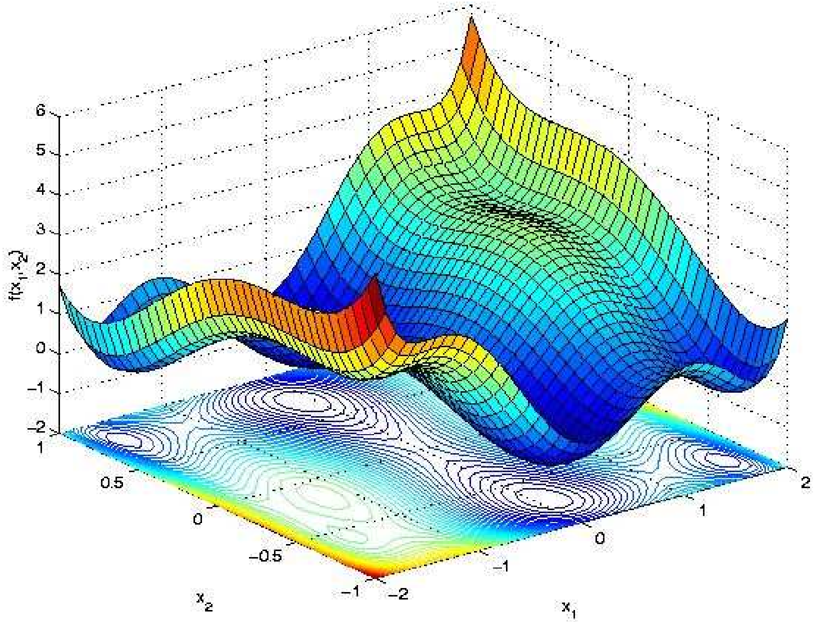
(制約付き) 非線形計画問題

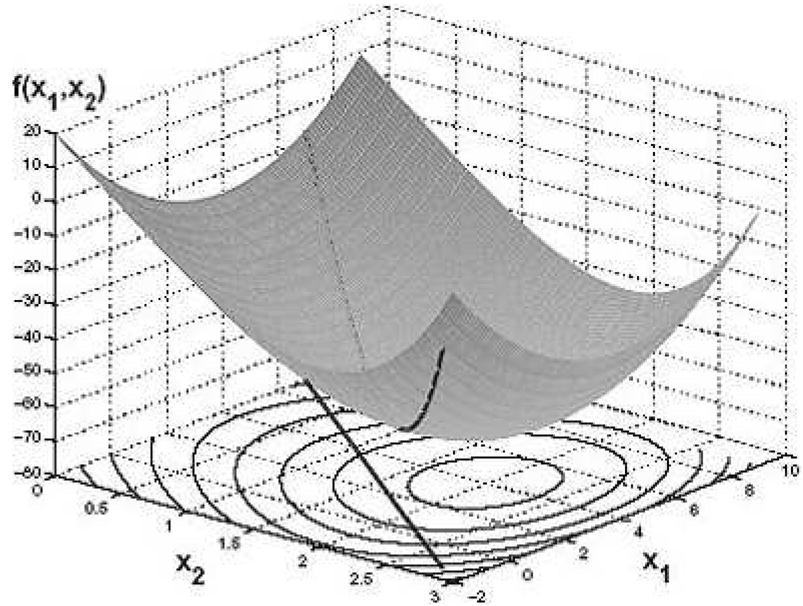
$$
\begin{cases}
\text{最小化：} & \boldsymbol{f(x)} \\
\text{条件：} & \boldsymbol{g_i(x) \leq 0, \;\; i = 1, 2, \ldots, \ell} \\
 & \boldsymbol{h_j(x) = 0, \;\; j = 1, 2, \ldots, m}
\end{cases}
$$

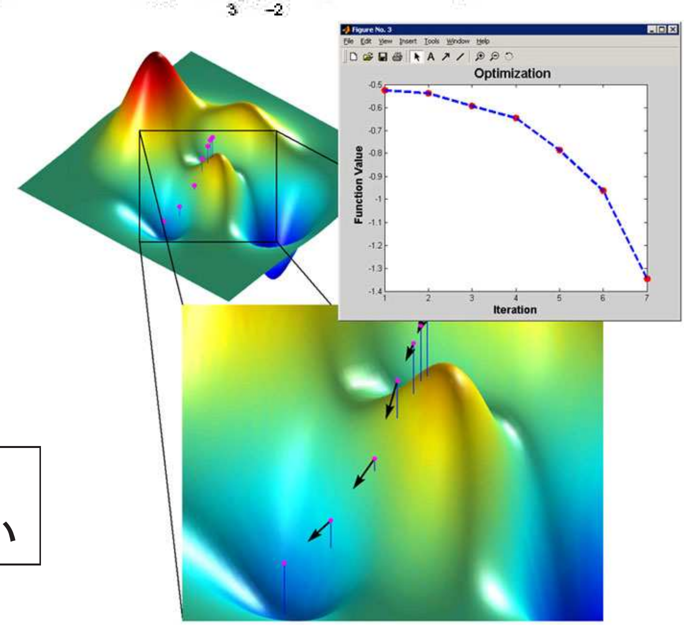
ただし， $\boldsymbol{f} : \mathbb{R}^n \longrightarrow \mathbb{R}$, $\boldsymbol{g_i} : \mathbb{R}^n \longrightarrow \mathbb{R}$, $\boldsymbol{h_j} : \mathbb{R}^n \longrightarrow \mathbb{R}$











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

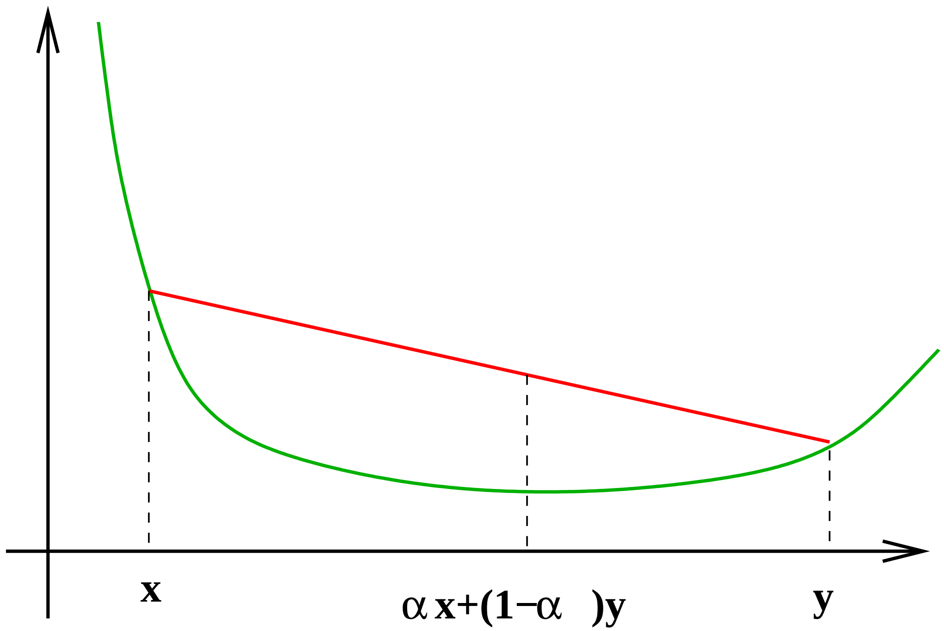
4.2 凸関数

<u>定義 1</u>

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.

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

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

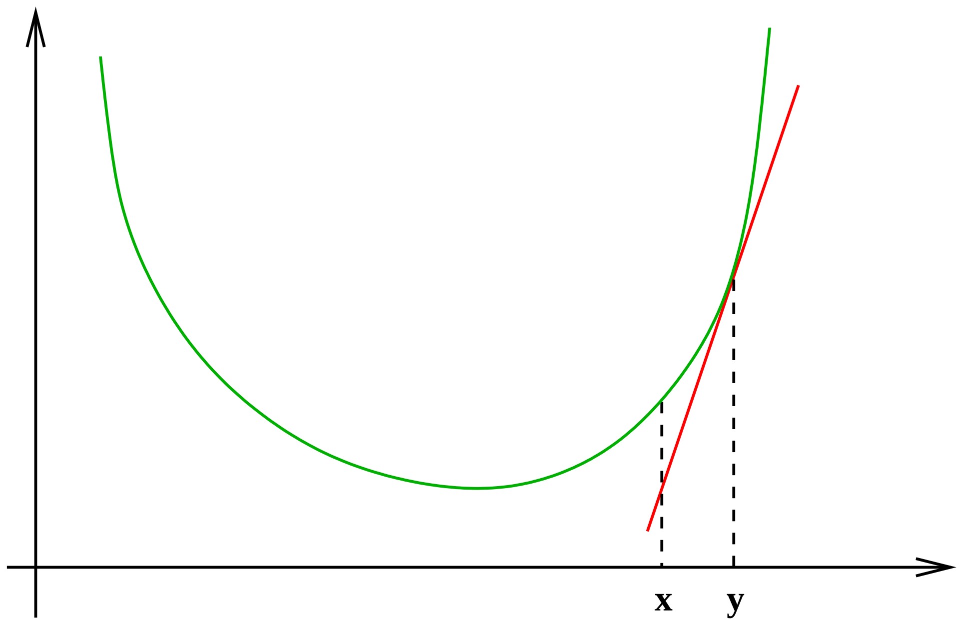


命題 2

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

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

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



4.3　最適性条件

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

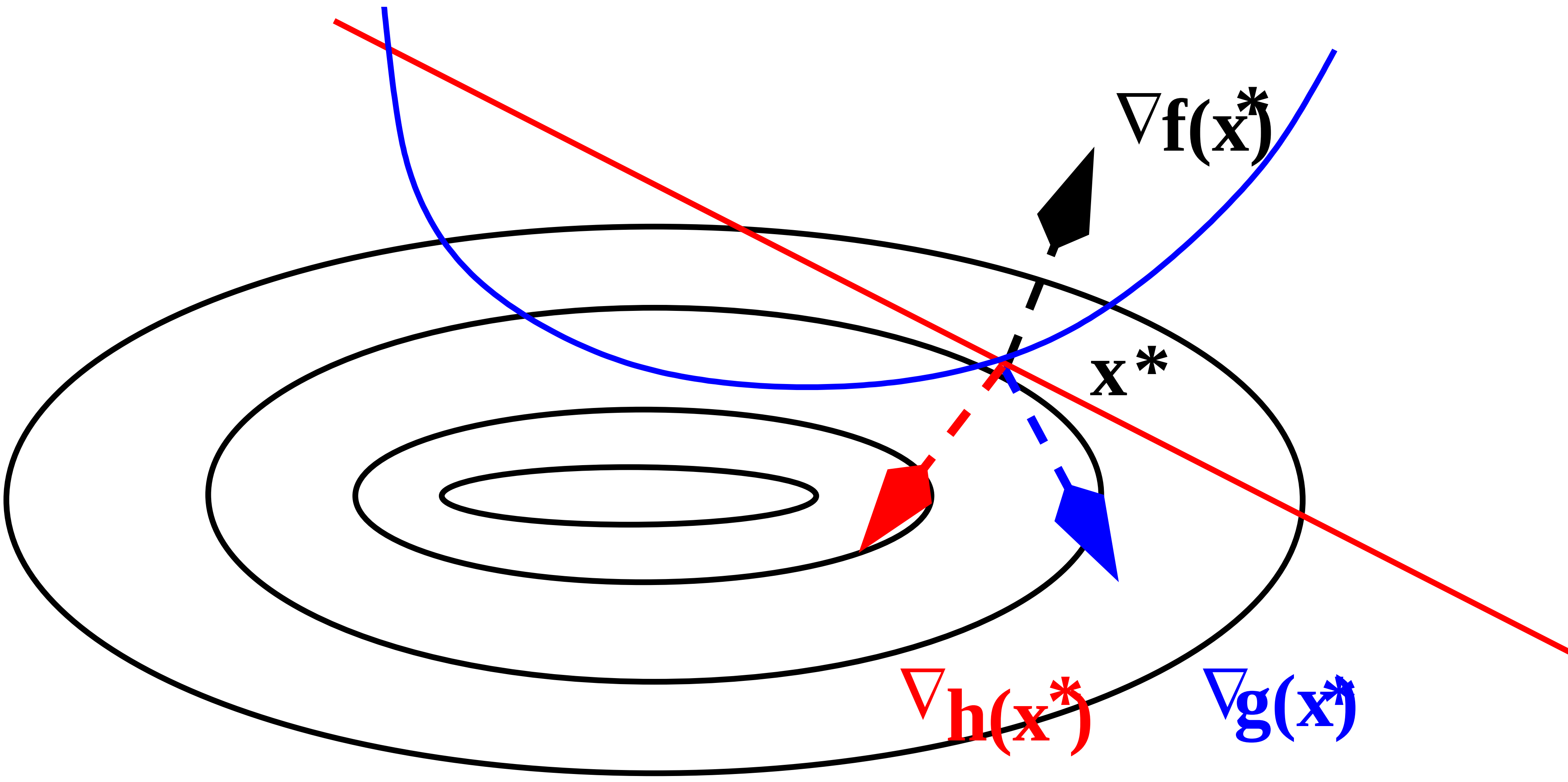
仮定

f, g_i, h_j は定義域 $\mathbb{R}^n$ にて微分可能な関数である

Karush-Kuhn-Tucker の最適性条件

次の条件を満たす $x^* \in \mathbb{R}^n$ と $(y^*, w^*) \in \mathbb{R}^\ell \times \mathbb{R}^m$ が存在する

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{\ell} y_i^* \nabla g_i(x^*) + \sum_{j=1}^{m} w_j^* \nabla h_j(x^*) &= 0 \\ g_i(x^*) \leq 0, \;\; i &= 1, 2, \ldots, \ell \\ h_j(x^*) = 0, \;\; j &= 1, 2, \ldots, m \\ y_i^* g_i(x^*) = 0, \;\; i &= 1, 2, \ldots, \ell \\ y^* &\geq 0 \end{aligned} \tag{4.1}$$


∇f(x*)
x*
∇h(x*)
∇g(x*)

(制約付き) 非線形計画問題

$$\begin{cases} \text{最小化：} & f(x) \\ \text{条件：} & g_i(x) \le 0, \quad i = 1, 2, \ldots, \ell \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, m \end{cases} \tag{4.2}$$

x^* が (4.2) の局所的最適解である

必要条件 $\Downarrow$ $\Uparrow$ 十分条件

次の条件を満たす (y^*, w^*) が存在する

Karush-Kuhn-Tucker の最適性条件

$$\nabla f(x^*) + \sum_{i=1}^{\ell} y_i^* \nabla g_i(x^*) + \sum_{j=1}^{m} w_j^* \nabla h_j(x^*) = 0$$

$$\begin{aligned} g_i(x^*) &\le 0, \quad i = 1, 2, \ldots, \ell \\ h_j(x^*) &= 0, \quad j = 1, 2, \ldots, m \\ y_i^* g_i(x^*) &= 0, \quad i = 1, 2, \ldots, \ell \\ y^* &\ge 0 \end{aligned}$$

4.3.1 最適性十分条件

定理3 (Karush-Kuhn-Tucker の最適性十分条件)

x^*, (y^*, w^*) が Karush-Kuhn-Tucker の最適性条件を満たしているとする．f, g_i, $i = 1, 2, \ldots, \ell$ が凸関数であり，h_j, $j = 1, 2, \ldots, m$ が線形関数だとすると，x^* は非線形計画問題 (4.2) の局所的最適解になる．

証明

[矢部2006] などを参照． □

定義4

非線形計画問題 (4.2) の不等式制約 g_i, $i = 1, 2, \ldots, \ell$ とある点 x^* に対して，$g_k(x^*) = 0$ が成り立つとき，この制約式は点 x^* で効いているといい，この制約式は x^* で有効制約式であるという．また $g_k(x^*) < 0$ のとき，この制約式は点 x^* で効いていないという．

x^* に対して有効な不等式制約の添字集合を $I(x^*)$ で表わす．

定理 5（2 次の最適性十分条件）

$\boldsymbol{x}^*$，$(\boldsymbol{y}^*, \boldsymbol{w}^*)$ が Karush-Kuhn-Tucker の最適性条件を満たしているとする．さらに不等式制約の中で有効制約式 $g_i(\boldsymbol{x}^*) = 0$，$i \in \boldsymbol{I}(\boldsymbol{x}^*)$ に対して $y_i^* > 0$ であり，かつ

$$\boldsymbol{u}^T \left(\nabla^2 f(\boldsymbol{x}^*) + \sum_{i \in \boldsymbol{I}(\boldsymbol{x}^*)} y_i^* \nabla^2 g_i(\boldsymbol{x}^*) + \sum_{j=1}^{m} w_j^* \nabla^2 h_j(\boldsymbol{x}^*) \right) \boldsymbol{u} > 0,$$

が

$$\nabla g_i(\boldsymbol{x}^*)^T \boldsymbol{u} = 0,\ i \in \boldsymbol{I}(\boldsymbol{x}^*), \quad \nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{u} = 0,\ j = 1, 2, \ldots, m$$

を満たす全ての $\boldsymbol{u} \neq \boldsymbol{0}$ に対していえるのならば，$\boldsymbol{x}^*$ は非線形計画問題 (4.2) の狭義局所的最適解である．

証明

[BERTSEKAS2003] など参照． $\square$

4.3.2 最適性必要条件

定義 6

非線形計画問題 (4.2) の実行可能解 x^*，つまり不等式制約にて $g_i(x^*) \leq 0,\ i = 1,2,\ldots,\ell,$ 等式制約にて $h_j(x^*) = 0,\ j = 1,2,\ldots,m$ を満たす点が線形独立制約想定を満たしているとは不等式制約の中で有効な制約式 $\nabla g_i(x^*),\ i \in I$ と等式制約 $\nabla h_j(x^*),\ j = 1,2,\ldots,m$ が線形独立であるときをいう．

定理 7 (Karush-Kuhn-Tucker の最適性必要条件)

x^* が非線形計画問題 (4.2) の局所的最適解だとする．さらに x^* が線形独立制約想定を満たしていると仮定する．すると Karush-Kuhn-Tucker の最適性条件を満たす $(y^*, w^*) \in \mathbb{R}^\ell \times \mathbb{R}^m$ が存在する．

証明

[矢部 2006] などを参照． □

定理８（2 次の最適性必要条件）

$\boldsymbol{x}^*$ が非線形計画問題 (4.2) の局所的最適解だとする．さらに $\boldsymbol{x}^*$ が線形独立制約想定を満たしていると仮定する．すると **Karush-Kuhn-Tucker** の最適性条件を満たす $(\boldsymbol{y}^*, \boldsymbol{w}^*) \in \mathbb{R}^\ell \times \mathbb{R}^m$ が存在し，

$$\boldsymbol{u}^T \left(\nabla^2 f(\boldsymbol{x}^*) + \sum_{i \in \boldsymbol{I}(\boldsymbol{x}^*)} y_i^* \nabla^2 g_i(\boldsymbol{x}^*) + \sum_{j=1}^{m} w_j^* \nabla^2 h_j(\boldsymbol{x}^*) \right) \boldsymbol{u} > 0,$$

が次の条件を満たす $\boldsymbol{u} \neq \boldsymbol{0}$ に対していえる

$$\nabla g_i(\boldsymbol{x}^*)^T \boldsymbol{u} = 0, \; i \in \boldsymbol{I}(\boldsymbol{x}^*), \quad \nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{u} = 0, \; j = 1, 2, \ldots, m.$$

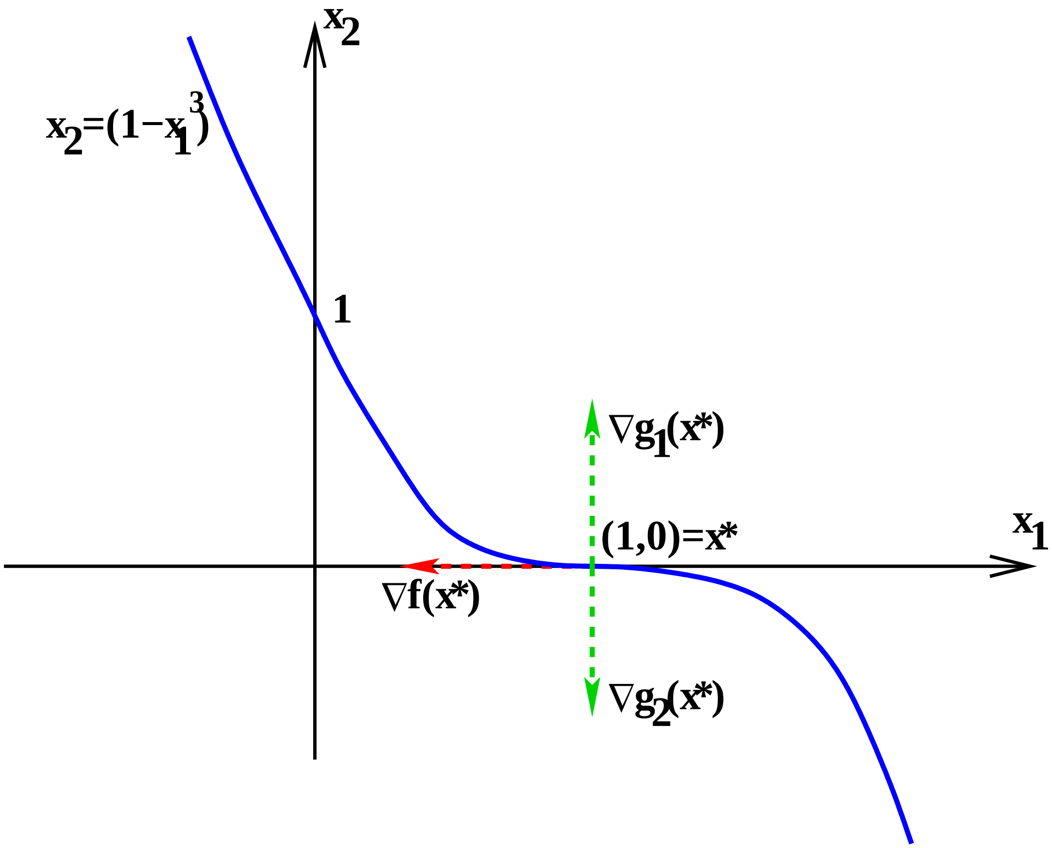
証明

[BERTSEKAS2003] など参照． □

次の最適化問題を考える

$$\begin{cases} \text{最小化：} & -x_1 \\ \text{条件：} & -(1-x_1)^3 + x_2 \leq 0 \\ & -x_2 \leq 0 \end{cases}$$

最適解は $(1,0)$ であり，最適値は -1 である．



(制約付き) 非線形計画問題

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

x^* が (4.2) の局所的最適解である

線形独立制約想定等の想定　必要条件 $\Downarrow$　　　$\Uparrow$ 十分条件　関数の凸性

次の条件を満たす (y^*, w^*) が存在する

Karush-Kuhn-Tucker の最適性条件

$$
\nabla f(x^*) + \sum_{i=1}^{\ell} y_i^* \nabla g_i(x^*) + \sum_{j=1}^{m} w_j^* \nabla h_j(x^*) = 0
$$

$$
\begin{aligned}
g_i(x^*) &\le 0, \quad i = 1, 2, \ldots, \ell \\
h_j(x^*) &= 0, \quad j = 1, 2, \ldots, m \\
y_i^* g_i(x^*) &= 0, \quad i = 1, 2, \ldots, \ell \\
y^* &\ge 0
\end{aligned}
$$

4.4 数値解法

- 数値解法では局所的最適解を求めることが目的
- 大域的最適解を求めることは一般的に難しい

数値解法例

- 拡張 **Lagrange** 法
- 逐次 **2** 次計画法
- 信頼領域法
- 主双対内点法
- フィルター法

4.4.1 拡張 Lagrange 法

- まずは等式制約付き非線形計画問題について考える

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

$\mu > 0$ をパラメータとした**拡張 Lagrange 関数**を定義する

$$\boldsymbol{\hat{\mathcal{L}}(x, w, \mu) \equiv f(x) + \sum_{j=1}^{m} w_j h_j(x) + \frac{\mu}{2}\sum_{j=1}^{m} h_j^2(x)} \tag{4.3}$$

- $\mu \longrightarrow \infty$ によって等式制約がより強く課せられる
- 無制約非線形計画問題として考え，x に対して微分をし，最適性条件を考える

$$\boldsymbol{\nabla_x \hat{\mathcal{L}}(x, w, \mu) = \nabla f(x) + \sum_{j=1}^{m}(w_j + \mu h_j(x))\nabla h_j(x) = 0}$$

さらに等式制約付き非線形計画問題に対する **KKT** の最適性条件 **(4.1)** と比較すると

$$\boldsymbol{w_j^* \approx w_j + \mu h_j(x)}$$

となることが推測できる．よって，この方法を反復的に行なって **(4.3)** を満たすような (x^*, w^*) を求めようとすると，

$$w_j^{k+1} := w_j^k + \mu^k h_j(x^k)$$

と **Lagrange 乗数**を更新することが自然である．

拡張 Lagrange 法

Step 0: 初期点 x^0，w^0 を決め，十分大きい $\mu^0 > 0$，$k := 0$ とする．

終了条件: もし x^k が局所的最適解の近似になっていると判断できたならば (w^k とともに KKT の最適性条件を満たしているならば) 終了する．

Step 1: (無制約) 非線形計画問題に対する手法を用いて $\hat{\mathcal{L}}(x, w^k, \mu^k)$ を x^k から出発して最小化することを試みて x^{k+1} を求める．ただし，w^k, μ^k は固定定数とみなす．

Step 2: Lagrange 乗数を $w_j^{k+1} := w_j^k + \mu^k h_j(x^{k+1})$ として更新する．

Step 3: $\mu^{k+1} > \mu^k$ を決め，$k := k + 1$ とし，終了条件を調べる．

定理 9

等式制約付き非線形計画問題の局所的最適解 $\boldsymbol{x}^*$ とそれに対応する Lagrange 乗数 $\boldsymbol{w}^*$ が 2 次の最適性十分条件を満たしていると仮定する (定理 5 参照)．さらに局所的最適解 $\boldsymbol{x}^*$ に対する線形独立制約想定が成り立っていると仮定する．この時，十分大きな $\mu^* > 0$ が存在して，$\mu > \mu^*$ に対して $\nabla_x^2 \hat{\mathcal{L}}(\boldsymbol{x}^*, \boldsymbol{w}^*, \mu)$ は正定値行列になる．

証明

[矢部 2006] などを参照． □

- 不等式制約付き非線形計画問題を考える

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

- この場合，スラック変数 $\mathbb{R}^\ell \ni \boldsymbol{s} > \boldsymbol{0}$ を導入して等式制約付き非線形計画問題に変換する

$$\begin{cases} \text{最小化：} & f(\boldsymbol{x}) \\ \text{条件：} & g_i(\boldsymbol{x}) + s_i = 0, \ \ i = 1, 2, \ldots, \ell \\ & s_i \geq 0, \qquad\qquad\ \ i = 1, 2, \ldots, \ell \end{cases}$$

- 等式制約問題なので，もとの等式制約の記述を用い，正確には箱型制約と等式制約付き非線形計画問題を扱う．ただし，$\mathbb{R}^n \ni \boldsymbol{b} \leq \boldsymbol{u} \in \mathbb{R}^n$ は変数 $\boldsymbol{x}$ が取り得る値の範囲である

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

上記の問題に対する拡張 Lagrange 関数は

$$\hat{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{w}, \mu) = f(\boldsymbol{x}) + \sum_{j=1}^{m} w_j h_j(\boldsymbol{x}) + \frac{\mu}{2} \sum_{j=1}^{m} h_j^2(\boldsymbol{x})$$

であり，拡張 Lagrange 法の部分問題では次の最適化問題を考える

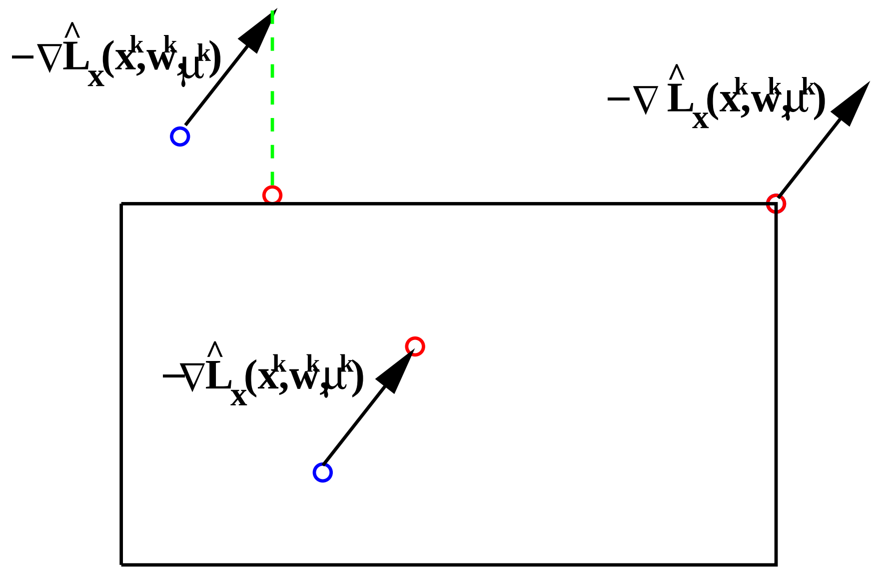
$$\boxed{\begin{cases} \text{最小化：} & \hat{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{w}, \mu) = f(\boldsymbol{x}) + \sum_{j=1}^{m} w_j h_j(\boldsymbol{x}) + \frac{\mu}{2} \sum_{j=1}^{m} h_j^2(\boldsymbol{x}) \\ \text{条件：} & \boldsymbol{b} \leq \boldsymbol{x} \leq \boldsymbol{u} \end{cases}}$$

上記の問題は例えば勾配ベクトル射影法で解くことができる．つまり次のような関係式を満たす $\boldsymbol{x}^k$ を求めればよい．

$$x^k := \mathcal{P}(x^k - \nabla \hat{\mathcal{L}}_x(x^k, w^k, \mu^k); b, u)$$

ただし，$\mathcal{P}(v; b, u)$ はベクトル v を箱型制約に射影した点である．

$$\mathcal{P}(v; b, u) = \begin{cases} b_i, & \text{if } v_i \le b_i, \\ v_i, & \text{if } b_i \le v_i \le u_i, \quad i = 1, 2, \ldots, n \\ u_i, & \text{if } u_i \le v_i \end{cases}$$



よって上記のルーチンを拡張Lagrange法アルゴリズムのStep 1に組みこむことによって不等式制約付き非線形計画問題も解けるようになる.

4.4.2 逐次2次計画法

- まずは等式制約付き非線形計画問題について考える

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

上記の最適化問題に対するKKTの最適性条件(4.1)から次の方程式に対するNewton法を考える

$$F(x, w) = \begin{pmatrix} \nabla f(x) + \sum_{j=1}^{m} w_j \nabla h_j(x) \\ h(x) \end{pmatrix} = 0$$

つまり，(x^k, w^k)をk反復目の点だとしたら，次点は次の式で求まる

$$\begin{pmatrix} x^{k+1} \\ w^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ w^k \end{pmatrix} + \begin{pmatrix} \Delta x \\ \Delta w \end{pmatrix}, \quad \text{ただし} \quad \begin{pmatrix} \Delta x \\ \Delta w \end{pmatrix} = -[F'(x^k, w^k)]^{-1} F(x^k, w^k),$$

$$F'(x^k, w^k) = \begin{pmatrix} \nabla^2 f(x^k) + \sum_{j=1}^{m} w_j^k \nabla h_j^2(x^k) & \nabla h(x^k) \\ \nabla h^T(x^k) & O \end{pmatrix}$$

なので，上記の行列の逆行列は次の仮定が満たされていれば存在する[NOCEDAL2006]．

仮定

(1) $\nabla h_j(x^k),\ j=1,2,\ldots,m$は線形独立である.

(2) $u^T\left(\nabla^2 f(x^k)+\sum_{j=1}^m w_j^k\nabla^2 h_j(x^k)\right)u>0$, が次の条件を満たす全ての$u\neq 0$に対していえる$\nabla h_j(x^k)^T u=0,\ j=1,2,\ldots,m.$

これを展開すると

$$\begin{cases}\left[\nabla^2 f(x^k)+\sum_{j=1}^m w_j^k\nabla h_j^2(x^k)\right]\Delta x+\sum_{j=1}^m(\Delta w)_j\nabla h_j(x^k)=-\nabla f(x^k)-\sum_{j=1}^m w_j^k\nabla h_j(x^k)\\ \nabla h_j^T(x^k)\Delta x=-h_j(x^k),\quad j=1,2,\ldots,m\end{cases}$$

逐次2次計画法

Step 0: 初期点x^0，w^0を決め，$k=0$とする.

終了条件: もし x^kが局所的最適解の近似になっていると判断できたならば(w^kとともにKKTの最適性条件を満たしているならば)終了する.

Step 1: 上記の式より$(\Delta x,\Delta w)$を求める.

Step 2: $(x^{k+1},w^{k+1})=(x^k,w^k)+(\Delta x,\Delta w)$，$k=k+1$とし，終了条件を調べる.

- 探索方向$(\boldsymbol{\Delta x}, \boldsymbol{\Delta w})$はまた下の問題の目的関数と制約式の近似モデルに対する**KKT**の最適化条件としても見なすことができる

$$\begin{cases} \text{最小化：} & \boldsymbol{\mathcal{L}(x, w^k) = f(x) + \sum_{j=1}^{m} w_j^k h_j(x)} \\ \text{条件：} & \boldsymbol{h_j(x) = 0,} \qquad\qquad \boldsymbol{j = 1, 2, \ldots, m} \end{cases}$$

$$\Downarrow \quad \boldsymbol{x^k} \text{ においての近似}$$

$$\begin{cases} \text{最小化：} & \boldsymbol{\nabla_x \mathcal{L}(x^k, w^k)\Delta x + \frac{1}{2}\Delta x \nabla_x^2 \mathcal{L}(x^k, w^k)\Delta x} \\ \text{条件：} & \boldsymbol{h_j(x^k) + \nabla h_j^T(x^k)\Delta x = 0,} \qquad\qquad \boldsymbol{j = 1, 2, \ldots, m} \end{cases}$$

$$\Downarrow \quad \boldsymbol{\Delta x} \text{ に対する}\mathbf{KKT}\text{の最適性条件}$$

$$\begin{cases} \boldsymbol{\nabla_x^2 \mathcal{L}(x^k, w^k)\Delta x + \nabla_x \mathcal{L}(x^k, w^k) + \sum_{j=1}^{m} (\Delta w)_j \nabla h_j(x^k) = 0} \\ \boldsymbol{h_j(x^k) + \nabla h_j^T(x^k)\Delta x = 0,} \qquad\qquad \boldsymbol{j = 1, 2, \ldots, m} \end{cases}$$

- 一般の非線形計画問題を考える

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

- 同様に目的関数と制約式の近似モデルを考える

$$
\begin{cases}
\text{最小化：} & \nabla_x \mathcal{L}(x^k, w^k)\Delta x + \dfrac{1}{2}\Delta x \nabla_x^2 \mathcal{L}(x^k, w^k)\Delta x \\
\text{条件：} & g_i(x^k) + \nabla g_i^T(x^k)\Delta x \leq 0, \qquad i = 1, 2, \ldots, \ell \\
& h_j(x^k) + \nabla h_j^T(x^k)\Delta x = 0, \qquad j = 1, 2, \ldots, m
\end{cases}
$$

- 上記の問題は変数 Δx に対する凸 2 次計画問題なので，単体法や内点法で簡単に解くことができる

- 以上のルーチンを逐次 2 次計画法の **Step 1** に組み込むことによって一般の非線形計画問題に対しても対応できる

- $\nabla_x^2 \mathcal{L}(x^k, w^k)$ の代わりにその近似行列 B^k を用いることもできる

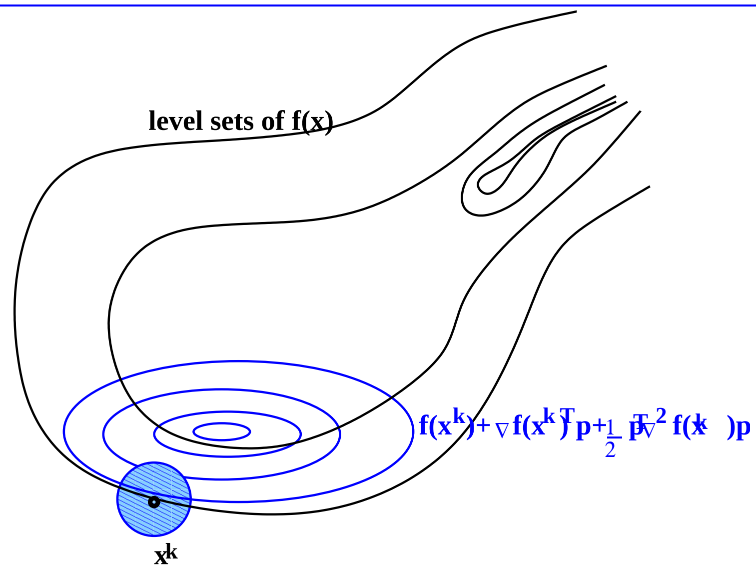
4.4.3　信頼領域法

- 無制約非線形計画問題を考える

$$
\begin{cases}
\text{最小化：} & f(x) \\
\text{条件：} & x \in \mathbb{R}^n
\end{cases}
$$

- k 反復目の点 x^k において信頼半径 Δ^k として $f(x)$ の 2 次近似モデルを考える

$$
\begin{cases}
\text{最小化：} & q^k(p) \equiv f(x^k) + \nabla f(x^k)^T p + \dfrac{1}{2} p^T \nabla^2 f(x^k) p \\
\text{条件：} & \|p\| \leq \Delta^k
\end{cases}
\tag{4.4}
$$



信頼領域法

Step 0: 初期点 $\boldsymbol{x}^0$，初期信頼半径 $\boldsymbol{\Delta}^0$ を与える．パラメータ $0 < \eta_1 \le \eta_2 < 1$, $0 < \gamma_1 < 1 < \gamma_2$ を決め，$k = 0$ とする．

終了条件: もし $\boldsymbol{x}^k$ が局所的最適解の近似になっていると判断できたならば終了する．

Step 1: 部分問題 (4.4) を解いて $\boldsymbol{p}^k$ を求める．

Step 2: もし $\dfrac{f(\boldsymbol{x}^k)-f(\boldsymbol{x}^{k+1})}{q^k(\boldsymbol{0})-q^k(\boldsymbol{p}^k)=-\nabla f(\boldsymbol{x}^k)^T\boldsymbol{p}^k-\frac{1}{2}(\boldsymbol{p}^k)^T\nabla^2 f(\boldsymbol{x}^k)\boldsymbol{p}^k} \ge \eta_1$ ならば，$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{p}^k$ とする．

Step 2.1: もし $\dfrac{f(\boldsymbol{x}^k)-f(\boldsymbol{x}^{k+1})}{q^k(\boldsymbol{0})-q^k(\boldsymbol{p}^k)=-\nabla f(\boldsymbol{x}^k)^T\boldsymbol{p}^k-\frac{1}{2}(\boldsymbol{p}^k)^T\nabla^2 f(\boldsymbol{x}^k)\boldsymbol{p}^k} \ge \eta_2$ ならば，$\boldsymbol{\Delta}^{k+1} \in [\boldsymbol{\Delta}^k, \gamma_2\boldsymbol{\Delta}^k]$ として信頼領域を拡大する．

Step 2.2: そうでなければ $\boldsymbol{\Delta}^{k+1} = \boldsymbol{\Delta}^k$ として現状を維持する．

Step 3: そうでなければ $\boldsymbol{\Delta}^{k+1} \in [\gamma_1\boldsymbol{\Delta}^k, \boldsymbol{\Delta}^k]$ として信頼領域を減少する．

Step 4: $k = k + 1$ とし，終了条件を調べる．

- 部分問題 (4.4) において $\nabla^2 f(\boldsymbol{x}^k)$ を準 Newton 法である BFGS 公式や DFP 公式を用いて更新することが一般的に実装されている

- 部分問題の解法は以下の定理により，固有値問題を解くことによって簡単に計算できる

> 定理１０
>
> $\boldsymbol{p}^k$ が部分問題 (4.4) の最適解である必要十分条件は次の条件を満たしている非負の値 $\boldsymbol{y}^k \geq 0$ が存在することである．
>
> $$(\nabla^2 f(x^k) + y^k I)p^k = -\nabla f(x^k)$$
> $$y^k(\Delta^k - \|p^k\|) = 0$$
> $$(\nabla^2 f(x^k) + y^k I) \text{ は正定値行列}$$

証明

[NOCEDAL2006] などを参照． □