2020年度「数値最適化」

第二部「制約の無い一般の最適化問題」
(2)多変数関数の最適化

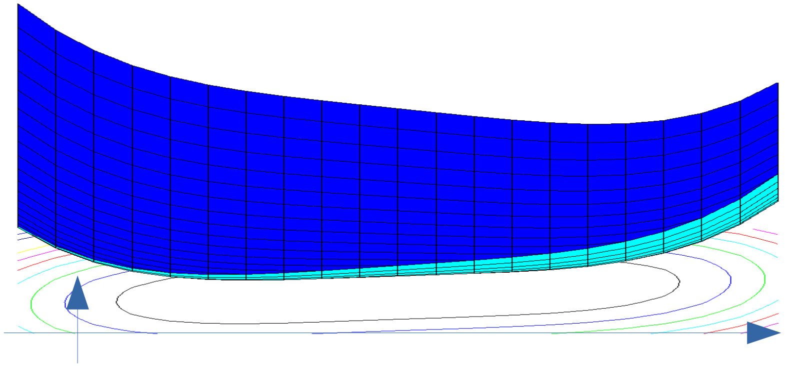
制約の無い非線形計画問題（多変数の場合）

制約の無い非線形計画問題:　$\min.\quad f(\boldsymbol{x}).$　(2)

制約関数の無い変数$\boldsymbol{x}$の非線形計画問題（2）に戻ります。

ベクトル$\boldsymbol{x}$の長さがn、すなわちn個の変数を持つ目的関数の非線形計画問題を考えます。

下図は$n=2$の目的関数$f(x_1,x_2)$を3次元透視図と等高線図で表した例です。



制約の無い非線形計画問題（多変数の場合）

多変数関数の停留点

非線形計画問題（2）の解$\boldsymbol{x}^*=[x^*{}_1,\ldots,x^*{}_n]^{\mathrm{T}}$において変数$x_j$の以外を固定した1変数関数$f_j(x_j)=f(x^*{}_1,\ldots,x^*{}_{j\text{-}1},x_j,x^*{}_{j+1},\ldots,x^*{}_n)$を考えます。

$x_j=x^*{}_j$はf_jの最適解なので停留点であり、$f_j{}'(x^*{}_j)=0$が成立します。

全ての変数で同じことが言えるので$\boldsymbol{x}^*$において次式が成立します。

$$\left.\frac{\partial f}{\partial x_1}\right|_{\boldsymbol{x}=\boldsymbol{x}^*} = \cdots = \left.\frac{\partial f}{\partial x_n}\right|_{\boldsymbol{x}=\boldsymbol{x}^*} = 0$$

上式の成立する点をfの停留点と呼びます。

注意：最適解は停留点ですが、停留点は最適解とは限りません。

勾配ベクトル

目的関数$f(\boldsymbol{x})=f(x_1,\ldots,x_n)$の偏導関数を用いて1変数の場合の1階導関数にあたる勾配ベクトル∇fを定めます。

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

勾配ベクトル∇fを用いたfの1次のテーラー展開式を考えることができます。$\boldsymbol{d}$は各成分が微小な変数の変化量です。

$$f(\boldsymbol{x}+\boldsymbol{d})=f(\boldsymbol{x})+\boldsymbol{d}^{\mathrm{T}}\nabla f+\mathrm{o}(||\boldsymbol{d}||)$$

勾配ベクトル

勾配ベクトル∇fがゼロベクトル$\mathbf{0}$でないとき、同じ向きの単位ベクトル$\boldsymbol{e}=\nabla f/||\nabla f||$を用いて微小ベクトル$\delta\boldsymbol{e}$による$f$の変化を

$$f(\boldsymbol{x}+\delta\boldsymbol{e})=f(\boldsymbol{x})+\delta\boldsymbol{e}^{\mathrm{T}}\nabla f+\mathrm{o}(||\delta\boldsymbol{e}||)=f(\boldsymbol{x})+\delta||\nabla f||+\mathrm{o}(||\delta||)$$

と表すことができます。（∵前ページテーラー展開式）

したがって、$\delta>0$ならfは増加し、$\delta<0$なら減少します。

このとき$-\nabla f$をfの最急降下方向、または最急降下方向ベクトルと呼びます。

同様にfを減少させるベクトルを降下方向、または降下方向ベクトルと呼びます。

多変数の降下法

1変数の場合と同様に、目的関数が減少する向き$\boldsymbol{v}_k$を選択して変数を更新し

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + r_k \boldsymbol{v}_k \qquad (\boldsymbol{v}_k \text{は降下方向ベクトル})$$

停留点を探索する方法を降下法と呼びます。

降下方向ベクトルを選択後、r_kを決定する過程を直線探索と呼びます。直線探索は1変数の最小化問題と見做すことができます。

$$\min.\ f(r_k) = f(\boldsymbol{x}_k + r_k \boldsymbol{v}_k)$$

すなわち直線探索では1変数の最小化問題で用いた計算法が利用できます。

最急降下法

降下方向として最急降下方向$-\nabla f$を選択した降下法

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k-r_k\nabla f(\boldsymbol{x}_k)$$

を最急降下法と呼びます。

最急降下法では直線探索に反復計算が必要で、その結果が最小化法全体の結果にも影響を与えますが、以下では1変数の最小化問題である直線探索は更新毎に適切に完了し最小解が得られるものと考えて説明を続けます。

$$f(\boldsymbol{x}_{k+1})=\min.\ f(r_k)=f(\boldsymbol{x}_k-r_k\nabla f(\boldsymbol{x}_k))$$

停留点と最小解･極小解

多変数の場合も同様に関数の凸性を定義できます。

凸関数

次の不等式を満たす関数を凸関数と呼びます。

任意の$\boldsymbol{x},\boldsymbol{y},0 \leqq t \leqq 1$に対し　$tf(\boldsymbol{x})+(1-t)f(\boldsymbol{y}) \geqq f(t\boldsymbol{x}+(1-t)\boldsymbol{y})$

※定義域が限定される場合はx,yは定義域からとります

狭義凸関数

次の不等式を満たす関数を狭義凸関数と呼びます。

任意の$\boldsymbol{x},\boldsymbol{y},0 \leqq t \leqq 1$に対し　$tf(\boldsymbol{x})+(1-t)f(\boldsymbol{y}) > f(t\boldsymbol{x}+(1-t)\boldsymbol{y})$

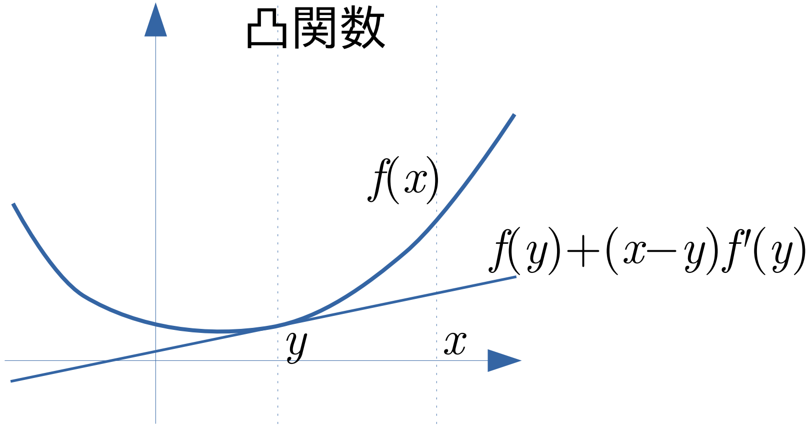
※条件式の不等号に等号が含まれません

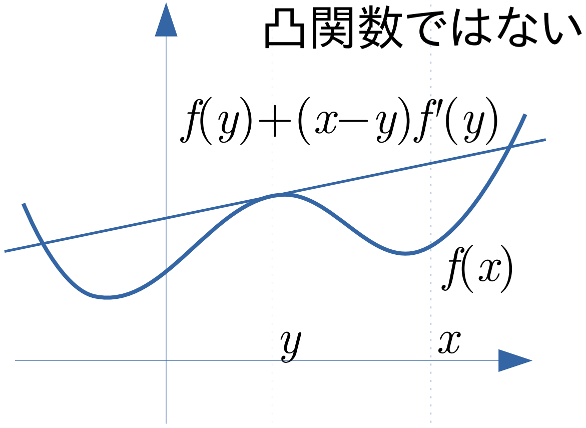
停留点と最小解·極小解

1階の導関数との関係は勾配ベクトルを用いたものになります。

次の不等式は1階連続微分可能な凸関数の必要十分条件です。

任意の$\boldsymbol{x},\boldsymbol{y}$に対し　$f(\boldsymbol{x}) \geqq f(\boldsymbol{y}) + [\nabla f(\boldsymbol{y})]^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y})$





図は1変数のものと同じものです。

停留点と最小解･極小解

凸関数の停留点の大域的最適性の定理も同様です。

定理：凸関数fの停留点は大域的最適解（＝最小解）である

任意の$\boldsymbol{x},\boldsymbol{y}$に対し　$f(\boldsymbol{x}) \geqq f(\boldsymbol{y}) + [\nabla f(\boldsymbol{y})]^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{y})$

$\boldsymbol{x}^*$が停留点であれば$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$

したがって、任意の$\boldsymbol{x}$について

$$f(\boldsymbol{x}) \geqq f(\boldsymbol{x}^*) + [\nabla f(\boldsymbol{x}^*)]^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) = f(\boldsymbol{x}^*)$$

すなわち$\boldsymbol{x}^*$は最小解となる。

停留点と最小解・極小解

多変数の目的関数$f(\boldsymbol{x})$において1変数の場合の2階導関数にあたらうものはヘシアン（Hesse行列）です。

$f(\boldsymbol{x})$の2階偏導関数を用いた行列HfをfのHesse行列と呼ぶ、

$$
Hf = \begin{bmatrix}
\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\
\vdots & \vdots & & \vdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2}
\end{bmatrix}
$$

停留点と最小解･極小解

Hesse行列Hfを用いてfの2次のテイラー展開を表します。

$$f(\boldsymbol{x}+\boldsymbol{d})=f(\boldsymbol{x})+(\nabla f)^{\mathrm{T}}\boldsymbol{d}+(1/2)\,\boldsymbol{d}^{\mathrm{T}}(Hf)\,\boldsymbol{d}+\mathrm{o}(||\boldsymbol{d}||^2)$$

fが凸関数なら任意の$\boldsymbol{x},\boldsymbol{y}$で

$$f(\boldsymbol{x})\geqq f(\boldsymbol{y})+[\nabla f(\boldsymbol{y})]^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y})$$

が成立するので、$\boldsymbol{x},\boldsymbol{y}$の代わりに$\boldsymbol{x}+\boldsymbol{d},\boldsymbol{x}$をとれば

$$f(\boldsymbol{x}+\boldsymbol{d})\geqq f(\boldsymbol{x})+[\nabla f(\boldsymbol{x})]^{\mathrm{T}}\boldsymbol{d}$$

これに2次のテイラー展開を併せて

$$(1/2)\,\boldsymbol{d}^{\mathrm{T}}[Hf(\boldsymbol{x})]\,\boldsymbol{d}+\mathrm{o}(||\boldsymbol{d}||^2)\geqq 0$$

停留点と最小解･極小解

十分小さい任意のdに対して

$$\boldsymbol{d}^{\mathrm{T}}[Hf(\boldsymbol{x})]\boldsymbol{d} \geqq 0$$

が言えることになるので、凸関数のHesse行列は半正定値であることが判る。（行列Aが半正定値⇔$\forall\, \boldsymbol{v},\ \boldsymbol{v}^{\mathrm{T}} A \boldsymbol{v} \geqq 0$）

逆に、fのHesse行列Hfが半正定値行列であれば、同様に2次のテイラー展開を通じてfが凸関数であることを言える。

（勾配ベクトルを用いた凸関数の必要十分条件が成立する）

Hesse行列の半正定値性により関数の凸性が言えて凸関数の性質により停留点が大域的最適解（最小解）と言える。

停留点と最小解･極小解

同様にfの停留点$\boldsymbol{x}^*$でfの2次のテイラー展開を考えれば

$$f(\boldsymbol{x}^*+\boldsymbol{d})=f(\boldsymbol{x}^*)+[\nabla f(\boldsymbol{x}^*)]^{\mathrm{T}}\boldsymbol{d}+(1/2)\,\boldsymbol{d}^{\mathrm{T}}[Hf(\boldsymbol{x}^*)]\,\boldsymbol{d}+\mathrm{o}(||\boldsymbol{d}||^2)$$

$$=f(\boldsymbol{x}^*)+(1/2)\,\boldsymbol{d}^{\mathrm{T}}[Hf(\boldsymbol{x}^*)]\,\boldsymbol{d}+\mathrm{o}(||\boldsymbol{d}||^2)$$

であるから停留点$\boldsymbol{x}^*$におけるHesse行列が半正定値であれば、
停留点が極小解であることが言える。

※ Hesse行列の半正定値性が成立する範囲に注意

降下法のバリエーション

目的関数fの1次のテイラー展開を用いて、勾配ベクトル∇fがゼロベクトル$\mathbf{0}$でないときの降下方向$\boldsymbol{v}$、すなわち目的関数を減少させる変数の変量を考えます。

$$f(\boldsymbol{x}+\boldsymbol{v})=f(\boldsymbol{x})+\boldsymbol{v}^{\mathrm{T}}\nabla f+\mathrm{o}(||\boldsymbol{v}||)$$

$\boldsymbol{v}\to 0$の極限を考えれば目的関数を減少させるための条件は

$$\boldsymbol{v}^{\mathrm{T}}\nabla f<0$$

です。このとき$\boldsymbol{v}=-\nabla f$を選べば減少の速度を最大化する（最急降下方向）であろうことは確かですが、条件を満たす降下方向には他の選択余地があることも判ります。

降下法のバリエーション

最急降下法の特徴

最急降下法において、直線探索が理想的に実行された場合、更新後の点は降下方向が等高線に沿うものになります。

したがって、その次の最急降下方向は直前の降下方向、すなわち等高線の接線に対して垂直になります。

これを繰り返すので、最急降下法では停留点探索の軌跡は1反復毎に直角に方向転換するものになります。

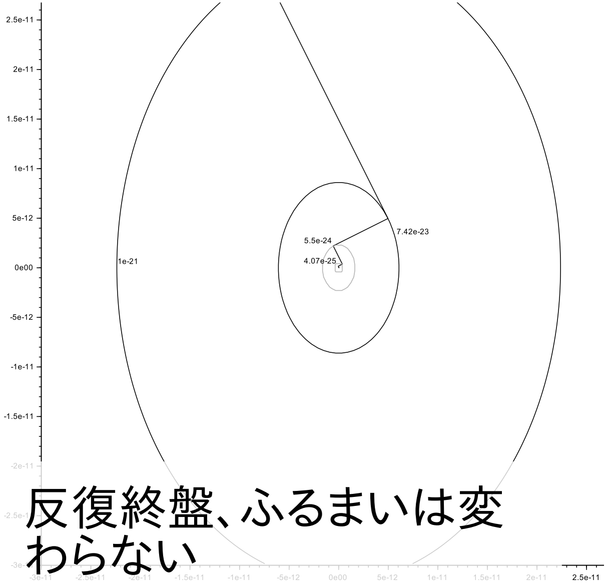
このような性質は反復の終盤で収束を遅らせる原因になります。

降下法のバリエーション

最急降下法の特徴

$f(x_1, x_2) = 2{x_1}^2 + {x_2}^2,\ \boldsymbol{x}_0 = [1, 1]^{\mathrm{T}}$に最急降下法を適用した例

反復初期段階、等高線で直角に方向転換している様子が判る



反復終盤、ふるまいは変わらない

ニュートン法

前ページで例示した目的関数の最小解が原点であることはすぐに判ります。

同様に

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}} \boldsymbol{x} + c$$

のように表現できる2次関数は勾配ベクトル

$$\nabla f(\boldsymbol{x}) = (A + A^{\mathrm{T}}) \boldsymbol{x} + \boldsymbol{b}$$

から停留点$\boldsymbol{x}^* = -(A + A^{\mathrm{T}})^{-1} \boldsymbol{b}$を容易に求めることができます。

このとき、Aが正定値で対称行列であればfは凸関数であるので、停留点は最小解となります。

ニュートン法

2次関数の停留点が容易に計算できることを利用するのが、ニュートン法です。

目的関数の2次のテイラー展開から近似関数が得られます。

$$f(\boldsymbol{x}+\boldsymbol{d}) \approx f(\boldsymbol{x}) + [\nabla f(\boldsymbol{x})]^{\mathrm{T}} \boldsymbol{d} + (1/2)\, \boldsymbol{d}^{\mathrm{T}} [Hf(\boldsymbol{x})]\, \boldsymbol{d}$$

近似関数の停留点は$\boldsymbol{d} = -[Hf(\boldsymbol{x})]^{-1}[\nabla f(\boldsymbol{x})]$とした$\boldsymbol{x}+\boldsymbol{d}$にあることが判ります。これを、降下法の更新過程で利用します。

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - [Hf(\boldsymbol{x}_k)]^{-1}[\nabla f(\boldsymbol{x})]$$

このような更新式を用いる方法をニュートン法と呼びます。

※fが$\boldsymbol{x}_k$で2回微分可能であれば$Hf(\boldsymbol{x}_k)$は対称行列です。

ニュートン法

目的関数の2次のテイラー展開からの説明とは別に、停留点探索の視点での説明を補足します。

停留点の条件式　$\nabla f(\boldsymbol{x})=\boldsymbol{0}$ をベクトル方程式と見做し、これにニュートン法による逐次解法を適用します。1次のテイラー展開

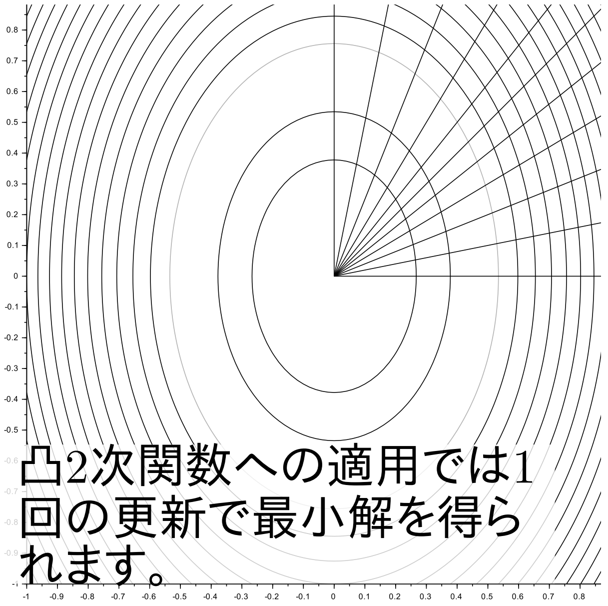
$$\nabla f(\boldsymbol{x}+\boldsymbol{d})=\nabla f(\boldsymbol{x})+\nabla[\nabla f(\boldsymbol{x})]^{\mathrm{T}}\boldsymbol{d}=\nabla f(\boldsymbol{x})+[Hf(\boldsymbol{x})]\boldsymbol{d}=\boldsymbol{0}$$

から連立得たdの連立一次方程式を解けば、2次関数の停留点を与えるものと同じ式を導出できます。

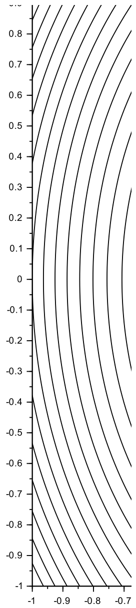
$$\boldsymbol{d}=-[Hf(\boldsymbol{x})]^{-1}[\nabla f(\boldsymbol{x})]$$

ニュートン法

$f(x_1, x_2) = 2{x_1}^2 + {x_2}^2$ にニュートン法（左）最急降下法（右）を適用した例（「最急降下法の特徴」と同じ目的関数です。）



凸2次関数への適用では1回の更新で最小解を得られます。



ニュートン法

最急降下法が反復の終盤に冗長な更新を繰り返すのに対して、ニュートン法は目的関数が十分に滑らか（3回微分可能）な終盤の最小解近傍で2次収束することが知られています。

収束の速度：

最小解$\boldsymbol{x}^*$との距離の1回の更新あたりの減少の様子で分類

1次収束：　$||\boldsymbol{x}_{k+1}-\boldsymbol{x}^*|| \leqq \alpha||\boldsymbol{x}_k-\boldsymbol{x}^*||$　となる$0<\alpha<1$が存在する。

2次収束：　$||\boldsymbol{x}_{k+1}-\boldsymbol{x}^*|| \leqq \beta||\boldsymbol{x}_k-\boldsymbol{x}^*||^2$　となる$0<\beta<1$が存在する。

※最急降下法は同様の条件で1次収束まで

直線探索付きニュートン法

ニュートン法が収束の速い良好な方法でなのは、2次のテイラー展開による目的関数の近似が良好で、凸関数である場合です。

一方で、降下法の要件$f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k)$を満たすためにはHesse行列$Hf(\boldsymbol{x}_k)$が半正定値であることが必要で、そうでない場合にはニュートン法の更新で目的関数値が減少しないこともあり得ます。

そこで、近似2次関数の停留点は更新方向のベクトル$\boldsymbol{d}$を得るためにめに用い、実際の更新点$\boldsymbol{x}_{k+1}$は直線探索を実行して決めるという方法が考えられます。

これは直線探索付きニュートン法と呼ばれ、降下法の要件を満たします。

準ニュートン法

ニュートン法の更新ベクトル

$$\boldsymbol{d} = -[Hf(\boldsymbol{x})]^{-1}[\nabla f(\boldsymbol{x})]$$

の計算には逆行列が現われます。少なくとも連立一次方程式の解法相当の計算が必要になります。また、Hesse行列が正則でなければ更新ベクトル$\boldsymbol{d}$が得られないことになります。

そこで反復毎のHesse行列の代わりに$[Hf(\boldsymbol{x}_k)]^{-1}$を近似する行列を利用するのが準ニュートン法です。

扱いの容易さから、代わりの行列B_k^{-1}には正定値対称行列になるものが選ばれます。

準ニュートン法

代わりの行列B_k^{-1}の選択基準としてセカント条件

$$B_{k+1}(\boldsymbol{x}_{k+1}-\boldsymbol{x}_k)=\nabla f(\boldsymbol{x}_{k+1})-\nabla f(\boldsymbol{x}_k)$$

が採用されます。これは勾配ベクトルのテイラー展開

$$\nabla f(\boldsymbol{x}_{k+1})=\nabla f(\boldsymbol{x}_k)+[Hf(\boldsymbol{x}_k)](\boldsymbol{x}_{k+1}-\boldsymbol{x}_k)+\mathrm{o}(||\boldsymbol{x}_{k+1}-\boldsymbol{x}_k||)$$

から微小項o(...)を無視して得られる、Hesse行列が満たすべき条件です。セカント条件にはHesse行列が現われないので、Hesse行列自体を計算する必要はありません。

セカント条件は連立一次方程式ではなく、B_kには様々な選択の余地が残されています。

準ニュートン法

Hesse行列$Hf(\boldsymbol{x}_k)$の代わりに用いるB_kは正定値行列であることが望ましく、またできるだけ計算が容易であること望ましいので、B_kからB_{k+1}を得る次のBFGS法によるB_kの更新がよく採用されます。

$$B_{k+1} = B_k - \frac{B_k \boldsymbol{s}_k (B_k \boldsymbol{s}_k)^{\mathrm{T}}}{(\boldsymbol{s}_k)^{\mathrm{T}} B_k \boldsymbol{s}_k} + \frac{\boldsymbol{y}_k (\boldsymbol{y}_k)^{\mathrm{T}}}{(\boldsymbol{s}_k)^{\mathrm{T}} \boldsymbol{y}_k}$$

ただし、$\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k,\ \boldsymbol{y}_k = \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$です。

このとき、B_kがセカント条件を満たす正定値であればB_{k+1}もセカント条件を満たし正定値となります。

準ニュートン法

BFGS法のような正定値対称行列を用いた準ニュートン法は、Hesse行列には一致しませんが、目的関数を凸2次関数で近似し停留点探索を実施するので良好な場合のニュートン法と同様に収束の速い方法として期待できます。

制約の無い多変数の非線形計画法

まとめ

多変数の目的関数による（制約の無い）非線形計画問題では、

勾配ベクトル$\nabla f(\boldsymbol{x})$が1変数の1階導関数、Hesse行列$Hf(\boldsymbol{x})$が2階導関数に対応することを学びました。

多変数の場合も、1変数と同様に関数の凸性を定義することができ、停留点の大域最適性、あるいは局所最適性に関連し、重要な性質であることを理解しました。

勾配ベクトル$\nabla f(\boldsymbol{x})$とHesse行列$Hf(\boldsymbol{x})$を利用して降下方向を求め、1変数の場合の降下法と同様に停留点を求める方法が利用できることを学びました。