数理解析研究所講究録
915 巻 1995 年 192-202

安定な準ニュートン法のクラスについて

八巻直一　　　　　　矢部　博
システム計画研究所　　　東京理科大学工学部

A Class of Stable Quasi-Newton Methods

Naokazu Yamaki* 　　Hiroshi Yabe**
*Research Institute of Systems Planning Ltd.
**Science University of Tokyo

1 はじめに

本稿では，次の無制約最小化問題に対する数値解法を考える.

[UMP] $f: R^n \rightarrow R$ のとき,$f(x)$を$x \in R^n$について最小化せよ.

この問題を解くための有力な手法として，ニュートン法が挙げられる．またその他の反復法の多くは，ニュートン法の考え方に基づいており，総称して準ニュートン法と呼ばれる．問題 [UMP] を解くための準ニュートン法のアルゴリズムのプロトタイプは，以下の通りである.

[NL]

Step 0. 初期点 $x_0 \in R^n$ と初期行列 $B_0 \in R^{n\times n}$を与える. $k=0$ とおく.

Step 1. $\|\nabla f(x_k)\|$ が非常に小さいならば停止する．さもなければ Step 2 へいく.

Step 2. 次の線形方程式を解いて，探索方向 d_kを作る

(1.1)
$$B_k d = -\nabla f(x_k).$$

Step 3. $x_{k+1}=x_k+\alpha_k d_k$とおく．ただし，α_kは適当なステップ幅である.

Step 4. 新しい行列 B_{k+1}を作り，$k=k+1$ とおいて，Step 1 へ行く. □

ここで，行列 B_kはヘッセ行列 $\nabla^2 f(x_k)$ 自身，もしくはその近似である．アルゴリズムの Step 4 はこの方法を特徴づけるもので，B_kを変えることによっていろいろな解法が得られる．たとえば，$B_k=\nabla^2 f(x_k)$ のときニュートン法そのものが得られるし，B_kを更新して B_{k+1}を生成すればセカント法が得られる [8]. 後者の場合，新しい行列 B_{k+1}は次式で生成される

$$B_{k+1}=B_k+\Delta B_k,$$

ただし，ΔB_kは補正行列で，満足すべき条件として，テイラー展開から導かれる次のセカント条件が課される.

(1.2)
$$B_{k+1}s_k=y_k,$$

ここで,

$$s_k = x_{k+1} - x_k, \qquad y_k = \nabla f(x_{k+1}) - \nabla f(x_k) \tag{1.3}$$

である.

ニュートン法は，現在でも最も広く用いられている手法の一つであるが，ヘッセ行列の計算がやっかいであることと，不安定挙動を示す場合があり得ることなどの欠点をもつ．このような欠点を克服し，かつニュートン法の利点すなわち収束の速さを確保することを目的として，準ニュートン法が提案された．準ニュートン法は，非線形関数の最小点を求める手法として，現在考えられる最有力な手法の一つである．準ニュートン法では $f(x)$ のヘッセ行列の近似行列を用いるが，ニュートン法の利点を保持するために，近似行列に対してセカント条件といわれる条件が課せられる．準ニュートン法においては，ヘッセ行列の近似行列の正定値対称性の保証が，アルゴリズムの安定性に本質的に望ましい．実際，有力な方法である BFGS 法や DFP 法では，f のヘッセ行列の近似行列の正定値対称性が保存される．B_kが正定値対称であれば，探索ベクトルは $f(x)$ の減少方向となり，したがって直線探索では$\alpha_k > 0$ となる．このような手法のクラスを，直線探索戦略という．本稿では，正定値性の保証された B_kを伴う直線探索戦略に基づく準ニュートン法を，安定な準ニュートン法と定義する．

一方，準ニュートン法におけるヘッセ行列の近似行列の更新公式には，正定値性を保証しないが有力な公式もいくつか知られている．最近注目されている信頼領域法では，正定値性の保存が必ずしも必要ではない．このような手法のクラスを，信頼領域戦略という．信頼領域戦略では直線探索の代わりに，現在の近似解の適当な近傍で目的関数の値を小さくするような，子問題を解くことによって次の近似解を得る．

本稿では，セカント条件と正定値対称性の，両方を満足する行列の一般形を導出し，その性質について検討する．すなわち，次の3つの条件を満足するような行列 B_+を考える．この条件を SSP 条件と呼ぶことにする．

[**SSP 条件**]

(1:S) $s^T z > 0$ を満たすベクトル $s, z \in R^n$が与えられたとき，行列 B_+は次の条件を満足する：

$$B_+ s = z. \tag{1.4}$$

(2:S) 行列 B_+は対称である.

(3:P) 行列 B_+は正定値である.

SSP 条件を満足する行列を，総称して SSP 行列と呼ぶ．ここで，$s^T z > 0$ は B_+が正定値であるための必要条件になっていることに注意しよう．また，条件 (1.4) にはヘッセ行列に関する適切な情報が含まれていなければならない．セカント法の場合，条件 (1.4) はセカント条件 (1.2) そのものになるが，我々はセカント法に限定せず，SSP 条件を満足する準ニュートン法について議論する．なお，以下で述べる我々の研究は，Nazareth[6] の分類によれば metric-based 法になる.

2 SSP 行列の一般形

本節では SSP 行列の一般形とその分解型を導出する．SSP 行列の一般形が，次の定理によって与えられる.

定理 1． 式 (2.1) によって与えられる行列 B_+は SSP 行列の一般形になる．ただし，Φは列フルランクな任意行列，u は $s^T \Phi^T u \neq 0$ を満たす任意ベクトルである.

$$B_+ = C - \frac{Css^TC}{s^TCs} + \frac{zz^T}{s^Tz} + (s^TCs)ww^T \tag{2.1}$$

$$= \left(I - \frac{su^T\Phi}{s^T\Phi^Tu}\right)^T C \left(I - \frac{su^T\Phi}{s^T\Phi^Tu}\right) + \frac{zz^T}{s^Tz}, \tag{2.2}$$

ただし,

$$w = \frac{\Phi^Tu}{s^T\Phi^Tu} - \frac{Cs}{s^TCs}, \qquad C = \Phi^T\Phi. \tag{2.3}$$

証明. 式 (2.1) で与えられる行列 B_+が SSP 条件を満足することは, 容易に示される.

逆に, $\widehat{B}_+$を SSP 条件を満足する任意の行列とする. $\widehat{B}_+$は正定値対称なので, 列フルランクな行列 $\widehat{L}_+ \in R^{m\times n}$が存在して, $\widehat{B}_+ = \widehat{L}_+^T\widehat{L}_+$となる. $\Phi = \widehat{L}_+$とおけば $C = \widehat{B}_+$となるので, $\Phi^Tu = z$となる u に対して, $w = 0$ となる. このとき (2.1) の右辺から

$$C - \frac{Css^TC}{s^TCs} + \frac{zz^T}{s^Tz} + (s^TCs)ww^T = \widehat{B}_+ - \frac{\widehat{B}_+ss^T\widehat{B}_+}{s^T\widehat{B}_+s} + \frac{zz^T}{s^Tz} = \widehat{B}_+$$

を得る. すなわち, $\widehat{B}_+$は (2.1) の形で表される. よって, (2.1) は SSP 行列の一般形である. したがって, 定理が証明された. □

SSP 行列の分解型に関しては, 次の系が得られる.

系 1. ベクトル $s, z \in R^n$ は $s^Tz > 0$ を満たすとする. このとき

$$\begin{aligned} N_+ &= \left(I - \frac{\Phi su^T}{s^T\Phi^Tu}\right)\Phi + \frac{uz^T}{\sqrt{s^Tz}\|u\|} \\ &= \Phi - \frac{\Phi su^T\Phi}{s^T\Phi^Tu} + \frac{uz^T}{\sqrt{s^Tz}\|u\|} \end{aligned} \tag{2.4}$$

で与えられる行列 $N_+ \in R^{m\times n}$ は行列方程式

$$\begin{aligned} N_+^T\frac{\sqrt{s^Tz}u}{\|u\|} &= z \\ N_+s &= \frac{\sqrt{s^Tz}u}{\|u\|} \end{aligned}$$

の一般解になる. ただし, $n \le m$ であり, $\Phi \in R^{m\times n}$と $u \in R^m$ は $s^T\Phi^Tu \neq 0$ を満たす任意行列と任意ベクトルである.

(証明は略す.)

この結果を用いて $B_+ = N_+^TN_+$とおけば, SSP 行列の一般形が得られる. すなわち,

$$\begin{aligned} N_+^TN_+ &= \left[\left(I - \frac{\Phi su^T}{s^T\Phi^Tu}\right)\Phi + \frac{uz^T}{\sqrt{s^Tz}\|u\|}\right]^T \left[\left(I - \frac{\Phi su^T}{s^T\Phi^Tu}\right)\Phi + \frac{uz^T}{\sqrt{s^Tz}\|u\|}\right] \\ &= \left(I - \frac{su^T\Phi}{s^T\Phi^Tu}\right)^T C \left(I - \frac{su^T\Phi}{s^T\Phi^Tu}\right) + \frac{zz^T}{s^Tz}. \end{aligned}$$

$H_+ = B_+^{-1}$に対応する一般形も, 同様に導出できる. この場合, SSP 条件の一番目は次式で置き代わる.

$$H_+z = s. \tag{2.5}$$

このとき，いままでと同様の議論をすれば，SSP 条件を満足する H_+の一般形とその分解型が得られる：

$$
\begin{aligned}
H_+ &= \tilde{C} - \frac{\tilde{C}zz^T\tilde{C}}{z^T\tilde{C}z} + \frac{ss^T}{z^Ts} + (z^T\tilde{C}z)\tilde{w}\tilde{w}^T \\
&= \left(I - \frac{z\tilde{u}^T\tilde{\Phi}}{z^T\tilde{\Phi}^T\tilde{u}}\right)^T \tilde{C}\left(I - \frac{z\tilde{u}^T\tilde{\Phi}}{z^T\tilde{\Phi}^T\tilde{u}}\right) + \frac{ss^T}{s^Tz}
\end{aligned} \tag{2.6}
$$

ただし，

$$
\tilde{w} = \frac{\tilde{\Phi}^T\tilde{u}}{z^T\tilde{\Phi}^T\tilde{u}} - \frac{\tilde{C}z}{z^T\tilde{C}z}, \qquad \tilde{C} = \tilde{\Phi}^T\tilde{\Phi},
$$

$$
\begin{aligned}
\tilde{N}_+ &= \left(I - \frac{\tilde{\Phi}z\tilde{u}^T}{z^T\tilde{\Phi}^T\tilde{u}}\right)\tilde{\Phi} + \frac{\tilde{u}s^T}{\sqrt{s^Tz}\|\tilde{u}\|} \\
&= \tilde{\Phi} - \frac{\tilde{\Phi}z\tilde{u}^T\tilde{\Phi}}{z^T\tilde{\Phi}^T\tilde{u}} + \frac{\tilde{u}s^T}{\sqrt{s^Tz}\|\tilde{u}\|},
\end{aligned} \tag{2.7}
$$

ただし，$\tilde{\Phi}$は列フルランクな任意行列，$\tilde{u}$は $z^T\tilde{\Phi}^T\tilde{u} \neq 0$ を満たす任意ベクトルである．逆行列の形式を用いれば，探索方向は直接

$$
d_k = -H_k\nabla f(x_k)
$$

から求まり，線形方程式を解く必要がない．

3 セカント法

本節では，無制約最小化問題を解くためのセカント法を扱う．この方法は，アルゴリズム [NL] の Step 4 のヘッセ行列の近似行列 B_{k+1}を現在の B_kを更新して得るもので，その際にセカント条件 (1.2) を満たすことが要請される．一般形 (2.1)，(2.4) を用いれば，SSP 条件を満足する B_kの更新公式とその分解型が得られる．式 (2.1) と (2.4) で $s = s_k$, $z = y_k$, $N_+ = L_{k+1}$, $\Phi = L_k \in R^{n\times n}$, $B_k = L_k^TL_k$，$B_{k+1} = L_{k+1}^TL_{k+1}$ とおけば，次の形式が得られる：

$$
B_{k+1} = B_k - \frac{B_ks_ks_k^TB_k}{s_k^TB_ks_k} + \frac{y_ky_k^T}{s_k^Ty_k} + (s_k^TB_ks_k)w_kw_k^T, \tag{3.1}
$$

$$
L_{k+1} = L_k - \frac{L_ks_ku_k^TL_k}{s_k^TL_k^Tu_k} + \frac{1}{\sqrt{s_k^Ty_k}}\frac{u_k}{\|u_k\|}y_k^T, \tag{3.2}
$$

ただし，u_kは $s_k^TL_k^Tu_k \neq 0$ を満たす任意のベクトルで，

$$
w_k = \frac{L_k^Tu_k}{s_k^TL_k^Tu_k} - \frac{B_ks_k}{s_k^TB_ks_k}. \tag{3.3}
$$

式 (3.1) は正定値対称なセカント法のクラスで，パラメータベクトル $u_k \in R^n$を含んでいる．そして (3.2) はその分解型である．ここで，行列 (3.1) はランクが高々 3 の補正項をもっていることに注意しよう．特別な u_kを選べば，ランクが 2 の更新公式になる．本節を通して行列 $L_k \in R^{n\times n}$は正則であると仮定する．

3.1 Broyden 公式族

パラメータ $\phi_k \geq 0$ に対して

$$
u_k = (1 - \sqrt{\phi_k})\frac{L_ks_k}{s_k^TB_ks_k} + \sqrt{\phi_k}\frac{(L_k^T)^{-1}y_k}{s_k^Ty_k} \tag{3.4}
$$

とおくと，一般形 (3.1) と (3.2) から Broyden 公式族とその分解型が得られる：

$$B_{k+1}^{Broyden} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \phi_k (s_k^T B_k s_k) v_k v_k^T, \tag{3.5}$$

$$v_k = \frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k}, \tag{3.6}$$

$$\begin{aligned} L_{k+1}^{Broyden} &= L_k + (1 - \sqrt{\phi_k}\,)\left(\frac{L_k s_k}{s_k^T B_k s_k}\right)(\sqrt{\lambda_k}\; y_k - B_k s_k)^T \\ &\quad + \sqrt{\phi_k}\; L_k(\sqrt{\lambda_k}\; B_k^{-1} y_k - s_k)\left(\frac{y_k}{s_k^T y_k}\right)^T, \end{aligned} \tag{3.7}$$

ただし，$\phi_k \geq 0$ であり

$$\lambda_k = \left[(1 - \phi_k)\frac{s_k^T y_k}{s_k^T B_k s_k} + \phi_k \frac{y_k^T B_k^{-1} y_k}{s_k^T y_k}\right]^{-1}$$

である．特別なϕ_kを選べばBFGS公式やDFP公式，およびそれらの分解型が得られる．すなわち，式 (3.5) と (3.7) で$\phi_k = 0$ とおけば BFGS 公式とその分解型が得られる．また同様に，式 (3.5) と (3.7) で$\phi_k = 1$ とおけば DFP 公式とその分解型が得られる．

もし B_kが正定値対称で $s_k^T y_k > 0$ が成り立ち，かつ

$$\phi_k > \phi_k^\sharp = -\left(\frac{(s_k^T B_k s_k)(y_k^T B_k^{-1} y_k)}{(s_k^T y_k)^2} - 1\right)^{-1} \tag{3.8}$$

となるϕ_kを選べば，Broyden 公式族 (3.5) が正定値対称になることが知られている．もし s_k と y_kが線形独立ならば，$\phi_k^\sharp < 0$ となることに注意しよう．パラメータ$\phi_k \geq 0$ に対する Broyden 公式族の分解型を前述したが，実際は$\phi_k > \phi_k^\sharp$に対して分解型が得られることが期待される．これについては，今後の課題である．

3.2 局所的超一次収束性

本節では，セカント法のクラス (3.1) あるいは (3.2) の，局所的超一次収束性を示す定理を与える．

$D \subset R^n$ を開凸集合とし，$f(x)$ の局所最適解 x_*を含むものとする．$f(x)$ は以下の仮定を満たすものとする．

(A1) 任意の $x \in D$に対して，次の条件を満たす正定数 ξ と p が存在する．

$$\|\nabla^2 f(x) - \nabla^2 f(x_*)\| \leq \xi \|x - x_*\|^p \tag{3.9}$$

(A2) $\nabla^2 f$ は x_*で正定値対称である．

また，行列 Q に対して $\|Q\|_F$ は Frobenius ノルム，$\|Q\|_{F,M} = \|M^{-1} Q M^{-1}\|_F$とする．

このとき，以下の結果を得る．

定理 2． 仮定 (A1) および (A2) が満たされているとし，B_k は (3.1) によって更新されるものとする．このとき，正の定数 η_1 と η_2があって，u_k が

$$\|(s_k^T B_k s_k)(w_k w_k^T - v_k v_k^T)\|_{F,M} \leq \eta_1 \frac{\|\widehat{s}_k - \widehat{B}_k \widehat{s}_k\|^2}{\|\widehat{s}_k\|^2} + \eta_2 \sigma_k^p \tag{3.10}$$

を満足するものとする．点列 $\{x_k\}$ は

$$x_{k+1} = x_k + s_k \quad , \quad B_k s_k = -\nabla f(x_k). \tag{3.11}$$

によって定義されるものとする．

このとき，任意の $\nu \in (0,1)$ に対して，正定数 ε と δ が存在して，

$$\|x_0 - x_*\| \leq \varepsilon, \qquad x_0 \in D$$

および

$$\|B_0 - \nabla^2 f(x_*)\|_{F,M} \leq \delta,$$

であれば，点列 $\{x_k\}$ は x_* に超一次収束する．

ここに，

$$M = \nabla^2 f(x_*)^{\frac{1}{2}} \tag{3.12}$$

および

$$\sigma_k = \max(\|x_{k+1} - x_*\|, \|x_k - x_*\|), \tag{3.13}$$

さらに

$$\widehat{B}_k = M^{-1} B_k M^{-1}, \quad \widehat{s}_k = M s_k \quad , \quad \widehat{B}_{k+1} = M^{-1} B_{k+1} M^{-1}. \tag{3.14}$$

である．

証明は省略するが，Broyden-Dennis-Moré の収束条件 [2] および，Dennis-Moré の超一次収束のための条件 [3] を満たすことが示される．

また，非負のパラメータ ϕ_k を持つ Broyden 公式族は，(3.1) あるいは (3.2) に含まれ，かつ定理 2 の条件を満たす．

4 SSP 変換とその応用例

一般形 (2.1) は，与えられた任意の正定値対称行列 C を SSP 行列に変換する作用素ともみなすことができる．そこで我々は，式 (2.1) を SSP 変換と呼ぶことにし，$SSP(C, s, z, u)$ と表すことにする．このとき，s, z は与えられたベクトル，u はパラメータである．

以下では，この変換を用いた例を 4 つ取り上げる．

4.1 ニュートン法

アルゴリズム [**NL**] において，$B_k = \nabla^2 f(x_k)$ とおくと，ニュートン法となる．ニュートン法の場合，$\nabla^2 f(x_k)$ が正定値でないとき，あるいは解析的に計算が困難なときアルゴリズムの継続に支障となる．そこで，例えば $\nabla^2 f(x_k)$ が正定値でなくなった場合，適当な $\lambda_k > 0$ を用いると，$\nabla^2 f(x_k) + \lambda_k I$ を正定値にすることができる．このとき，$SSP(C, s, z, u)$ において $C = \nabla^2 f(x_k) + \lambda_k I, C = L^T L, s_{k-1} = x_k - x_{k-1}$，および $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$ とおくと，$s_{k-1}^T y_{k-1} > 0$ のとき，セカント条件を満たしかつ安定なアルゴリズムが得られる．

$$\begin{aligned} B_k &= SSP(C, s_{k-1}, y_{k-1}, \frac{L s_{k-1}}{s_{k-1}^T C s_{k-1}}) \\ &= \nabla^2 f(x_k) + \lambda_k I - \frac{(\nabla^2 f(x_k) + \lambda_k I) s_{k-1} s_{k-1}^T (\nabla^2 f(x_k) + \lambda_k I)}{s_{k-1}^T (\nabla^2 f(x_k) + \lambda_k I) s_{k-1}} + \frac{y_{k-1} y_{k-1}^T}{s_{k-1}^T y_{k-1}}. \end{aligned} \tag{4.1}$$

また，ヘッセ行列 $\nabla^2 f(x)$ が計算可能部分 $P(x)$ と，解析的に計算することが困難な部分 $E(x)$ の和で与えられるとき，すなわち $\nabla^2 f(x) = P(x)+E(x)$ と書ける場合，適当な$\lambda_k \geq 0$ を用いると，$P(x_k)+\lambda_k I$を正定値にすることができる．このとき，$SSP(C,s,z,u)$ において $C = P(x_k)+\lambda_k I, C = L^T L, s_{k-1} = x_k - x_{k-1}$, および $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$ とおくと，$s_{k-1}^T y_{k-1} > 0$ のとき，セカント条件を満たしかつ安定なアルゴリズムが得られる．

$$
\begin{aligned}
B_k &= SSP(C, s_{k-1}, y_{k-1}, \frac{L s_{k-1}}{s_{k-1}^T C s_{k-1}}) \\
&= P(x_k) + \lambda_k I - \frac{(P(x_k)+\lambda_k I) s_{k-1} s_{k-1}^T (P(x_k)+\lambda_k I)}{s_{k-1}^T (P(x_k)+\lambda_k I) s_{k-1}} + \frac{y_{k-1} y_{k-1}^T}{s_{k-1}^T y_{k-1}}.
\end{aligned} \tag{4.2}
$$

4.2 スパース構造を持つ準ニュートン法

大規模な無制約最適化問題に準ニュートン法を適用する場合，B_kのサイズが大きくなる．このとき，通常のセカント法の公式で得られる B_kは密行列であり，ある程度以上のサイズになると記憶容量がデッドロックとなる．他方，大規模な問題では，ヘッセ行列がスパース構造を持つことが一般的であるので，B_kにヘッセのスパース構造を持たせるような公式が考えられた．[9][11]

しかしながら，これらの B_kは正定値性が保証されない．今のところ，スパース構造と正定値性を両立させる公式は発見されていない．そこで，SSP 変換を用いて正定値を保証するスパース準ニュートン法の導出を試みる．

$\tilde{B}_k$を，スパース公式によって得た行列とする．このとき，適当な$\lambda_k > 0$ に対して，$\tilde{B}_k + \lambda_k I$を正定値と出来る．したがって，例えば $SSP(C,s,z,u)$ において，$C = \tilde{B}_k + \lambda_k I$，とおくと，

$$
\begin{aligned}
B_k &= SSP(C, s_k, y_k, u) \\
&= C - \frac{C s_k s_k^T C}{s_k^T C s_k} + \frac{y_k y_k^T}{s_k^T y_k} + s_k^T C s_k w w^T,
\end{aligned} \tag{4.3}
$$

ただし，

$$
\begin{aligned}
w &= \frac{\Phi^T u}{s_k^T \Phi^T u} - \frac{C s_k}{s_k C s_k}, \\
C &= \Phi^T \Phi
\end{aligned}
$$

が得られる．ここで $u = \Phi s_k$および$\Phi^T u = y_k$とおけば，それぞれ，BFGS 型形式と DFP 型形式が得られる．

(4.3) そのものはスパースではないが，方程式

$$
B_k d = -\nabla f(x_k) \tag{4.4}
$$

は，BFGS 型公式や DFP 型公式を適用し，SOR 法などの反復法を用いることにより，スパース行列 Cの他にベクトル s_kと y_kの 2 本を用意するだけで解くことができる．ただし，λ_kをどうやって決定するかは，今後の課題である．

4.3 Limited memory 法

limited memory 法は，大規模な無制約最適化問題を解くために Nocedal [7] らによって提案された方法である．この考え方は，ヘッセ行列の逆行列 $\nabla^2 f(x_k)^{-1}$を H_kで近似することに基づく．本節では，この方法を SSP 変換の立場から統一的にながめてみたい．

式 (2.6) で $z = y_k$ とおけば，次の一般形を得る：

$$H_{k+1} = SSP(\tilde{\Phi}^T\tilde{\Phi}, y_k, s_k, \tilde{u}) = \left(I - \frac{y_k\tilde{u}_k^T\tilde{\Phi}}{y_k^T\tilde{\Phi}^T\tilde{u}_k}\right)^T \tilde{\Phi}^T\tilde{\Phi}\left(I - \frac{y_k\tilde{u}_k^T\tilde{\Phi}}{y_k^T\tilde{\Phi}^T\tilde{u}_k}\right) + \frac{s_k s_k^T}{y_k^T s_k}. \tag{4.5}$$

初期行列が，次のように分解型で与えられていると仮定する．

$$H_0 = L_0^T L_0, \qquad L_0 \in R^{n\times n}.$$

このとき，あらかじめ定められた数 t に対して，

$$V_k = I - \frac{y_k\tilde{u}_k^T\tilde{\Phi}}{y_k^T\tilde{\Phi}^T\tilde{u}_k} \tag{4.6}$$

および

$$\tilde{\Phi} = \begin{pmatrix} L_0 V_{k-t} \cdots V_{k-1} \\ \frac{1}{\sqrt{s_{k-t}^T y_{k-t}}} s_{k-t}^T (V_{k-t+1} \cdots V_{k-1}) \\ \frac{1}{\sqrt{s_{k-t+1}^T y_{k-t+1}}} s_{k-t+1}^T (V_{k-t+2} \cdots V_{k-1}) \\ \vdots \\ \frac{1}{\sqrt{s_{k-2}^T y_{k-2}}} s_{k-2}^T V_{k-1} \\ \frac{1}{\sqrt{s_{k-1}^T y_{k-1}}} s_{k-1}^T \end{pmatrix} \in R^{(n+t)\times n} \tag{4.7}$$

とおくと，

$$\begin{aligned} H_{k+1} &= (V_{k-t}V_{k-t+1} \cdots V_{k-1}V_k)^T H_0 (V_{k-t}V_{k-t+1} \cdots V_{k-1}V_k) \\ &\quad + (V_{k-t+1}V_{k-t+2} \cdots V_{k-1}V_k)^T \frac{s_{k-t}s_{k-t}^T}{s_{k-t}^T y_{k-t}} (V_{k-t+1}V_{k-t+2} \cdots V_{k-1}V_k) \\ &\quad + \cdots + (V_{k-1}V_k)^T \frac{s_{k-2}s_{k-2}^T}{s_{k-2}^T y_{k-2}} (V_{k-1}V_k) + V_k^T \frac{s_{k-1}s_{k-1}^T}{s_{k-1}^T y_{k-1}} V_k + \frac{s_k s_k^T}{s_k^T y_k} \end{aligned} \tag{4.8}$$

を得る．この結果から，$V_{k-t}, \ldots, V_k$ に含まれる $2(t+1)$ 本のベクトル $s_i, y_i, i = k-t, \ldots, k$ を保存しておけば，探索方向は $d_{k+1} = -H_{k+1}\nabla f(x_{k+1})$ として得られる．

ただし，$k \leq t$ の場合には，上式は

$$\begin{aligned} H_{k+1} &= (V_0V_1 \cdots V_{k-1}V_k)^T H_0 (V_0V_1 \cdots V_{k-1}V_k) \\ &\quad + (V_1V_2 \cdots V_{k-1}V_k)^T \frac{s_0 s_0^T}{s_0^T y_0} (V_1V_2 \cdots V_{k-1}V_k) + \cdots \\ &\quad + (V_{k-1}V_k)^T \frac{s_{k-2}s_{k-2}^T}{s_{k-2}^T y_{k-2}} (V_{k-1}V_k) + V_k^T \frac{s_{k-1}s_{k-1}^T}{s_{k-1}^T y_{k-1}} V_k + \frac{s_k s_k^T}{s_k^T y_k} \end{aligned} \tag{4.9}$$

とみなす．もし$\Phi^T u_k = s_k$を満たすu_kを選んで$t = k$とおけば，行列H_kとH_{k+1}の関係は次のようになる

$$
\begin{aligned}
H_{k+1} &= V_k^T H_k V_k + \frac{s_k s_k^T}{s_k^T y_k} \\
&= H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} + \left(1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k}, \qquad (4.10)
\end{aligned}
$$

ただし，

$$V_k = I - \frac{y_k s_k^T}{s_k^T y_k}.$$

実は，この式はBFGS公式である．したがって，$\Phi^T u_k = s_k$の場合には，以上の方法はNocedalによって提案されたlimited memory BFGS法になる．また，いろいろなu_kを選べば，いろいろな種類のlimited memory法が得られる（ただし，形式的に定式化できるという意味であって，必ずしも計算機の記憶容量が大幅に節約できるような実用性をもっているとは限らない）．なお，初期行列H_0が正定値対称で，すべてのiに対して$s_i^T y_i > 0$が成り立ち，かつ行列$\tilde{\Phi}$が列フルランクならば，(4.8)および(4.9)で定義される行列はSSP条件を満足することが証明できることを注意しておく．

4.4 非線形最小2乗問題に対するGauss-Newton法

本節では，次の非線形最小2乗問題を扱う．

$$f(x) = \frac{1}{2}\sum_{j=1}^{\nu}(r_j(x))^2 \text{ を最小化せよ } (n \le \nu), \qquad (4.11)$$

ただし，$r_j{:}R^n \to R,\ j = 1, 2, \ldots, \nu$は非線形関数である．以下では，残差ベクトルを

$$r(x) = (r_1(x), \ldots, r_\nu(x))^T$$

と表す．以上の問題では，目的関数の勾配ベクトルとヘッセ行列は特別な構造をもっていて，

$$\nabla f(x) = J(x)^T r(x), \qquad (4.12)$$

$$\nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^{\nu} r_j(x) \nabla^2 r_j(x) \qquad (4.13)$$

となる．ただし，$J(x)$は残差ベクトル$r(x)$のヤコビ行列である．このとき，ヘッセ行列$\nabla^2 f(x)$の第1項$J(x)^T J(x)$は1階偏導関数の情報しかもっていないのに比べて，第2項$\sum_{j=1}^{\nu} r_j(x)\nabla^2 r_j(x)$は2階偏導関数の情報を含んでいるので，計算の手間がかかる．

そこで，ヘッセ行列の第2項を無視して，アルゴリズム[NL]で$B_k = J(x_k)^T J(x_k)$とおいた解法が考えられる．これがGauss-Newton法である．この方法は，次の線形方程式の解を探索方向d_kとするものである

$$J(x_k)^T J(x_k) d = -J(x_k)^T r(x_k). \qquad (4.14)$$

一般に，Gauss-Newton法は残差がゼロもしくは非常に小さい問題に対しては効率がよいが，大きな問題に対しては効率が悪くなることがあることが知られている．これは，ヘッセ行列の第2項を無視したことによる．この問題を解決する方法として，SSP変換をGauss-Newton法に適用することが考えられる．このとき変換されてできあがった行列は，正定値対称性を保存し，かつ，ヘッセ行列$\nabla^2 f(x)$の第2項の情報をもっている．したがって，こうして修正された方法は$J(x_k)^T J(x_k)$よりもヘッセ行列の情報をより多

くもっており，その結果として，Gauss-Newton 法よりも効率が上がることが期待される．以上の目的のために，Dennis [4] によって導出された次のセカント条件を考えよう

$$B_k s_{k-1} = z_{k-1}, \tag{4.15}$$

ただし，

$$z_{k-1} = J(x_k)^T J(x_k) s_{k-1} + (J(x_k) - J(x_{k-1}))^T r(x_k). \tag{4.16}$$

この条件はテイラー展開の 1 次近似

$$\left(J(x_k)^T J(x_k) + \sum_{j=1}^{\nu} r_j(x_k) \nabla^2 r_j(x_k) \right) s_{k-1} \approx J(x_k)^T J(x_k) s_{k-1} + (J(x_k) - J(x_{k-1}))^T r(x_k)$$

から導かれる．

もしヤコビ行列 $J(x_k)$ が列フルランクならば，行列 $J(x_k)^T J(x_k)$ は正定値対称になるので，SSP 行列の一般形で $B_k = L_k^T L_k$ とおける．よって，(2.1) と (2.4) において

$$B_+ = B_k, \quad N_+ = N_k, \quad \Phi = J(x_k), \quad s = s_{k-1}, \quad z = z_{k-1}$$

とおけば，次式が得られる：

$$\begin{aligned} B_k &= SSP(J(x_k)^T J(x_k), s_{k-1}, z_{k-1}, u) \\ &= J(x_k)^T J(x_k) - \frac{J(x_k)^T p_{k-1} p_{k-1}^T J(x_k)}{\|p_{k-1}\|^2} + \frac{z_{k-1} z_{k-1}^T}{s_{k-1}^T z_{k-1}} + \|p_{k-1}\|^2 w_{k-1} w_{k-1}^T, \end{aligned} \tag{4.17}$$

$$L_k = J(x_k) - \frac{p_{k-1} u^T J(x_k)}{p_{k-1}^T u} + \frac{u z_{k-1}^T}{\sqrt{s_{k-1}^T z_{k-1}} \|u\|}, \tag{4.18}$$

ただし，$p_{k-1} = J(x_k) s_{k-1}$,

$$w_{k-1} = \frac{J(x_k)^T u}{p_{k-1}^T u} - \frac{J(x_k)^T p_{k-1}}{\|p_{k-1}\|^2}$$

である．ここで $u = p_{k-1}$ および $J(x_k)^T u = z_{k-1}$ とおけば，それぞれ，BFGS 型形式と DFP 型形式が得られる: 上記の BFGS 型形式と DFP 型形式は，それぞれ Al-Baali and Fletcher [1](もしくは Dennis et al. [5] も参照) によって提案された GN-BFGS 法と GN-DFP 法に相当する．

形式族 (4.17) を用いた解法に関する好ましい性質については，[10] を参照されたい．

5 おわりに

本稿では，安定な準ニュートン法というクラスを規定し，SSP 行列の一般形を導出することによって包括的な定義に成功した．

SSP 行列は，多くの示唆を含む形式をしている．例えば Broyden 公式族との密接な関係より，準ニュートン法の基本的な性質の体系化，あるいは解明に，上の一般形は大きな意味を持つものと期待される．4 章で示されたいくつかの応用の内，例えば，大規模な非線形最適化問題に対する準ニュートン公式の導出が，興味ある結果である．大規模問題に対して今後注目されるであろう手法の一つとして，スパース構造を持つ B_k の構成が挙げられる．そのとき本稿で提案された公式は，これまで困難とされていた正定値性とセカント条件の両立という意味で重要である．また，Limited memory 法が，このクラスの例として導出されることも，新しい手法の導出の方法論として注目される．

今後，ここで示されたクラスのより統一的体系化の研究が，準ニュートン法の発展に大きく寄与するであろう．

参考文献

[1] M.Al-Baali and R.Fletcher, Variational methods for non-linear least squares, *Journal of the Operational Research Society* 36 (1985) 405-421.

[2] C.G. Broyden, J.E. Dennis, Jr. and J.J. Moré , On the local and superlinear convergence of quasi-Newton methods, *Journal of the Institute of Mathematics and its Applications*, 12 (1973) 223-245.

[3] J.E. Dennis, Jr. and J.J. Moré , A characterization of superlinear convergence and its application to quasi-Newton methods, *Mathematics of Computation*, 28 (1974) 549-560.

[4] J.E.Dennis,Jr., A brief survey of convergence results for quasi-Newton methods, *SIAM-AMS Proceedings*, 9 (1976) 185-199.

[5] J.E.Dennis,Jr., S.Sheng and P.A.Vu, A memoryless augmented Gauss-Newton method for nonlinear least-squares problems, Technical Report 85-1, Department of Mathematical Sciences, Rice University, Texas, USA (1985).

[6] J.L.Nazareth, The Newton and Cauchy perspectives on computational nonlinear optimization. *SIAM Review*, 36 (1994) 215-225.

[7] J.Nocedal, Updating quasi-Newton matrices with limited storage, *Mathematics of Computation*, 35 (1980) 773-782.

[8] J.Nocedal, Theory of algorithms for unconstrained optimization, *Acta Numerica* (Cambridge University Press, Cambridge, 1992) 199-242.

[9] Ph.L.Toint, On sparse and symmetric matrix updating subject to a linear equation, *Mathematics of Computation*, 31 (1977) 773-782.

[10] H.Yabe and N.Yamaki, A modification of Gauss-Newton method for nonlinear least squares problems, *RIMS Kokyuroku*, 880 (1994) 202-210.

[11] H.Yabe , S.Takahashi and N.Yamaki, Secant updates of quasi-Newton methods for solving sparse nonlinear equations, *TRU Mathematics* , 19,1 (1983) 75-87.