OR 研究の最前線

準ニュートン法の研究とその展望

山下　信雄

1.　はじめに

本稿では，「OR 研究の最前線」の一環で，非線形計画の古典的手法のひとつである準ニュートン法の解説と著者からみた最近の話題の提供を行う．はじめから言い訳になるが，著者は，歴史ある準ニュートン法のすべてを知っているわけではない．また，準ニュートン法は，（著者のみるところ）非線形計画の研究におけるホットトピックではない．本稿の内容は，著者が準ニュートン法において勉強してきたことの「最前線」である．

本稿では，制約なし最小化問題：

$$\min_{x\in R^n} \quad f(x) \tag{1}$$

を考える．ここで，関数 f は R^n から R の 2 回連続的微分可能な関数とする．

準ニュートン法の説明に入る前に，準ニュートン法の現在おかれている状況，なぜ，非線形計画の主要な研究テーマとなっていないかについて説明する．そのためには，直線探索法と信頼領域法を知る必要がある．問題(1)に対する多くの手法は各反復で目的関数値を減じるように点列 $\{x_k\}$ を生成する降下法という枠組みに入る．降下法において，大域的収束[1] を保証する主な技法が，**直線探索法**と**信頼領域法**である[7]．両技法とも，通常は目的関数を近似したモデル関数

$$m_k(x_k+d)=f(x_k)+\nabla f(x_k)^T d+\frac{1}{2}d^T B_k d$$

を用いて，x_k から次の反復点 x_{k+1} への「方向 $\left(\frac{x_{k+1}-x_k}{\|x_{k+1}-x_k\|}\right)$」と「長さ ($\|x_{k+1}-x_k\|$)」を決める ($B_k=\nabla^2 f(x_k)$ とすると，m_k は f の 2 次近似関数となる)．直線探索法では，モデル関数 m_k の制約なしの最小解 d_k を「方向」とし，つぎに $f(x+td_k)<f(x_k)$ となるように「長さ（ステップ幅）」$t\in(0,1]$ を定める．なお，そのようなステップ幅が求まるためには，B_k が**正定値行列**となる必要がある．一方，信頼領域法では，まず，「長さ（信頼半径）」Δ_k を決め，次にその長さよりも小さいベクトルで，モデル関数を最小にする「方向」s_k を求め，$x_{k+1}=x_k+s_k$ とする．この s_k を求める問題 $\min_{\|s\|\le\Delta_k} m_k(x_k+s)$ は非凸な問題である．

非線形計画法の歴史において，初期のころは，直線探索法が主流であった．それは，当時はヘッセ行列の計算が容易ではなかったが，その一方で準ニュートン法によって正定値となる近似ヘッセ行列 B_k が計算できたからである．しかし，直線探索法では，B_k が正定値となる必要があるため，ヘッセ行列が計算できるときでも，$B_k=\nabla^2 f(x_k)$ とすることができなかった．そこで，いつでも $B_k=\nabla^2 f(x_k)$ とすることができる（よりよいモデル関数を使うことができる）信頼領域法が考案された．信頼領域法の部分問題は非凸であるが，効率のよい厳密解法および近似解法が提案されている[1]．また，そのような解法では $\nabla^2 f(x_k)$ の疎性を利用して，比較的大規模な問題でも，高速に解くことができる．そして，近年の自動微分の開発によりヘッセ行列の計算が容易になると，信頼領域法が主流とみなされるようになった．

準ニュートン法は，通常は直線探索法とともに用いられるため，（現在非主流の）直線探索法派に所属する．準ニュートン法では，モデル関数 m_k を定義する B_k に“近似”ヘッセ行列を用いる ($B_k=\nabla^2 f(x^k)$ としたものが純粋なニュートン法である)．準ニュートン法による探索方向は $d_k=-H_k\nabla f(x_k)$ で与えられる．ただし，H_k は B_k の逆行列である．たいていの準ニュートン法では，大域的収束性と高速性を両立するために，(i) **正定値行列**であり，(ii) **セカント条件**を満たす近似ヘッセ行列 B_k を用いる．B_k が正定値行列で

やました　のぶお
京都大学 大学院情報学研究科
〒606-8501 京都市左京区吉田本町

[1] ここでの大域的収束とは，$\lim_{k\to\infty}\nabla f(x^k)=0$ となることである．

あれば，ステップ幅 t_k を十分小さくとれば，$f(x_{k+1}) < f(x_k)$ とすることができる．(ii)のセカント条件とは，$B_{k+1}s_k = y_k$ となる性質である．ここで，$s_k = x_{k+1} - x_k$, $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ である．テーラーの定理より $\nabla f(x_{k+1})^2 s_k = y_k + O(\|s_k\|^2)$ となるため，セカント条件をみたせば，B_{k+1} はヘッセ行列のよい近似とみなせ，ニュートン法のような高速な収束が期待できる．条件(i)と(ii)を満たす B_k の更新法として，**DFP 更新**や**BFGS 更新**がある．しかし，これらの手法で更新された行列 B_{k+1} は密な行列となるため，大規模な問題には適用することができない（次節参照）．そのため，特に大規模な問題に対しては，BFGS 更新（あるいは DFP 更新）と直線探索法を組み合わせた手法は，ヘッセ行列をそのまま用いた信頼領域法に比べて，魅力の薄い手法とみなされている．準ニュートン法の研究においては，いかに大規模な問題，あるいは信頼領域法の不得手な問題を扱うかを考えることが重要となる．

本稿では，次節で，古典的な準ニュートン更新である DFP 更新と BFGS 更新を紹介し，その性質を述べる．次に，3 節において，大規模な問題にも適用することができる準ニュートン法をいくつか紹介する．さらに，4 節では，今後，準ニュートン法の研究において，重要となる（主流になれる）であろうテーマの私見を述べる．

2. 古典的な準ニュートン法

古典的な準ニュートン更新である DFP 更新および BFGS 更新では，次式で $\{H_k\}$ を更新する．

$$H_{k+1}^{\mathrm{DFP}} = H_k - \frac{H_k y_k (H_k y_k)^T}{y_k^T H_k y_k} + \frac{s_k s_k^T}{s_k^T y_k} \tag{2}$$

$$H_{k+1}^{\mathrm{BFGS}} = H_k - \frac{H_k y_k s_k^T + s_k (H_k y_k)^T}{s_k^T y_k} + \left(1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k} \tag{3}$$

B_k の更新式も同じように書けるがここでは省略する．H_k が正定値行列であり，$s_k^T y_k > 0$ のときには，H_{k+1} も正定値行列となり，セカント条件も満たす[7]．なお，Wolfe の基準[7]をみたすようにステップ幅を定めれば，$s_k^T y_k > 0$ は必ず満たされる．

DFP 更新や BFGS 更新は多くの教科書で紹介されているが，この更新規則がどのようにして導出されたかはほとんど知られていない．Fletcher[5]は DFP 更新による H_{k+1} は次の凸計画問題の解となることを示した[2]．

$$\begin{array}{ll} \min_H & \phi(H_k^{-1} H) \\ \text{subject to} & H y_k = s_k \\ & H = H^T,\ H \succeq 0 \end{array} \tag{4}$$

ただし，$\phi(A) = \mathrm{trace}(A) - \ln\det(A)$ である．行列 A の固有値を $\lambda_i, i=1,\cdots,n$ とすると，$\phi(A) = \sum_{i=1}^n (\lambda_i - \ln \lambda_i)$ となるから，$\phi(A)$ は A が単位行列のとき最小値をとる．$\phi(H_k^{-1}H)$ は，$H = H_k$ のとき最小となる狭義凸関数であるため，H と H_k との“距離”を表している．そのため，DFP 更新による H_{k+1} は，セカント条件を満たす正定値対称行列のうち，H_k からの“距離”が最小となるものと考えることができる．

ここで，DFP 更新と BFGS 更新の性質をふれておこう．まず，これらの更新規則によって生成された行列は，特殊な場合を除いて，真のヘッセ行列には収束しない．また，DFP 更新では，初期行列 B_0 が解 x^* におけるヘッセ行列 $\nabla^2 f(x^*)$ に十分近似していなければ大域的収束を保証することができない．そのため，DFP 更新では，収束が著しく遅くなることがある．一方，BFGS 更新では，適当な仮定のもとで，初期行列を任意にとっても，大域的収束および超一次収束することが示されている[7]．s_k や y_k にノイズが入るとき，B_{k+1} は理論とかけ離れたものになる．そのようなときでも，BFGS 更新では，何回か反復を繰り返すうちに，よりよい行列に戻ってくるようである．このことおよび数値実験の検証により，BFGS 更新は，DFP 更新に比べて，誤差に強く，また，収束が速いことが知られている．

一方，B_{k+1} は一般に密な行列となり，計算容量は $O(n^2)$ となる．そのため，大規模な問題には BFGS 更新（あるいは DFP 更新）は適用することができない．これが，古典的な準ニュートン法の最大の欠点である．

3. 少ない記憶容量で実装できる準ニュートン法

BFGS 更新の計算容量に関する欠点を克服する手法として，記憶制限つき BFGS 法（L-BFGS 法），partially separable BFGS 法（PSB 法），正定値行列補完を用いた準ニュートン更新がある．はじめの 2 つはすでに古典的といえる手法であるが，最後の手法は比較的新しいものである．

まず，L-BFGS 法を紹介する．BFGS 更新は，行列 $V_k = I - \frac{s_k y_k^T}{s_k^T y_k}$，$S_k = \frac{s_k s_k^T}{s_k^T y_k}$ を用いて，$H_{k+1} = V_k H_k (V_k)^T + S_k$ と表せる．ここで，行列 V_k, S_k に関

[2] BFGS 更新による B_{k+1} は，問題(4)において，H と B, s_k と y_k を入れ替えた問題の解となる．

する演算はベクトル s_k, y_k を用いて計算できるため，V_k, S_k を保持する必要はない．さらに，

$$\begin{aligned} H_{k+1} &= V_k H_k V_k + S_k \\ &= V_k V_{k-1} H_{k-1} V_{k-1} V_k + S_k + V_k S_{k-1} V_k \\ &\vdots \quad \vdots \end{aligned}$$

と展開できるため，任意の $0 \le m \le k$ となる m に対して，H_{k-m} と $m+1$ 個のペア $(s_{k-m}, y_{k-m}), \dots, (s_k, y_k)$ を用いて，H_{k+1} に関する演算を行うことができる．L-BFGS 法では，記憶するペア (s_i, y_i) の数 $m+1$ を制限し，上記の関係を用いて H_{k+1} に関する演算を行う．そのとき，H_{k-m} をそのまま用いれば，BFGS 更新と同じになるが，それでは行列 H_{k-m} を保持しなければならない．そこで H_{k-m} の代わりに対角行列，例えば $(\|s_k\|^2/y_k^T s_k)I$ を用いることによって，記憶容量を抑える．その結果，L-BFGS 法が必要とする記憶容量は $O(mn)$ ですむ．さらに，探索方向 $-H_k \nabla f(x^k)$ の計算も，V_k，S_k がベクトルの積で表現されていることを利用して，$O(mn)$ でできる．ただし，m が小さいときは取り込めるヘッセ行列の情報も少なくなるため，理論的には1次収束しか保証できない．実際に，数値実験においても，解におけるヘッセ行列の条件数が悪い問題においては，収束が遅くなることが報告されている．しかし，1回の反復にかかる計算時間が少ないため，高精度の解が必要とされないときには，十分実用的な解法である．

L-BFGS 法は，問題の構造を特に利用していない．次に，目的関数の構造を利用した準ニュートン法を紹介しよう．目的関数 f が

$$f(x) = \sum_{i=1}^{p} f_i(x_{C_i})$$

と表すことができるとき，f は partial separable であるという．ここで，$C_i, i=1, \dots, p$ は $\{1, 2, \dots, n\}$ の部分集合であり，f_i は $R^{|C_i|}$ から R への関数である．さらに，x_{C_i} は $x_i, i \in C_i$ を縦に並べた $|C_i|$ 次元ベクトルである．

PSB 法は，各 f_i の近似ヘッセ行列 $B_k^i \in R^{|C_i| \times |C_i|}$ を BFGS 更新で計算し，それらを足し合わせることによって，f の近似ヘッセ行列 B_k を生成する方法である[7]．一般に $|C_i| \ll n$ であることから，PSB 法は少ないメモリーで実装できる．また，f を分割した各 f_i に対する近似ヘッセ行列を用いるため，f に対してそのまま BFGS 更新をつかう場合と比べて，より近似精度のよい行列を得ることができる．一方，逆行列 H_k の更新式ではないため，探索方向を求めるために線形方程式 $B_k d = -\nabla f(x^k)$ を解かなければならない．また，f_i が凸関数でないときには，B_k^i が正定値行列にならないことがある．そのようなときは，B_k も正定値行列とならないため，信頼領域法と組み合わせて用いる必要がある．

最後に，著者が最近提案したヘッセ行列の疎構造を利用した準ニュートン法を紹介する．以下では，ヘッセ行列の疎構造を $E = \{(i, j) | \nabla^2 f(x)_{i,j} \ne 0 \text{ for some } x \in R^n\}$ とし，その拡張を $F \supset E$ とする．H_k にヘッセ行列と同等の疎性を持つようにするには，問題(4)に制約 $(H^{-1})_{ij} = 0, (i, j) \notin F$ を加えればよい[7]．しかし，この制約とセカント条件を満たした行列は数値的に不安定になることが知られている．そこで，文献[8]では次のようにして近似ヘッセ行列を更新することを提案している．

第1段階：既存の準ニュートン法（BFGS 更新等）で，$(i, j) \in F$ に対する H_{ij}^{QN} を計算する．

第2段階：$H_{ij}^{QN}, (i, j) \in F$ をパラメータにもつ次の最適化問題の解を H_{k+1} とする．

$$\begin{array}{ll} \min & \phi(H_k^{-1} H) \\ \text{subject to} & H_{ij} = H_{i,j}^{QN}, (i, j) \in F \\ & (H^{-1})_{ij} = 0, (i, j) \notin F \\ & H = H^T, H \succeq 0 \end{array} \tag{5}$$

問題(5)は，問題(4)に制約条件 $(H^{-1})_{ij} = 0, (i, j) \notin F$ を追加し，セカント条件の代わりに「セカント条件を満たした行列 H^{QN} と部分的に一致する」という条件に置き換えた問題である．この手法の実装において重要となるのは，問題(5)の解き方である．集合 F が特別な形をしているとき，問題(5)の解は陽に与えることができる．

まず，$\{1, \dots, n\}$ を頂点集合，$F \backslash \{(i, i) | i=1, \dots, n\}$ を辺集合としたグラフを G とする．グラフ G がコーダルグラフのとき，問題(5)は $H_{i,j}^{QN}$ の正定値行列補完問題

$$\begin{array}{ll} \max & \det(H) \\ \text{subject to} & H_{ij} = H_{i,j}^{QN}, (i, j) \in F \\ & H = H^T, H \succeq 0 \end{array} \tag{6}$$

と等価となり，その解は次のように陽に与えることができる[6]．$\{C_r\}$ を次の条件を満たすコーダルグラフ G の最大クリーク族とする．

$$\exists s > r \text{ such that } C_r \cap (C_{r+1} \cup C_{r+2} \cdots \cup C_l) \subset C_s.$$

さらに，$\{S_r\}, \{U_r\}$ を次のように定義する．

$$S_r = C_r \backslash (C_{r+1} \cup C_{r+2} \cup \cdots \cup C_l), \ r=1, \dots, l$$
$$U_r = C_r \cap (C_{r+1} \cup C_{r+2} \cup \cdots \cup C_l), \ r=1, \dots, l$$

このとき，問題(6)の解，つまり問題(5)の解 H_{k+1} は次式で与えられる[6]．

$$H_{k+1}=P^T L_1^T L_2^T \cdots L_l^T D L_l L_{l-1} \cdots L_2 L_1 P$$

ここで，P はある置換行列であり，$\{L_r\}$ と D は次式で定義される行列である．For $r=1, \dots, l-1$

$$[L_r]_{ij}=\begin{cases} 1 & i=j \\ \left((H^{QN}_{U_rU_r})^{-1}H^{QN}_{U_rS_r}\right)_{ij} & (i,j)\in U_r\times S_r \\ 0 & \text{otherwise} \end{cases}$$

$$D=\begin{pmatrix} D_{S_1S_1} & & \\ & \ddots & \\ & & D_{S_lS_l} \end{pmatrix}$$

$$D_{S_rS_r}=\begin{cases} H^{QN}_{S_rS_r}-H^{QN}_{S_rU_r}(H^{QN}_{U_rU_r})^{-1}H^{QN}_{U_rS_r} & r\leq l-1 \\ H^{QN}_{S_rS_r} & r=l \end{cases}$$

G がコーダルグラフとなるように $F\supset E$ を決めれば，H_{k+1} を疎な行列 L_r および D で表すことができる．さらに，実際の探索方向の計算は H_{k+1} を用いず，L_r および D を用いて行えるため，少ないメモリーと計算時間で，探索方向を計算することができる．例えば，ヘッセ行列が3重対角となる場合，H_{k+1} の更新および探索方向の計算は $O(n)$ でできる．

4．準ニュートン法の研究の展望

1節でも述べたように，最近では準ニュートン法（の研究）は，非線形計画に関する研究の主流ではないように思われる．しかし，計算技術の発達とともに，扱う問題が複雑かつ超大規模になってきている今日，準ニュートン法にも活躍の場がまた増えてきているようである．ここでは，準ニュートン法の活躍する状況や，そのために必要な技術に関する私見を述べたい．

4.1 超大規模な問題

超大規模な問題においては，信頼領域法の部分問題を解いたり，線形方程式を解くことは困難である．そのような問題に対して，現在の選択肢は，最急降下法あるいは共役勾配法と直線探索法の組合せである．このとき，計算のボトルネックとなるのは，目的関数およびその勾配の計算であり，そのような計算が少ない回数ですむ手法が良い手法と考えられる．ここでは，最近また注目を集めているBarzilai-Borwein法(BB法)[3]を紹介しよう．

最急降下法では，$d_k=-\nabla f(x_k)$ とするが，問題のスケールが考慮されていないため，適切なステップ幅を求めるために，多くの目的関数の評価が必要となる．そこで，$d_k=-(1/\alpha_k)\nabla f(x_k)$ として，α_k によって，スケーリングすることを考える．このスケーリングは $B_k=\alpha_k I$ とした“準ニュートン法”と考えることができるため，α_k を決めるのに，セカント条件を利用する．実際には，セカント条件を満たすことは不可能であるから，セカント条件がなるべく満たされるように $\|\alpha s_k-y_k\|^2$ を最小とする $\alpha_k=\frac{s_k^T y_k}{\|s_k\|^2}$ を用いて探索方向を計算する手法がBB法である．

超大規模な問題では，計算時間のボトルネックとなる目的関数の計算を省きたいため，直線探索（ステップ幅の計算）をしたくない．BB法における $1/\alpha_k$ はステップ幅と考えることもできる．しかし，各反復で $f(x_k+d_k)<f(x_k)$ が成り立つとは限らないので，$t_k=1$ のままでは大域的収束を保証することができない．そこで，非単調直線探索を用いる．非単調直線探索とは，各反復では，目的関数値の増加を許す代わりに，ある定数反復ごとに，必ず目的関数値が減少することを保証する手法である[7]．非単調直線探索を組み合わせたBB法は，共役勾配法に匹敵するようである．また，共役勾配法は誤差に弱いが，BB法はBFGS更新と同様に誤差に強いようである[3]．なお，L-BFGS法は，各反復で $s_k^T y_k>0$ を保証するために，BB法や共役勾配法と比較して，目的関数を評価する回数が多くなる傾向がある．

4.2 微分を使わない最適化法

応用問題においては，関数 f の値は計算（評価）できるが，その微分は解析的に与えられていない場合がある．例えば，f の関数値が複雑なシミュレーションや数値計算などで与えられる場合などである．そのような問題では，自動微分は利用することができない．そのため，目的関数値のみを用いた最適化手法，Derivative-Free Optimization (DFO)[2]を考える必要がある．DFOに関する研究の歴史は長いが，近年，より一層，注目を集めている．その理由は，計算機の性能向上に伴い，複雑なモデルのシミュレーションが高速に行えるようになり，シミュレーションによってモデルの評価を行うだけでなく，そのモデルの最適化が可能になってきているからである．

DFOとしては，古くから，有限差分近似による勾配を用いた準ニュートン法が考えられている．しかし，通常の有限差分近似では，各反復において，x_k のまわりで n 回関数評価をしなければならない．また，差分を常に微小の値に固定するため，関数にノイズが入るような問題には，有限差分近似による近似勾配は役にたたない．そこで，Kellyはimplicit filiterlingとよばれるsimplex gradientを用いた手法を提案し

ている．simplex gradient g は以下のように計算される．まず，$n+1$ 個の点 $x_k, x_k+y_1, \ldots, x_k+y_n$ とその関数値がわかっているものとしよう．このとき simplex gradient g は，

$$g=Y^{-1}\begin{pmatrix} f(x_k+y_1)-f(x_k) \\ \vdots \\ f(x_k+y_n)-f(x_k) \end{pmatrix}$$

で与えられる[2]．ただし，$Y=[y_1\cdots y_n]^\top$ である．なお，微小量 t を用いて，$Y=tI$ となるように y_i を選んだ simplex gradient は通常の有限差分近似である．

simplex gradient の計算にはすでに関数値が評価された点の情報を再利用することができるため，各反復で n 回関数評価をする必要がない．Implicit filiterling では，simplex gradient から BFGS 更新を使って近似ヘッセ行列を計算し，高速化をはかっている．このとき，BFGS 更新の誤差に対する頑健性から，実際の勾配を用いていないのにも関わらず，良好な収束性をもつようである．

4.3 問題構造の解析

疎構造や partial separable などの問題構造が事前にわかっていれば，それを利用した準ニュートン法は実用価値が高い．しかし，定式化の段階では疎構造を持たなかったり，あるいは，DFO のようにそもそも疎構造を調べることができない応用問題もある．そのような問題に対して，隠れている問題構造を解析する手法の研究が今後必要だと思われる．例えば，

$$f(x)=\left(\sum_{i=1}^{n} x_i\right)^4+\sum_{i=1}^{n}(x_i^2+x_i)$$

という目的関数を考えてみよう．この関数は第1項のせいで，疎構造をもたない．ここで，$x_1=z_1$，$x_i=z_i-z_{i-1}(i=2,\ldots,n)$ という変換をすると，

$$f(x)=\sum_{i=1}^{n-1}(2z_i^2-2z_iz_{i+1})+z_n^4+z_n^2-z_n$$

と表すことができる[4]．f を z の関数とみたとき，ヘッセ行列は3重対角行列となり，疎構造を利用した準ニュートン法などで高速に解くことができる．このような変換を自動的にするヒューリスティクスが，最近文献[4]で提案されている．しかし，この手法は計算時間がかかるため，超大規模な制約なし最小化問題には向いていない．準ニュートン法に適した疎構造を高速に発見する手法の開発が必要である．

また，微分を使わない最適化問題でふれたような問題においては，そもそも関数の形さえ特定されていない．そのようなときにでも，何らかの手法，例えば統計的な手法によって，疎構造が推定できれば，implicit filitering においても有効となるはずである．ただ，著者には具体的なアイデアがあるわけではない．

5. おわりに

本稿では，制約なし最小化問題に対する古典的な準ニュートン法から，疎構造を利用した準ニュートン法までを紹介した．

現在のところ，自動微分を用いたニュートン法＋信頼領域法に比べて，準ニュートン法と直線探索法を組み合わせた手法の研究は，分が悪いように思われる．ただし，すべての問題に対して万能な手法は存在しない．関数形が特定できない問題や超大規模な問題には，信頼領域法は向いていない．そのような問題に対しては，最後の節でふれたように，準ニュートン法（の発展形）の存在価値があるように思う．

なお，研究テーマとして準ニュートン法は現在の主流でないかもしれないが，（普通の）準ニュートン法は実装が容易なため，1,000 変数程度の問題や，他のプログラムのサブルーチンとして用いる場合，十分現役の手法であることを最後に明記しておく．

参考文献

[1] A. R. Conn, N. I.. M. Gould and Ph. L. Toint, Trust Region Methods, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2000.

[2] A. R. Conn, K. Scheinberg and L. N. Vicente, Introduction to Derivative-Free Optimization, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2009.

[3] R. Fletcher, *On the Barzilai-Borwein method*, Optimization and Control with Applications, Springer series in Applied Optimization 96, Springer-Verlag, pp. 235-256, 2005.

[4] S. Kim, M. Kojima and P. Toint, *Recognizing underlying sparsity in optimization*, Mathematical Programming, 119, pp. 273-303, 2009.

[5] R. Fletcher, *A new result for quasi-Newton formulae*, SIAM Journal on Optimization, 1, pp. 18-21, 1991.

[6] M. Fukuda, M. Kojima, K. Murota and K. Nakata, *Exploiting sparsity in semidefinite programming via matrix competion I : General framework*, SIAM J. Optimization, 11, pp. 647-674, 2001.

[7] J. Nocedal and S. J. Wright, Numerical Optimization, Springer-Verlag, New York, 1999.

[8] N. Yamashita, *Sparse quasi-Newton updates with positive definite matrix completion*, Mathematical Programming, 115, pp. 1-30, 2008.