統計数理（1998）
第46巻 第2号 317-334

特集「計算と最適化」
［原著論文］

半正定値計画問題に関する情報幾何を用いた考察

大阪大学大学院* 小 原 敦 美

（受付 1998年4月16日；改訂 1998年8月24日）

要 旨

本稿では半正定値計画（以下SDP）問題の性質を情報幾何の立場から論じる．SDP問題の許容領域を正定対称行列の部分多様体とみなし，情報幾何構造から導かれる第二基本形式と呼ばれるこの部分多様体の埋め込み曲率を考えると，この第二基本形式が小さいほどSDP問題をある種の内点法で解くのに必要な計算コストが少なくなることを示す．

特に第二基本形式がその許容領域上で0になるとき（この部分多様体を二重自己平行であるとよぶことにする），対応するSDP問題は任意の許容解がひとつわかればNewton反復なしで最適解を求めることができ，解の陽な公式も得ることができる．

またこのような性質を持つ正定対称行列の二重自己平行な許容領域は，対称行列のJordan部分代数と関係があることを示し，このことを用いて二重自己平行となるような許容領域の具体的な例のいくつかを構成する．

キーワード：半正定値計画問題，情報幾何，二重自己平行，Jordan代数，陽な解表示．

1. はじめに

近年，半正定値計画（Semidefinite Programming，以下SDP）問題と呼ばれる凸計画問題が盛んに研究されている．すでによく知られているように，この問題は信号処理，通信などを含むシステム制御工学に線形行列不等式アプローチとして広い応用をもち（Boyd et al.（1994）参照），最近では組み合わせ最適化などへの適用（小島（1997）参照）も活発である．またこの問題は線形計画（Linear Programming）問題の素直な拡張となっており，Karmarkar法を端緒にLP問題で研究された多項式時間解法である内点法のメカニズムが適用でき，求解のために必要な反復回数などが解析されている．

一方，Amari（1985），甘利・長岡（1993）らによる情報幾何は統計理論を中心として成果を上げているが，そこに現れる双対性の一側面はLegendre変換を通して現れるということからわかるようにもともと凸解析と関係が深い．特に凸計画問題の内点法に現れる手法やアルゴリズムは情報幾何の観点から興味深い解釈ができる．

本稿ではこの観点から考察をすすめることで以下の結果を得た．まず，考えているSDP問題の許容領域を正定対称行列集合の部分多様体と考え，主・双対両接続という2つの基準の意味で自己平行と呼ばれる性質（二重自己平行性）を持つとき，そのSDP問題がNewton反復無しで解けることを示す．この特別な場合として接空間がJordan代数と呼ばれるものを用いて代

* 基礎工学研究科 システム科学分野：〒560-8531 大阪府豊中市待兼山1-3.

数的に特徴づけられる部分多様体があらわれる．これらの場合の解の陽な表示も与える．また，両接続の意味で自己平行でなくとも，非許容領域での予測子-修正子法を実行することで，許容領域の双対接続に関してのある種の曲率（第二基本形式）が小さければバリアパラメータを大きく更新でき，結果的に全体の計算量が少なくなることを示す．

得られた結果では，計算量の評価が可能なSDP問題の中でもさらに計算量のかからない問題の目安として，双対接続の第二基本形式という幾何学的な量が意味を持ってくることを明らかにした点に工学的な意義があると思われる．またJordan代数という代数的な概念が関係することも内点法の数理的な側面を明らかにするという点で興味深い．なお，異なった角度からの内点法とJordan代数の関連についてはFaybusovich（1997）も参照されたい．

本稿の構成は，2節で一般の凸集合上の主パス追跡法と情報幾何について概説し，3節でNewton反復なしで解けるSDP問題のクラスを二重自己平行性で特徴づける．許容領域がもとの行列で見ても逆行列で見ても線形制約を持っていることがこのクラスの必要十分条件である．4節では，二重自己平行な許容領域の例の構成と最適解の陽な表示を与える．5節では，二重自己平行でない場合の予測子の誤差，バリアパラメータ更新の大きさ，双対接続に関する第二基本形式の三者間に成り立つ関係を導く．

2. 内点法と情報幾何に関する準備

凸計画問題は一般性を失うことなく次のようにある線形関数最小化問題に変換できる：

問題 C. 与えられたベクトル $c\in\boldsymbol{R}^m$ と開凸領域 $\mathscr{M}\subset\boldsymbol{R}^m$ に対して

$$\min c^T x\,,\qquad \text{s.t.}\quad x\in\overline{\mathscr{M}}\subset\boldsymbol{R}^m\,. \tag{2.1}$$

この最小化問題Cの解を求めるために凸領域$\mathscr{M}$で3回連続微分可能，Hesse行列が正定，$\mathscr{M}$の境界で発散するような関数 $\psi(x)$（バリア関数と呼ばれる）を選び，次のようなパラメータ付き最小化問題を考える．

問題 C′.

$$\min \Psi_t(x)\,,\qquad \text{s.t.}\quad x\in\overline{\mathscr{M}}\,,\quad \Psi_t(x):=tc^Tx+\psi(x)$$

この問題C′の最小解を $x(t)$，元の問題Cの最小解のひとつを x^* とすると，$x(t)$ は$\mathscr{M}$の内点を通って

$$x(t)\to x^*\qquad (t\to+\infty)$$

となる．実際には適当に離散化した増大列 $\{t_i\}$ に対して，$\Psi_{t_i}(x)$ の最小化の反復により各 $x(t_i)$ を生成し，十分大きな t_i に対する $x(t_i)$ を x^*（の近似解）として採用することになる．以上が内点法の中で主パス追跡法と呼ばれる方法の概略である．

Nesterov and Nemirovskii（1994）らは，この方法で $\psi(x)$ がself-concordant barrierと呼ばれるクラスに属するとき，$t_i\geq\vartheta/\epsilon$ とすれば $c^Tx(t_i)-c^Tx^*\leq\epsilon$ なる x^* の ϵ 解 $x(t_i)$ が求まることと，うまく t_i をアップデートすれば x^* の ϵ 解を求めるのに必要な反復回数の上界が評価できることを示した．ただし，ここで ϑ は $\psi(x)$ から求まるある定数である．$\mathscr{M}$の内点を通るこの曲線 $x(t)$ は中心曲線と呼ばれ他の内点法でも重要な役割を果たす．また$\mathscr{M}$を $\psi(x)$ のHesse行列をRiemann計量として持つRiemann多様体と考えたとき，中心曲線は凸計画問題Cの目的関数 c^Tx を最小化し，x^* に収束する次の勾配系

(2.2)
$$\dot{x}=-g^{-1}(x)c$$

の解 $x(t)$ のうちで，解析的中心と呼ばれる特別な初期値をとった場合と見なすことができる．ただし，g は (2.3) で与えられる Riemann 計量．この勾配系を解く内点法アルゴリズムはアファインスケーリング法と呼ばれる．

次に$\mathscr{M}$上で情報幾何構造を考えてみよう．これにはある条件を満たす Riemann 計量 g と2つの接続のペア ∇, ∇^* が$\mathscr{M}$上に定まれば良いのだが，ここでは $\psi(x)$ を用いて次のように定めることにする．まず g はその成分を

(2.3)
$$g_{ij}(x)=g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right):=\frac{\partial^2\psi}{\partial x^i\partial x^j}$$

とする．2つの接続 ∇, ∇^* の作用をその係数 Γ_{ijk}, Γ^*_{ijk} によってそれぞれ

(2.4)
$$\Gamma_{ijk}(x)=g\left(\nabla_{\partial/\partial x^i}\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}\right):=0\,,$$

(2.5)
$$\Gamma^*_{ijk}(x)=g\left(\nabla^*_{\partial/\partial x^i}\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}\right):=\frac{\partial^3\psi}{\partial x^i\partial x^j\partial x^k}$$

と定める．このように$\mathscr{M}$に Riemann 計量 g と接続のペア ∇, ∇^* を定めたことで導入された情報幾何構造を ($\mathscr{M}$, g, ∇, ∇^*) と表すことにする．

この情報幾何構造 ($\mathscr{M}$, g, ∇, ∇^*) から ∇-, ∇^*-測地線と呼ばれる$\mathscr{M}$上の曲線が次の微分方程式の解としてそれぞれ定義できる．

(2.6)
$$\sum_{j=1}^{m}g_{ij}\ddot{x}^j+\sum_{j,k=1}^{m}\Gamma_{ijk}\dot{x}^j\dot{x}^k=\sum_{j=1}^{m}g_{ij}\ddot{x}^j=0\,,\qquad i=1,\dots,m\,,$$

(2.7)
$$\sum_{j=1}^{m}g_{ij}\ddot{x}^j+\sum_{j,k=1}^{m}\Gamma^*_{ijk}\dot{x}^j\dot{x}^k=0\,,\qquad i=1,\dots,m\,,$$

$g=(g_{ij})$ は正則であるから，∇-測地線は座標 x で見たとき，直線になっていることに注意しよう．

主座標 x に対して双対座標と呼ばれる$\mathscr{M}$上の新しい座標 y (これに対して x を主座標と呼んでおく) を

(2.8)
$$y_i=\frac{\partial\psi(x)}{\partial x^i}\,,\qquad i=1,\dots,m\,,$$

とし，$\psi(x)$ の共役関数を

(2.9)
$$\psi^*(y)=\sup_{x\in\mathscr{M}}\left\{\sum_{i=1}^{m}x^iy_i-\psi(x)\right\}$$

と定める．

(2.8)，(2.9) をそれぞれ x, ψ の Legendre 変換と呼ぶ．(2.8) と g の定義から Legendre 変換の Jacobi 行列は g と一致する，すなわち

(2.10)
$$\frac{\partial y_i}{\partial x^j}=g_{ij}\,,\qquad i, j=1,\dots,m$$

となることに注意する．この関係より

$$(2.11)\qquad g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y_j}\right)=g\left(\frac{\partial}{\partial x^i}, \left(\sum_{k=1}^m g^{jk}\frac{\partial}{\partial x^k}\right)\right)=\delta^i_k, \qquad \text{ただし } (g^{ij})=(g_{ij})^{-1}$$

が得られ，主座標曲線 x^i と双対座標曲線 y_j は直交していることわかる．また y から x への逆 Legendre 変換は

$$(2.12)\qquad x^i=\frac{\partial\psi^*(y)}{\partial y_i}, \qquad i=1,\ldots, m,$$

となる．(2.10) より

$$\sum_{j=1}^m g_{ij}\dot{x}^j=\dot{y}_i$$

であるから，この両辺を t で微分して (2.7) を用いると

$$(2.13)\qquad \ddot{y}_i=0, \qquad i=1,\ldots, m.$$

従って，双対座標 y では ∇^*-測地線は直線として表されることが分かる．
　さて以上の準備のもとに，双対座標 y を用いると，$\Psi_t(x)$ に関する極値条件は

$$(2.14)\qquad y_i(t)=-c_i t, \qquad t>0, \quad i=1,\ldots, m$$

となり，中心曲線 $x(t)$ は ∇^*-測地線，すなわち双対座標では $y(t)$ は直線として表されることがわかる．さらに一般にアファインスケーリング法の勾配方程式 (2.2) も (2.10) を用いて双対座標へ変換すると

$$\dot{y}_i=-c_i, \qquad i=1,\ldots, m$$

となるから，次のようにまとめることが出来る．

命題 2.1. アファインスケーリング法の軌跡（以後 AS 軌跡）は，双対座標では $y(t)=ct+y(t_0)$, $t_0\leq t$ という半直線で表される．ただし，$y(t_0)$ は勾配方程式 (2.2) の初期値 $x(t_0)\in\mathscr{M}$ の双対座標である．

この命題は，Tanabe (1987)，Bayer and Lagarias (1989) などによって最初に指摘されたようである．この事実から，主座標で逐次 $x(t_i)$ を生成して中心曲線を追跡するために最適化を反復しなくとも，次のような手続きによって，x^* の ϵ 解が求まることが直ちにわかる．

手続き 2.2.
step 1.　任意の $\mathscr{M}$ の点 $x(t_0)$ に対してこれを $y(t_0)$ に Legendre 変換する．
step 2.　$t\geq\vartheta/\epsilon$ なる t に対して $y(t)=-ct+y(t_0)$ を計算する．
step 3.　$y^*(t)$ を主座標へ逆 Legendre 変換して，$x(t)$ を求める．

しかし一般に共役関数 $\psi^*(y)$ を陽な形で求めることが難しいため，step 3 で逆 Legendre 変換を行うには実際は y に対して (2.8) の非線形方程式を満たす x を求める必要がある．特にここで Newton 法を用いれば，この方法は結局は本質的にアファインスケーリング法と一致する．従ってこのアイデアによっても，一般には全体の計算量はアファインスケーリング法と少なくとも同程度はかかることになる．しかし次節で見るように，ある条件のもとでは逆 Legendre 変換に Newton 反復計算は不必要となる．

3. Newton 反復無しで解ける半正定値計画問題

以後，n 次の対称行列，正定値対称行列集合をそれぞれ $Sym(n)$，$PD(n)$ とあらわす．また $Sym(n)$ をベクトル空間と考えた時の次元を $N=n(n+1)/2$ とする．

与えられた $c\in\boldsymbol{R}^m$ と $E_i\in Sym(n), i=0,\dots, m$（但し $\{E_i\}_{i=1}^m$ は線形独立，$m<N$），に対して，SDP 問題は次のような問題である．

SDP 問題．

(3.1) $$\min_x c^T x, \qquad \text{s.t. } P(x)=E_0+\sum_{i=1}^m x^i E_i \geq O, \quad O \text{ は零行列}$$

この SDP 問題で

$$\mathcal{V}:=\mathrm{span}\{E_i\}_{i=1}^m, \qquad E_0+\mathcal{V}:=\{X|X-E_0\in\mathcal{V}\}$$

と表すことにする．$x\in\boldsymbol{R}^m$ は $P(x)\in E_0+\mathcal{V}$ と 1 対 1 対応であり，SDP 問題の $P(x)$ の許容領域は

(3.2) $$\mathcal{L}:=PD(n)\cap(E_0+\mathcal{V})$$

の閉包となる．$\mathcal{L}$ のバリア関数としては

$$\psi(x)=-\log\ \det\ P(x)$$

が self-concordant barrier となることが知られている．また前節に従って，$\psi(x)$ を用いて $\mathcal{L}$ の情報幾何構造（$\mathcal{L}$, g, ∇, ∇^*）が定義されるが，$\mathcal{L}$ 上で $\psi(x)$ の共役関数を陽な形で求めることはやはり難しい．従って逆 Legendre 変換が反復計算を必要とするという状況は前節の一般の場合と同じである．

ところが $\mathcal{L}$ を部分集合とする凸錐 $PD(n)$ 全体では共役関数が陽に計算でき，主，双対座標間の（逆）Legendre 変換も容易に行える（Ohara et al. (1996) 参照）．　具体的には $\{E_i\}_{i=m+1}^N$ を $\mathcal{V}$ の補空間の任意の基底行列とし，任意の対称行列 P を θ で

$$P=P(\theta):=\sum_{i=1}^N \theta^i E_i$$

と表す．この θ を $PD(n)$ の主座標と呼ぶことにする．一方，$\{E^j\}_{j=1}^N$ を

(3.3) $$-\mathrm{tr}\{E^i E_j\}=\delta_j^i, \qquad i, j=1,\dots, N$$

を満たす $Sym(n)$ の別の基底行列とすると

$$P=P(\eta):=\left(\sum_{i=1}^N \eta_i E^i\right)^{-1}$$

のように P を表すことができる．この η が θ に対する双対座標となる．すなわち $PD(n)$ 全体では Legendre 変換の本質的な部分は以下の公式のように P の逆行列の計算となる：

$$\eta_i=-\mathrm{tr}(E_iP^{-1})=-\mathrm{tr}\left(E_i\left(\sum_{i=1}^N \theta^i E_i\right)^{-1}\right),$$
$$\theta^i=-\mathrm{tr}(E^iP)=-\mathrm{tr}\left(E^i\left(\sum_{i=1}^N \eta_i E^i\right)^{-1}\right).$$

また $PD(n)$ のバリア関数

$$\psi(\theta)=-\log\ \det\ P(\theta)$$

に対して，その共役関数は

$$\psi^*(\eta) = -\log\ \det\ P(\eta)^{-1}$$

となる．

この $\psi(\theta)$ を用いて前節のように $PD(n)$ に定めた情報幾何構造を $(PD(n), g, \nabla, \nabla^*)$ と表す．θ と η の間にも (2.3) から (2.13) までと同様な関係が成立する．$(PD(n), g, \nabla, \nabla^*)$ は $\mathcal{L}$ 上では $(\mathcal{L}, g, \nabla, \nabla^*)$ に一致することは，$\psi(\theta)$ と $\psi(x)$ の定義から明らかであろう．

このように $\mathcal{L}$ を $PD(n)$ に埋め込んで考えることで，(逆) Legendre 変換に Newton 反復は必要なくなった代わりに，ϵ 解 $P(x)=P(\theta)$ を求める手続きが一般に容易ではなくなる．つまり，$\mathcal{L}$ 内部だけで考えているときは，手続き 2.2 では ϵ 解を求めるのに $\mathcal{L}$ の ∇^*-測地線として $\mathcal{L}$ の双対座標 y で初期点から初期方向へ直線を延ばすだけで足りていた．ところが $PD(n)$ に拡大して余分な次元を導入するとその部分は $\mathcal{L}$ の $PD(n)$ への埋め込まれ方に関係するので，一般には y 座標での直線は $PD(n)$ の ∇^*-測地線，すなわち η 座標で直線にならない．直感的には 2 次元球面の測地線 (例えば経線) が 3 次元 Euclid 空間の測地線 (直線) になっていないことと同じである．このことは次のように正確に言及できる．

補題 3.1. i) E_0 の θ 座標を $(\theta_0^1, \ldots, \theta_0^N)$ とする．$\mathcal{L}$ 上で，$\mathcal{L}$ の座標系 x と $PD(n)$ の座標系 θ の関係は

$$\theta^i = x^i + \theta_0^i, \qquad i=1, \ldots, m,$$
$$\theta^\kappa = \theta_0^\kappa, \qquad \kappa = m+1, \ldots, N.$$

ii) $\mathcal{L}$ 上で，$\mathcal{L}$ の座標系 y と $PD(n)$ の座標系 η の関係は

$$\eta^i = y^i, \qquad i=1, \ldots, m,$$
$$\eta^\kappa = \left.\frac{\partial \psi}{\partial \theta^\kappa}\right|_{\theta=(x^1+\theta_0^1, \ldots, x^m+\theta_0^m, \theta_0^{m+1}, \ldots, \theta_0^N)}, \qquad \kappa = m+1, \ldots, N.$$

iii) $\mathcal{L}$ の ∇^*-測地線である (2.2) の解 AS 軌跡は，θ 座標，η 座標でそれぞれ次の微分方程式の解である．

$$\dot{\theta} = -\begin{pmatrix} G_1^{-1}c \\ 0 \end{pmatrix}, \tag{3.4}$$

$$\dot{\eta} = -\begin{pmatrix} c \\ G_2^T G_1^{-1} c \end{pmatrix}. \tag{3.5}$$

ただし

$$G = (g_{pq}) := \left(\frac{\partial^2 \psi}{\partial \theta^p \partial \theta^q}\right) = \begin{pmatrix} G_1 & G_2 \\ G_2^T & G_3 \end{pmatrix},$$

で G_1，G_2，G_3 のサイズはそれぞれ $m \times m$，$m \times (N-m)$，$(N-m) \times (N-m)$ である．

証明. i)，ii) は座標系のとり方と $\mathcal{L}$ 及び $PD(n)$ の情報幾何構造の定義の仕方から明らか．iii) で (3.4) は，まず $\theta^i, i=1, \ldots, m$ に関しては i) と (2.2) より明らか．次に $\theta^\kappa, \kappa=m+1, \ldots, N$ に関しては AS 軌跡が $\mathcal{L}$ 上の曲線なので $\dot{\theta}^\kappa = 0,\ m+1 \le \kappa \le N$ より導かれる．さらに (3.5) は (2.10) と同様に $PD(n)$ でも，

$$\dot{\eta} = G\dot{\theta}$$

の関係が成立することから，(3.4) を座標変換して導かれる．□

上の結果 ii）あるいは iii）から，$\mathcal{L}$ 上の AS 軌跡を η 座標で見たとき η_κ, $\kappa=m+1,\dots,N$ の成分が一般に直線になっていない（なぜなら G_i, $i=1,2,3$ は η の関数）ことが理解できる．この曲がりは本質的には，部分多様体 $\mathcal{L}$ が $PD(n)$ の中で接続 ∇^* の基準でみて曲がっていることから起因するものである．このような一般の場合には，もはや初期点と初期方向ベクトルを与えただけでは AS 軌跡を $PD(n)$ の曲線として陽な解表示を与えるのは困難であり反復計算が必要になる．

しかしながら，$\mathcal{L}$ 上の AS 軌跡が $PD(n)$ の双対座標 η でみても直線（すなわち ∇^*-測地線）になっているような特殊な SDP 問題では，２節の手続きと同様に初期点から初期方向ベクトルの方向へ直線を伸ばすことで任意の t での η 座標での解表示を与えることが出来る．以上の考察をまとめると次のように結論できる：

命題 3.2. SDP 問題において，$\mathcal{L}$ 上のある点 P が得られているとする．その点 P を通る AS 軌跡が（$PD(n)$, g, ∇, ∇^*）の ∇^*-測地線にもなっている時，$\mathcal{L}$ 上の逆 Legendre 変換を $PD(n)$ 上で行うことで前節で述べた手続きによりその SDP 問題は Newton 反復無しで解ける．

次にどのような SDP 問題が上の命題のクラスに属するかを特徴づけることが問題になる．測地線を一本ずつ扱うのは難しいので，命題の条件を少し強めて次の条件を考える．

条件 3.3. $\mathcal{L}$ 上のすべての ∇^*-測地線が $PD(n)$ の ∇^*-測地線にもなっている．

このようにより強い条件を要求することで上の命題で述べたクラスよりも限定されるが，P は $\mathcal{L}$ 上の任意の点，すなわち許容解であれば良いことになる．∇^* は捩率 0 であることを考慮すると，条件 3.3 は $\mathcal{L}$ が ∇^* に関して自己平行（autoparallel）であることが必要十分条件となる (Kobayashi and Nomizu (1969))．以後この幾何学的な条件を ∇^*-自己平行と呼ぶことにして，この条件を行列の言葉で書き直すと以下のようになる．

定理 3.4. $\mathcal{L}$ 上の点が少なくとも一つわかっているとする．この時 SDP 問題の許容領域 $\mathcal{L}$ が ∇^*-自己平行，すなわち

$$E_iP^{-1}E_j+E_jP^{-1}E_i\in\mathcal{V}, \qquad \forall E_i, E_j, \quad i,j=1,\dots,m, \quad \forall P\in\mathcal{L} \tag{3.6}$$

ならば，その SDP 問題は Newton 反復無しで解ける．

証明. $\mathcal{L}$ が ∇^*-自己平行であることの定義は，$\mathcal{L}$ の接ベクトル場 $\partial_i(=\partial/\partial x^i)$, $i=1,\dots,m$ に対して任意の点 $P\in\mathcal{L}$ で $PD(n)$ の接続 ∇^* による共変微分 $\nabla^*_{\partial_i}\partial_j|_P$ が P での $\mathcal{L}$ の接ベクトル空間 $T_P\mathcal{L}$ に含まれる，すなわち

$$\nabla^*_{\partial_i}\partial_j|_P\in T_P\mathcal{L}, \qquad \forall P\in\mathcal{L}, \quad i,j=1,\dots,m$$

が成立することである（Kobayashi and Nomizu（1969））．ただし，$\partial_i|_P$, $T_P\mathcal{L}$ などの $_P$ は点 P における接ベクトルや接ベクトル空間であることを表す．

一方，対称行列 $E_i\in Sym(n)$ と $PD(n)$ の接ベクトル $\partial_i|_P\in T_PPD(n)$ を同一視できる（これを $\partial_i|_P\equiv E_i$ と表すことにする）ことを用いて，前節で情報幾何構造を導入した方法にしたがって $\nabla^*_{\partial_i}\partial_j|_P$ を計算すると

$$\nabla^*_{\partial_i}\partial_j|_P\equiv -E_iP^{-1}E_j-E_jP^{-1}E_i$$

と行列表現できる (Ohara et al. (1996)).

$\mathcal{L}$の定義より $T_P\mathcal{L}=\mathrm{span}\{E_i\}_{i=1}^m=\mathcal{V}$であるから, $\mathcal{L}$が∇^*-自己平行となる必要十分条件として (3.6) が得られる. □

注. SDP問題では, $\mathcal{L}$はその作り方からもともと∇-自己平行である(すなわち$\mathcal{L}$の座標xでみた任意の直線がそのまま$PD(n)$の座標θでも直線になっている)し, 逆に任意の∇-自己平行な$PD(n)$の部分多様体は適当な$E_i, i=1,\ldots,p$で (3.2) のように作れる. 従って, 定理の条件 (3.6) は$\mathcal{L}$が∇, ∇^*の両接続の意味で自己平行であることを意味している. 以後 (3.6) を満たす$\mathcal{L}$を**二重自己平行 (doubly autoparallel)** と呼ぶことにする.

定理3.4の条件 (3.6) はすべてのPでチェックする必要があるため, 実用上は判定が難しい条件である. しかし次のような特殊な状況では, 簡単な代数的な条件で判定可能となる.

定理 3.5. $E_0\in\mathcal{V}$で$I\in\mathcal{L}$とする. この時$\mathcal{L}$が∇^*-自己平行であることの必要十分条件は$T_P\mathcal{L}$ $(=\mathcal{V})$ が$Sym(n)$の Jordan 部分代数となる, すなわち$\mathcal{V}$が$Sym(n)$の部分ベクトル空間でかつ

$$E_i * E_j := (E_iE_j+E_jE_i)/2\in\mathcal{V}, \qquad \forall E_i, E_j, \quad i, j=1,\ldots, m$$

が満たされることである.

証明. $E_0\in\mathcal{V}$なので$\mathcal{L}\subset\mathcal{V}$であることに注意する. A, $B\in T_P\mathcal{L}=\mathcal{V}$とすると$\mathcal{L}$が$\nabla^*$-自己平行の時, すべての$P\in\mathcal{L}$に対して

$$AP^{-1}B+BP^{-1}A\in\mathcal{V}$$

となるので, $P=I$とすると$\mathcal{V}$が$Sym(n)$の Jordan 部分代数であることがわかる.

逆に$\mathcal{V}$を Jordan 部分代数とする (但し, $A*B=(AB+BA)/2$としている). $P\in\mathcal{L}$, A, $B\in\mathcal{V}$に対して

$$A*(A*B)=\frac{1}{4}(A^2B+BA^2+2ABA)\in\mathcal{V}$$

と

$$(A*A)*B=\frac{1}{2}(A^2B+BA^2)\in\mathcal{V}$$

から$ABA\in\mathcal{V}$. これと$I\in\mathcal{L}$であればP^{-1}が再び$\mathcal{V}$に, したがって$\mathcal{L}$に含まれるという Jordan 部分代数の性質 (例えば Malley (1994) 参照) から

$$\begin{aligned}(A+B)P^{-1}(A+B)-AP^{-1}A-BP^{-1}B\\ =AP^{-1}B+BP^{-1}A\in\mathcal{V}.\end{aligned}$$

これは$\mathcal{L}$が∇^*-自己平行であることを意味する. □

注. このような$\mathcal{L}$として, 二重対称行列や固有ベクトルが固定された正定対称行列がある. $Sym(n)$の Jordan 部分代数については例えば(Malley (1994))やその参考文献を参照されたい.

4. Newton 反復なしで解けるSDP問題のクラスについての考察

3節ではSDP問題の許容領域$\mathcal{L}$が二重自己平行であるなら，そのSDP問題はNewton反復なしで解けることを示した．

またその一例として$E_0 \in \mathcal{V}$で$\mathcal{V}$が$Sym(n)$のJordan部分代数になっているものがあることを示した．しかしこの場合許容領域$\mathcal{L} = \mathcal{V} \cap PD(n)$は$PD(n)$の部分錐となるので，目的関数$c^T x$の$c$が特別な方向でないと最適解$P(x^*)$は原点か，解が存在しないかであるようなつまらない問題となる．

そこでここでは，部分錐とならないようなより一般的な二重自己平行な$\mathcal{L}$の例を構成し，Newton反復なしで解ける非自明なSDP問題のクラスが存在することを示す．ここで求めたクラスのひとつは，（Vanderbei and Yang (1995)）らが全く別な考えに基いて求めていた，やはりNewton反復なしで解けるSDP問題のクラスと一致することがわかる．最後に$\mathcal{L}$が二重自己平行の場合の陽な解表示を与える．

4.1 部分錐にならないような二重自己平行な$\mathcal{L}$のクラス

前節までに述べてきたように座標系を取ったとき，∇-(∇^*-)自己平行性は，それぞれx (y)座標での直線がθ (η)座標でも直線になっていることである．したがって$\mathcal{L}$が二重自己平行であることは適当な対称行列$\{E_i\}_{i=0}^m$，$\{F^i\}_{i=0}^m$があって

$$\mathcal{L} = \left\{ P \middle| P = E_0 + \sum_{i=1}^m x^i E_i > O,\ P^{-1} = F^0 + \sum_{i=1}^m y_i F^i \right\}$$

となることと同値である．

このように逆行列の要素の間にも線形制約が成立するような正定対称行列はサンプル共分散行列からの真の共分散行列の最尤推定が凸計画問題となるクラスにもなっているので，統計学，信号処理でも興味深い対象であると思われる．特に部分錐になる（$E_0 \in \mathrm{span}\{E_i\}_{i=1}^m$）場合はJordan部分代数を用いて調べられてきたが（Malley (1994) 及びその参考文献参照)，それ以外の場合については筆者の知る限りないようである．

以下，$Sym(n)$のJordan部分代数でn次単位行列を含むものをまとめて$\mathcal{JS}(n)$，行列サイズを問題にしないときは省略して単に$\mathcal{JS}$とあらわす．$\mathcal{JS}$の具体例としては，$Sym(n)$自身や一次元の部分空間という自明なもの以外に二重対称行列（主対角と反主対角に対して対称な行列）集合や固有ベクトルが固定された対称行列集合など様々なものがあるが，$\mathcal{JS}$と表したときにはそのうちの任意のいずれかを指していることにする．

以下の仮定を設ける：

仮定． 仮定として，変数行列Xは$\mathcal{JS}$の任意の要素に値をとれる，すなわち$\{J_i\}_{i=1}^p$を$\mathcal{JS}$の基底行列として

$$X = \sum_{i=1}^p x^i J_i, \qquad \text{s.t. } \mathrm{span}\{J_i\}_{i=1}^p = \mathcal{JS},\ p = \dim \mathcal{JS}$$

とする．

例 1. Xを変数行列で上記の仮定を満たすとする．A，Bを定数行列とした時，

$$\mathcal{A}_1 := \left\{ \begin{pmatrix} X & A \\ A^T & B \end{pmatrix} \middle| \det B \neq 0,\ AB^{-1}A^T \in \mathcal{JS} \right\}$$

とする．$\mathcal{L}_1 := \mathcal{A}_1 \cap PD(n)$は空集合でないなら二重自己平行である．

証明. $P\in\mathcal{L}_1$ とすると

$$P^{-1}=\begin{pmatrix} X & A \\ A^T & B \end{pmatrix}^{-1}$$
$$=\begin{pmatrix} (X-AB^{-1}A^T)^{-1} & (X-AB^{-1}A^T)^{-1}AB^{-1} \\ -B^{-1}A^T(X-AB^{-1}A^T)^{-1} & B^{-1}+B^{-1}A^T(X-AB^{-1}A^T)^{-1}AB^{-1} \end{pmatrix}>O$$

となるので，仮定より $X-AB^{-1}A^T\in\mathcal{JS}\cap PD(n)$ となる．したがって定理3.5より $X-AB^{-1}A^T$ の集合全体は二重自己平行となり，$(X-AB^{-1}A^T)^{-1}$ はある定数対称行列の線形結合で表される．したがって P^{-1} 全体も適当な定数対称行列のアファイン結合で表す事ができる．□

例 2. X を変数行列で上記の仮定を満たすとする．A，B を定数行列とした時，

$$\mathcal{A}_2:=\{A-BXB^T|\det A\neq 0,\ B^TA^{-1}B\in\mathcal{JS}\}$$

とする．このとき，$\mathcal{L}_2:=\mathcal{A}_2\cap PD(n)$ は空集合でないなら二重自己平行である．

証明. 逆行列に関するよく知られた以下の公式

$$(A-LXR)^{-1}=A^{-1}+A^{-1}L(X^{-1}-RA^{-1}L)^{-1}RA^{-1} \tag{4.1}$$

を用いると

$$(A-BXB^T)^{-1}=A^{-1}+A^{-1}B(X^{-1}-B^TA^{-1}B)^{-1}B^TA^{-1}$$

となる．$X^{-1}\in\mathcal{JS}$ となる（例えば Malley (1994) 参照）ことから $X^{-1}-B^TA^{-1}B\in\mathcal{JS}$ であり，さらに $(X^{-1}-B^TA^{-1}B)^{-1}\in\mathcal{JS}$ となる．よって $(A-BXB^T)^{-1}$ は適当な定数対称行列のアファイン結合で表す事ができる．□

例1のクラスは例2のクラスの特殊な例となっていることが次のように分かる．

$$\begin{pmatrix} X & A \\ A^T & B \end{pmatrix}=\begin{pmatrix} C & A \\ A^T & B \end{pmatrix}-\begin{pmatrix} I \\ O \end{pmatrix}(C-X)(I\quad O)$$

と書き直して

$$\tilde{A}:=\begin{pmatrix} C & A \\ A^T & B \end{pmatrix},\qquad \tilde{B}:=\begin{pmatrix} I \\ O \end{pmatrix},\qquad \tilde{X}:=C-X$$

とおき，定数行列 $C\in\mathcal{JS}$ を $\tilde{A}$ の正則性条件を満たすように適当にとってやれば例2の特別な場合となる．

例2をさらに一般化して次の命題を得る．

命題 4.1. 上記の例の仮定と同様に，$i=1,\dots,q$ について，X_i 全体の集合 $\{X_i\}$ は $\mathcal{JS}_i$ に一致すると仮定する．ただし，ここで $\mathcal{JS}_i$, $i=1,\dots,q$ はそれぞれ異なる Jordan 部分代数であってもよい．

$$\mathcal{A}_3:=\left\{A-\sum_{i=1}^{q}B_iX_iB_i^T\ \middle|\ \det A\neq 0,\ B_i^TA^{-1}B_i\in\mathcal{JS}_i,\ B_i^TA^{-1}B_j=O\ (i\neq j)\right\}$$

とする．このとき，$\mathcal{L}_3:=\mathcal{A}_3\cap PD(n)$ は空集合でないなら二重自己平行である．

証明. $P_n:=A-\sum_{i=1}^{n}L_iX_iR_i$ とする．一般に，$R_iA^{-1}L_j=O\ (i\neq j)$ の時，

$$P_n^{-1}=\left(A-\sum_{i=1}^{n}L_iX_iR_i\right)^{-1}=A^{-1}+\sum_{i=1}^{n}A^{-1}L_i(X_i^{-1}-R_iA^{-1}L_i)^{-1}R_iA^{-1}$$

とあらわせる事を証明する．まず，$n=1$ の時

$$\begin{aligned}P_1^{-1}&=(A-L_1X_1R_1)^{-1}\\&=A^{-1}+A^{-1}L_1(X_1^{-1}-R_1A^{-1}L_1)^{-1}R_1A^{-1}\end{aligned}$$

となり，成立している．次に，$n=k$ の時に成立していると仮定する．すなわち，

$$P_k^{-1}=\left(A-\sum_{i=1}^{k}L_iX_iR_i\right)^{-1}=A^{-1}+\sum_{i=1}^{k}A^{-1}L_i(X_i^{-1}-R_iA^{-1}L_i)^{-1}R_iA^{-1},$$

ただし，

$$R_iA^{-1}L_j=O \qquad (i\neq j).$$

$n=k+1$ の時に

$$\begin{aligned}P_{k+1}^{-1}&=\left(A-\sum_{i=1}^{k+1}L_iX_iR_i\right)^{-1}\\&=\left(A-\sum_{i=1}^{k}L_iX_iR_i-L_{k+1}X_{k+1}R_{k+1}\right)^{-1}\\&=(P_k-L_{k+1}X_{k+1}R_{k+1})^{-1}\\&=P_k^{-1}+P_k^{-1}L_{k+1}(X_{k+1}^{-1}-R_{k+1}P_k^{-1}L_{k+1})^{-1}R_{k+1}P_k^{-1}\\&=P_k^{-1}+\left(A^{-1}+\sum_{i=1}^{k}A^{-1}L_i(X_i^{-1}-R_iA^{-1}L_i)^{-1}R_iA^{-1}\right)\\&\quad\times L_{k+1}(X_{k+1}^{-1}-R_{k+1}P_k^{-1}L_{k+1})^{-1}R_{k+1}\\&\quad\times\left(A^{-1}+\sum_{i=1}^{k}A^{-1}L_i(X_i^{-1}-R_iA^{-1}L_i)^{-1}R_iA^{-1}\right).\end{aligned}$$

ここで，$R_iA^{-1}L_j=O$ より，$\sum_{i=1}^{k}R_iA^{-1}L_{k+1}=O$，$\sum_{i=1}^{k}R_{k+1}A^{-1}L_i=O$．これを代入して，

$$\begin{aligned}P_{k+1}^{-1}&=P_k^{-1}+A^{-1}L_{k+1}(X_{k+1}^{-1}-R_{k+1}A^{-1}L_{k+1})^{-1}R_{k+1}A^{-1}\\&=A^{-1}+\sum_{i=1}^{k}A^{-1}L_i(X_i^{-1}-R_iA^{-1}L_i)^{-1}R_iA^{-1}\\&\quad+A^{-1}L_{k+1}(X_{k+1}^{-1}-R_{k+1}A^{-1}L_{k+1})^{-1}R_{k+1}A^{-1}\\&=A^{-1}+\sum_{i=1}^{k+1}A^{-1}L_i(X_i^{-1}-R_iA^{-1}L_i)^{-1}R_iA^{-1}.\end{aligned}$$

以上より帰納法により

$$P_n^{-1}=\left(A-\sum_{i=1}^{n}L_iX_iR_i\right)^{-1}=A^{-1}+\sum_{i=1}^{n}A^{-1}L_i(X_i^{-1}-R_iA^{-1}L_i)^{-1}R_iA^{-1}.$$

ここで $L_i=B_i$, $R_i=B_i^T$ として，例 2 と同様に考えれば $\mathcal{L}_3$ が二重自己平行であることが導かれる．□

注. Vanderbei and Yang (1995) らは，

$$\min\quad \mathrm{tr}(CX),\qquad \text{s.t. } MXM^T=B,\quad M:\text{full rank},\quad X\geq O$$

という SDP 問題が QR 分解を用いて Newton 反復なしで解けることを明らかにしている．この SDP 問題の双対問題は

$$\max \quad \mathrm{tr}(BY), \qquad \mathrm{s.t.}\ C - M^T YM = Z > O$$

であり，例2の$\mathcal{L}_2$で$\mathcal{JS}$を対称行列全体にとったときに対応する（Cが非正則でも主問題の目的関数を定数分だけ変更して正則なCにとりかえることができることに注意）．本稿と Vanderbei and Yang (1995) らの考え方は全く異なり，その関係は興味深いがまだ明らかではない．

同様に Wolkovicz (1993) も別の SDP 問題のクラスがやはり Newton 反復なしで解けることを報告している．そのクラスと本稿で述べたクラスの包含関係は現在のところ明らかでない．

4.2 二重自己平行な場合の解の陽な表示

最後に，(3.2) で与えられた許容領域$\mathcal{L}$が二重自己平行な場合の最適解 $P(x^*)$ の陽な表示を導出する．まず$\mathcal{L}$上の与えられた点を $P(x_0)$ とする．$\mathcal{L}$の二重自己平行性より任意の $P\in\mathcal{L}$はある $\{F^i\}_{i=0}^m$ を用いて

$$P^{-1} = F^0 + \sum_{i=1}^m z_i F^i$$

と表される．このうち $\{F^i\}_{i=1}^m$ は $PP^{-1}=I$ を $P(x_0)$ で微分することで計算できて

$$F^i = -P(x_0)^{-1} E_i P(x_0)^{-1}$$

とすればよい．また $F^0 = P(x_0)^{-1}$ とおけばよい．

次に初期点 $P(x_0)$ から行列C方向へ η 座標で直線 (∇^*-測地線) を伸ばして θ 座標へ変換することは，3節で述べたことより行列の計算としては

$$\lim_{t\to+\infty} (P(x_0)^{-1} - Ct)^{-1} \tag{4.2}$$

を計算することになる．ここでCは

$$c^T x = \mathrm{tr}\, CP(x) + \mathrm{constant}, \qquad C \in \mathrm{span}\{F^i\}_{i=1}^m \tag{4.3}$$

となる行列で，線形方程式

$$H\tilde{c} = c, \qquad H := (\mathrm{tr}(E_i F^j))$$

の解 $\tilde{c} = (\tilde{c}_i)$ を用いて（H は常に正則）

$$C = \sum_{i=1}^m \tilde{c}_i F^i$$

と計算される．(4.2) を計算するために

$$C = (V_1 \quad V_2)\begin{pmatrix} \Sigma_1 & O \\ O & O \end{pmatrix}\begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = V_1 \Sigma_1 V_1^T$$

と特異値分解する．これを (4.2) に代入して整理すると

$$\begin{aligned} P(x^*) &= \lim_{t\to+\infty} (P(x_0)^{-1} - Ct)^{-1} \\ &= \lim_{t\to+\infty} (P(x_0) + P(x_0) V_1 \{t^{-1}\Sigma_1^{-1} - V_1^T P(x_0) V_1\}^{-1} V_1^T P(x_0)) \\ &= P(x_0) - P(x_0) V_1 \{V_1^T P(x_0) V_1\}^{-1} V_1^T P(x_0) \end{aligned}$$

が得られ，これが解の陽な表示となる．x^* の各要素は (3.3) の関係から

$$x^{*i} = -\mathrm{tr}(E^i P(x^*))$$

と求まる．

5. 一般の半正定値計画問題の解析

5.1 非許容領域を通る予測子–修正子法

3節で見たように$\mathcal{L}$が二重自己平行でない場合には，一般に$G_2^T G_1^{-1} c$が定数ベクトルでなくなり$\mathcal{L}$のAS軌跡は$PD(n)$に埋め込まれるとη座標でみて曲線になるのであった．従ってこの場合にはAS軌跡上の点からその接方向へη座標の意味での直線（$PD(n)$の∇^*-測地線）を伸ばすと，伸ばすにつれてその線形近似直線は（局所的には）AS軌跡から離れ誤差は大きくなる．

この問題に対処するためには，誤差がある大きさ以上になるくらいtを伸ばしたら，これをAS軌跡の方向へ引き戻して修正するという手続きを繰り返すことでAS軌跡を追跡する方法が考えられる．よく用いられる予測子–修正子法であるが，予測子は$\mathcal{L}$からの外に出てしまうので非許容領域を通るアルゴリズムとなる．

本節では，このようなη座標での予測子–修正子法によりAS軌跡を追跡するアルゴリズムについて基本的な部分を解析する．その結果パラメータδtだけ伸ばしたときの予測誤差が，$\mathcal{L}$の接続∇^*に関する第二基本形式（the second fundamental form）と呼ばれる曲がり方を表す量と関係することがわかる．接続∇^*の第二基本形式が0となるとき$\mathcal{L}$はちょうど∇^*-自己平行である（Kobayashi and Nomizu (1969)）．この場合にはδtを無限大にしてよいという結果が得られ，3節の結果を含んだものとなっている．

$\mathcal{L}$上の任意の初期点$P_0 \in \mathcal{L}$（そのη座標を$\eta_0 = (\eta_{0i})$と表す）が与えられていると仮定し，P_0を通るAS軌跡$P(t)$を追跡する．

まず予測子について述べる．AS軌跡の従う微分方程式は補題3.1から

$$\dot{\eta} = -\begin{pmatrix} c \\ G_2^T(\eta) G_1^{-1}(\eta) c \end{pmatrix}, \qquad \eta(t_0) = \eta_0 \tag{5.1}$$

のようになるが，この解$P(t)$とP_0での線形近似$\hat{P}(t)$との誤差を考えよう．$\hat{P}(t)$は以下の微分方程式の解である：

$$\dot{\eta} = -\begin{pmatrix} c \\ \underline{c} \end{pmatrix}, \qquad \eta(t_0) = \eta_0 \tag{5.2}$$

ただし，$\underline{c}$は以下の定数ベクトル

$$\underline{c} = (c_\kappa) := G_2^T(\eta_0) G_1^{-1}(\eta_0) c\,, \qquad \kappa = m+1, \ldots, N \tag{5.3}$$

としている．従って，パラメータtをδtだけ増やした予測子$\hat{P}(t+\delta t)$のη座標は以下のように求まる．

$$\hat{\eta}(t_0 + \delta t) = \eta_0 - \begin{pmatrix} c \\ \underline{c} \end{pmatrix} \delta t\,, \tag{5.4}$$

または行列で書くとη座標の定義より

$$\hat{P}^{-1}(t_0 + \delta t) = P_0^{-1} - C \delta t\,, \tag{5.5}$$

と表される．ただし，Cは

$$C = \sum_{i=1}^{N} c_i E^i\,, \qquad \text{ここで}(c_i)\text{は}(c^T \quad \underline{c}^T)^T\text{の各要素} \tag{5.6}$$

と表される定数行列である．

次に，修正子について述べる前に次の部分多様体を導入しておく．

定義 5.1. $\underline{\eta}:=(\eta_\kappa)$, $\kappa=m+1,\ldots,N$ とする．与えられた $E^0\in PD(n)$ とその η 座標 $(\eta_i^0;\eta_\kappa^0)$, $i=1,\ldots,m$, $\kappa=m+1,\ldots,N$ の一部を用いて，$\mathcal{L}^\perp(E^0)\subset PD(n)$ を次のように定める．

$$\mathcal{L}^\perp(E^0):=\left\{P(\underline{\eta})\middle|P^{-1}(\underline{\eta})=(E^0)^{-1}+\sum_{\kappa=m+1}^{N}(\eta_\kappa-\eta_\kappa^0)E^\kappa,\ P(\underline{\eta})\in PD(n)\right\}.$$

$\underline{\eta}$ はこの部分多様体の座標系になっていることに注意する．$\mathcal{L}^\perp(E^0)$ は逆行列の空間で $\mathcal{L}$ と同様なアファイン制約を持つ集合である．この部分多様体 $\mathcal{L}^\perp(E^0)$ のいくつかの性質については (Ohara et al. (1996)) を参照．

修正子は次の最適化問題を解くことで得られる．

補題 5.2. 予測子 $P^+:=\widehat{P}(t_0+\delta t)$ に対する (AS 軌跡上の厳密な) 修正子 $P(t_0+\delta t)$ は，次の最小化問題の唯一解である：

$$P(t_0+\delta t)=\arg\min_{P\in\mathcal{L}^\perp(P^+)} D(P_0,P), \tag{5.7}$$

ただし，$D(\cdot,\cdot): PD(n)\times PD(n)\to \boldsymbol{R}$ はダイバージェンスと呼ばれる量で次のように定義される．

$$D(P_1,P_2):=\log\ \det\ P_2-\log\ \det\ P_1+\mathrm{tr}(P_2^{-1}P_1)-n \tag{5.8}$$

証明. Ohara et al. (1996) 参照. □

予測子 P^+ の η 座標を $(\eta_i^+;\eta_\kappa^+)$, $i=1,\ldots,m$, $\kappa=m+1,\ldots,N$, 初期点 P_0 の θ 座標を $(\theta_0^i;\theta_0^\kappa)$, $i=1,\ldots,m$, $\kappa=m+1,\ldots,N$ とする．$\mathcal{L}^\perp(P^+)$ 上の P を定義 5.1 のように $\underline{\eta}$ で表し，P_0 が定数行列であることに注意すると，上の最小化問題は次のような変数 $\underline{\eta}$ の凸最適化問題となる：

[修正子を求める最適化問題]

$$\eta(t_0+\delta t)=\arg\min_{\underline{\eta}}\ \varphi(\underline{\eta}),\qquad \text{s.t. } P(\underline{\eta})\in\mathcal{L}^\perp(P^+), \tag{5.9}$$

$$\text{ここで}\quad \varphi(\underline{\eta}):=-\log\ \det\left((P^+)^{-1}+\sum_{\kappa=m+1}^{N}(\eta_\kappa-\eta_\kappa^+)E^\kappa\right)-\sum_{\kappa=m+1}^{N}\eta_\kappa\theta_0^\kappa.$$

この最適化問題の目的関数 φ の第一項は self-concordant 関数であり，第二項は $\underline{\eta}$ に関する線形関数なので，φ も self-concordant 関数となる (Nesterov and Nemirovskii (1994))．したがって，修正子を求める問題は damped Newton 法で効率よく解くことができる．

注. $\mathcal{L}^\perp(P^+)$ の作り方より $\mathcal{L}\cap\mathcal{L}^\perp(P^+)$ に含まれる点は一点しかないが，上の補題はその点が AS 軌跡上にあることを主張している．また実際には厳密に AS 軌跡上に修正子が戻ってくる必要はない．しかし任意の δt に対して $\mathcal{L}^\perp(P^+)=\mathcal{L}^\perp(P(t+\delta t))$ であるので，上に述べたアルゴリズムを $P(t)$ 上の点 P_0 から始めると，結局予測子-修正子法を部分多様体 $\cup_{0<t}\mathcal{L}^\perp(P(t))$ 上で行っていることになる．

5.2 増分，誤差及び第二基本形式の関係

ここではバリアパラメータの増分 δt, η 座標での予測誤差, ∇^*-接続の第二基本形式の関係を導く．

第二基本形式を具体的に計算するために，まず $PD(n)$ に新しい座標系 $\tilde{\eta}=(\tilde{\eta}_p)$ を導入する．$\tilde{\eta}$ はそのベクトル場が以下のように定義されている座標系である：

$$\frac{\partial}{\partial\tilde{\eta}_i}:=\sum_{j=1}^{m}g_1^{ij}\frac{\partial}{\partial\theta^j},\qquad (g_1^{ij})=G_1^{-1} \tag{5.10}$$

(5.11)
$$\frac{\partial}{\partial\tilde{\eta}_\kappa}:=\frac{\partial}{\partial\eta_\kappa}.$$

(5.10) より $\partial/\partial\tilde{\eta}_i, i=1,\ldots,m$ は $\partial/\partial\theta^i$ の線形結合であるので，$\mathcal{L}$の各点で$\mathcal{L}$にそれぞれ接している．一方，(2.11) より $g(\partial/\partial\theta^i, \partial/\partial\eta_\kappa)=\delta_i^\kappa$ であるので，$\partial/\partial\tilde{\eta}_\kappa, \kappa=m+1,\ldots,N$ は$\mathcal{L}$の各点で$\mathcal{L}$に直交するよう作られていることに注意する．

(2.10) より，θ から η への座標変換の Jacobi 行列は $(\partial\eta_p/\partial\theta^q)=G$ であったので，$\tilde{\eta}$ とこれらの座標の間の Jacobi 行列は以下のように計算できる．

(5.12)
$$\left(\frac{\partial\eta_p}{\partial\tilde{\eta}_q}\right)=\begin{pmatrix} I & G_1^{-1}G_2 \\ 0 & I\end{pmatrix},\quad \left(\frac{\partial\tilde{\eta}_q}{\partial\eta_p}\right)=\begin{pmatrix} I & -G_1^{-1}G_2 \\ 0 & I\end{pmatrix}$$

(5.13)
$$\left(\frac{\partial\theta^p}{\partial\tilde{\eta}_q}\right)=\begin{pmatrix} G_1^{-1} & 0 \\ -S^{-1}G_2^TG_1^{-1} & S^{-1}\end{pmatrix},\quad ここで\ S:=G_3-G_2^TG_1^{-1}G_2.$$

さて，この座標系 $\tilde{\eta}$ を用いると$\mathcal{L}$の接続∇^*に関する第二基本形式 $\tilde{H}_\kappa^{*ij}$ は $\tilde{\eta}$ に関する Christoffel 記号 $\tilde{\Gamma}_\kappa^{*ij}$ に一致し次のように計算できる：

(5.14)
$$\tilde{H}_\kappa^{*ij}=\tilde{\Gamma}_\kappa^{*ij}:=\sum_{p=1}^N\tilde{g}_{\kappa p}\tilde{\Gamma}^{*ijp},\qquad i,j=1,\ldots,m,\quad \kappa=m+1,\ldots,N.$$

ここで $\tilde{g}_{pq}$ は行列（$\tilde{g}^{pq}$）の逆行列の成分で，$\tilde{g}^{pq}$，$\tilde{\Gamma}^{*pqr}$ はそれぞれ

(5.15)
$$\tilde{g}^{pq}:=g\left(\frac{\partial}{\partial\tilde{\eta}_p},\frac{\partial}{\partial\tilde{\eta}_q}\right)$$

(5.16)
$$\tilde{\Gamma}^{*pqr}:=g\left(\nabla^*_{\partial/\partial\tilde{\eta}_p}\frac{\partial}{\partial\tilde{\eta}_q},\frac{\partial}{\partial\tilde{\eta}_r}\right)$$

で定義される量である．このように定められた第二基本形式 $\tilde{H}_\kappa^{*ij}$ は$\mathcal{L}$が $PD(n)$ の中に ∇^*-接続の意味で（局所的に）どれくらい曲がって埋め込まれているかという情報を与える．

次の結果が $\tilde{H}_\kappa^{*ij}$，パラメータの増分 δt，解曲線からの誤差 $\hat{P}(t_0+\delta t)-P(t_0+\delta t)$ の η 座標 $(\tilde{\eta}_i(t_0+\delta t)-\eta_i(t_0+\delta t))$ の関係を与える．

定理 5.3. $\hat{P}(t_0+\delta t)\in PD(n)$ であるような $\delta t>0$ に対して，予測誤差 $\hat{P}(t_0+\delta t)-P(t_0+\delta t)$ の η 座標は $i=1,\ldots,m,\ \kappa=m+1,\ldots,N$ に対して以下のように与えられる：

(5.17)
$$\begin{cases}\tilde{\eta}_i(t_0+\delta t)-\eta_i(t_0+\delta t)=0,\\ \tilde{\eta}_\kappa(t_0+\delta t)-\eta_\kappa(t_0+\delta t)=-\dfrac{\delta t^2}{2}\displaystyle\sum_{i,j=1}^m\tilde{H}_\kappa^{*ij}(\eta(t_0+\xi\delta t))c_ic_j,\quad \exists\xi\in(0,1).\end{cases}$$

証明. (5.1) と (5.3) を用いるとある $\xi\in(0,1)$ について次のような $\eta(t)$ の Taylor 展開が得られる：

(5.18)
$$\begin{cases}\eta_i(t_0+\delta t)=\eta_i(t_0)-c_i\delta t, & i=1,\ldots,m\\ \eta_\kappa(t_0+\delta t)=\eta_\kappa(t_0)-c_\kappa\delta t+\dfrac{1}{2}\ddot{\eta}_\kappa(t_0+\xi\delta t)\delta t^2, & \kappa=m+1,\ldots,N,\end{cases}$$

したがって，(5.2) から

(5.19)
$$\begin{cases}\tilde{\eta}_i(t_0+\delta t)-\eta_i(t_0+\delta t)=0,\\ \tilde{\eta}_\kappa(t_0+\delta t)-\eta_\kappa(t_0+\delta t)=-\dfrac{1}{2}\ddot{\eta}_\kappa(t_0+\xi\delta t)\delta t^2\end{cases}$$

となる．

$P(t)$ はすべての t について$\mathcal{L}$上に拘束されるので，$\mathcal{L}$の直交方向への微分は 0，すなわち，$P(t)$ 上で $\dot{\tilde{\eta}}_\kappa=0,\ \kappa=m+1,\ldots,N$ となる．したがって，$\ddot{\eta}$ は次のように表される：

$$(5.20)\qquad \ddot{\eta}_\kappa=\sum_{i=1}^{m}\frac{\partial\dot{\eta}_\kappa}{\partial\tilde{\eta}_i}\dot{\tilde{\eta}}_i=-\sum_{i,j,l=1}^{m}\frac{\partial g_{\kappa l}g_1^{lj}}{\partial\tilde{\eta}_i}c_j\dot{\tilde{\eta}}_i\,,\qquad (g_{\kappa l})=G_2^T,\quad (g_1^{lj})=G_1^{-1}\,.$$

Christoffel 記号の座標変換による公式（例えば Kobayashi and Nomizu (1963) 参照）を用いると，η から $\tilde{\eta}$ へ座標変換することで $\tilde{\Gamma}_\kappa^{*ij}$ は Γ_r^{*pq} を用いて次のように計算できる：

$$(5.21)\qquad \tilde{H}_\kappa^{*ij}=\tilde{\Gamma}_\kappa^{*ij}=\sum_{p,q,r=1}^{N}\frac{\partial\eta_p}{\partial\tilde{\eta}_i}\frac{\partial\eta_q}{\partial\tilde{\eta}_j}\frac{\partial\tilde{\eta}_\kappa}{\partial\eta_r}\Gamma_r^{*pq}+\sum_{q=1}^{N}\frac{\partial^2\eta_q}{\partial\tilde{\eta}_i\partial\tilde{\eta}_j}\frac{\partial\tilde{\eta}_\kappa}{\partial\eta_q}\,.$$

η 座標系で ∇^*-アファインであるので $\Gamma_r^{*pq}=0$ となることと (5.12) を考慮して

$$(5.22)\qquad \sum_{l=1}^{m}\frac{\partial g_{\kappa l}g_1^{lj}}{\partial\tilde{\eta}_i}=\tilde{H}_\kappa^{*ij}\,,\qquad \text{for}\quad i,j=1,\dots,m,\ \kappa=m+1,\dots,N$$

を得る．$P(t)$ 上では $\dot{\tilde{\eta}}_i=\dot{\eta}_i=-c_i,\ i=1,\dots,m$ となることから題意が成立する．□

注. i）第二基本形式の各要素 $\tilde{H}_\kappa^{*ij}$ は，実際に (5.21) と同様な変換則と (5.13) を用いて θ 座標に関する幾何学量から ξ とは関わりなく各点で次のように計算可能である．

$$\tilde{H}_\kappa^{*ij}=\tilde{\Gamma}_\kappa^{*ij}=\sum_{l,k=1}^{m}\sum_{\mu=m+1}^{N}g_1^{il}g_1^{jk}s_{\kappa\mu}\Gamma_{lk}^{*\mu}\,,$$

ただしここで

$$(s_{\kappa\mu})=S=G_3-G_2^TG_1^{-1}G_2\,,\qquad \Gamma_{lk}^{*\mu}=\sum_{p=1}^{N}g^{\mu p}\Gamma_{lkp}^*\,.$$

ii）この結果より，$|\sum_{i,j=1}^m\tilde{H}_\kappa^{*ij}c_ic_j|$ が $\mathcal{L}$ 上で一様に（従って ξ に関わりなく）小さければ小さいほど，η 座標での予測子 $\hat{P}(t_0+\delta t)$ と解曲線 $P(t_0+\delta t)$ の誤差はより小さくなり，あるいは一定の誤差を許すなら δt をより大きくとれることが分かる．特に $\mathcal{L}$ 上で

$$\tilde{H}_\kappa^{*ij}=0\,,\qquad i,j=1,\dots,m,\quad \kappa=m+1,\dots,N\,,$$

すなわち $\mathcal{L}$ が二重自己平行である時は，δt を無限大にしても予測誤差を生じないことを示しており3節の結果を含んだものになっている．

iii）さらに，$\mathcal{L}$ が二重自己平行でなくとも

$$\left|\sum_{i,j=1}^{m}\tilde{H}_\kappa^{*ij}c_ic_j\right|=0\,,\qquad \kappa=m+1,\dots,N\,,$$

であれば，やはり Newton 反復なしで解けることをも意味している．このようなケースは $\mathcal{L}$ が曲がっていない方向と c によって決まる ∇^*-測地線を伸ばす方向が一致した場合などに起こり得る．実際，このような SDP 問題の例を構成することができる (Ohara (1997))．

iv）ここでは AS 軌跡上の点を初期値として，予測子と AS 軌跡の誤差を計算したが，実際の修正子は一般には厳密に AS 軌跡上には戻ってこない．したがって，以降の予測子を計算するには初期値誤差も考慮する必要がある．本稿では省略するが，初期値誤差による誤差も微分方程式論の簡単な誤差解析により評価することができる．それは AS 軌跡近傍で $|\sum_{i,j=1}^m\tilde{H}_\kappa^{*ij}c_ic_j|$ が小さければ，やはり小さいという結果が得られる．これらの誤差の和と self-concordant 関数最適化に関する結果 (Nesterov and Nemirovskii (1994)) を用いて，修正子が AS 軌跡の近傍に戻ってくるまでに必要な反復回数の上界を与えたとき，δt をどれだけ伸ばせるかという評価式を導くことができる．

6. おわりに

SDP 問題を情報幾何の観点から考察し，非許容領域での予測子-修正子法アルゴリズムにより許容領域の ∇^*-接続に関する第二基本形式が計算の手間と結びついていることを示した．特にこの第二基本形式が 0 のとき，すなわち二重自己平行の場合 Newton 反復が不要となり最適解の陽な表示を導出できた．4 節では二重自己平行な許容領域の例をいくつか構成したが，このような例から少し摂動したような許容領域を持つ SDP 問題なら大規模な問題でも，本稿で述べたアルゴリズムを基本にしたもので比較的少ない計算量で解ける可能性がある．その際には 5 節のような誤差評価から δt を求めるより適応的に定めた方が得策と思われる．

謝　辞

貴重な参考文献やコメントをご教示下さった査読者，大阪大学岩田覚氏，紀田馨氏，上智大学村松正和氏に感謝いたします．

参 考 文 献

Amari, S. (1985). *Differential-geometrical Methods in Statistics*, Springer, Berlin.

甘利俊一，長岡浩司 (1993).『情報幾何の方法』，岩波書店，東京．

Bayer, D. A. and Lagarias, J. C. (1989). The nonlinear geometry of linear programming I, II, *Trans. Amer. Math. Soc.*, **314**, 499-526, 527-581.

Boyd, S., El Ghaoui, L., Feron, E. and Balakrishnan, V. (1994). *Linear Matrix Inequalities in System and Control Theory,* SIAM, Philadelphia.

Faybusovich, L. (1997). Linear systems in Jordan algebras and primal-dual interior-point algorithms, *J. Comput. Appl. Math.*, **86**, 149-175.

Kobayashi, S. and Nomizu, K. (1963). *Foundation of Differential Geometry I*, Wiley, New York.

Kobayashi, S. and Nomizu, K. (1969). *Foundation of Differential Geometry II*, Wiley, New York.

小島政和 (1997). 半正定値計画の組合わせ最適化への応用に向けて，オペレーションズ・リサーチ，**42**(4), 216-221.

Malley, J. D. (1994). *Statistical Applications of Jordan Algebras*, Springer, Berlin.

Nesterov, Yu. and Nemirovskii, A. (1994). *Interior-point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia.

Ohara, A. (1997). Information geometric analysis of an interior point method for semidefinite programming, 数理解析研究所講究録, **1004**, 71-89.

Ohara, A., Suda, N. and Amari, S. (1996). Dualistic differential geometry of positive definite matrices and its applications to related problems, *Linear Algebra Appl.*, **247**, 31-53.

Tanabe, K. (1987). Center flattening transformation and a centered Newton method for linear programming, Manuscript presented at MP seminar, The Operations Research Society of Japan, July.

Vanderbei, R.J. and Yang, B. (1995). The simplest semidefinite programs are trival, *Math. Oper. Res.*, **20**(3), 590-596.

Wolkowicz, H. (1993). Explicit Solutions for Interval Semidefinite Linear Programs, Tech. Report, CORR 93-29, Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Canada.

Information Geometric Analysis of Semidefinite Programming Problems

Atsumi Ohara

(Division of Systems Science, Faculty of Engineering Science, Osaka University)

In this paper, we consider semidefinite programming (SDP) problems from points of view of information geometry. Regarding the feasible region of an SDP problem as a submanifold imbedded in the set of positive definite matrices, we elucidate that a certain imbedding curvature of the submanifold, called the second fundamental form derived from information geometric structure, relates to computational cost to solve the SDP problem.

Specifically, when the second fundamental form vanishes everywhere on the submanifold, in such case we call the feasible region is doubly autoparallel, we show the corresponding SDP problem can be solved without any Newton iterations and give a formula for an optimal solution.

In addition, we study the properties of doubly autoparallel submanifolds in the set of positive definite matrices. It turns out this specific structure has a relation with Jordan subalgebra of symmetric matrices. Using this fact, we construct some concrete examples of doubly autoparallel submanifold.

Key words: Semidefinite programming, information geometry, doubly autoparallel, Jordan algebra, explicit optimal solution.