

意味概念位相空間論

副題: 半正定値計画問題と因果推論

□**自然言語処理**¹NLP は、今や統計学や確率論、線形代数学、解析学だけでなく、微分幾何学から派生した情報幾何学²など、数学の多分野にまで発展してきた。従来の NLP は、文脈自由文法や変形生成文法、格文法、概念依存文法、モンタギュー文法、語彙機能文法(LFG)、句構造生成文法(GPSG)、主辞駆動句構造文法(HPSG)など「文法」を考察する学問で、数学とは無縁な学問であった。

□NLP が**数学の分野**に取り込まれるようになったのには、単語のバラツキを統計的確率論で構造と性質を求める研究が欧米から盛んになってきたことである。その背景には、ネット社会のビッグデータが容易に手に入る時代になったことや新しい統計学³や確率論、深層学習⁴や機械学習⁵理論などの成熟性、そして高速なパラレル GPU コンピュータが安価になったことなどが上げられる。特に欧米は、分ち書き(単語切り)の必要がなく、文書を単語集合やベクトル化するのに容易であったことも起因している。

□NLP の歴史からみても**人工知能 AI**との関係が深い。人工知能型言語 LISP で創った ELIZA など人工無能型対話心療診断ソフトウェアなどが 1964 年頃に世界を騒がせたこともあり、NLP はアトムのような人間型自律ロボットの人工頭脳として期待されてきた。そして、NLP と AI が数学の世界に入ってきた現代に入り、その概念や理論が急激に進化し、第三次 AI ブームの真価が問われることになってきた。

□NLP と AI が合体した **AI-NLU** は、人工知能の究極な課題である知識生成と構築そして推論を研究する分野であり、その核 Kernel となるものが意味概念位相空間⁶SCTS である。これは表記された文書などテキストの背景にある深層レベルの意味概念上で意味解析/理解⁷を処理する為の体系である。

□SCTS の**構造と性質**を研究するのが情報幾何学で、それは確率分布空間が「双対性」という構造から「空間の歪み」を曲率や振率として求めることである。これは熱力学の自由エネルギーやエントロピーにも表れる双対性という性質であり、NLP でもこの Symmetric な構造が意味の概念空間上に現れる。この構造と性質が意味を解析/理解する為の基本となる。

□SCTS は**行列**で表せる。行列は線形代数の分野だけではなく、リー群という概念とも結びついており群論や幾何学とは関係が深く、中でも正定値対称と呼ばれる行列が情報幾何学と NLP の Symmetric な関係で中心課題になっている。正定値対称行列⁸は、統計科学やデータ解析分野では「多変量解析の分散共分散行列や相関行列」として頻繁に使われているもので、力学、物性、制御、信号処理、量子情報、機械学習などの分野でも基本的な役割を果たしている。

¹ Natural language processing : NLP

² 確率分布空間の構造や性質などを考察する学問

³ ベイズ統計学や条件付き確率分布 CRF などの「分布」という概念的理論の発展

⁴ Deep Learning : DL

⁵ Machine Learning : ML

⁶ Semantic Conceptual Topological Space : SCTS

⁷ 意味解析は共通した意味概念で、意味理解は個々の意味概念として区別している

⁸ Positive definite symmetric matrix : PDSM

□SCTS は意味概念行列⁹という半正定値対称行列¹⁰で表せるので、まずは**正定値対称行列**の幾何構造を考える。 n 次実対称行列の集合を $Sym(n)$ と表し、

2×2 の対称行列の集合 $Sym(2)$ は、 $X = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix}$ となる。 但し、 $x_i \in \mathbb{R}$, $i = 1, 2, 3$

この行列はスカラー倍や和に対して閉じている。すなわち、

$$\alpha X_1 + \beta X_2 \in Sym(2) \quad \text{但し、} \alpha, \beta \in \mathbb{R}, X_1, X_2 \in Sym(2)$$

上記の2次元行列 X は 3次元ベクトル (x_1, x_2, x_3) になる実ベクトル空間である。一般には、 n 次元行列は、 $n(n+1)/2$ 次元の実ベクトル空間になる。

ベクトル空間に内積が定義されたリーマン空間へ拡張されたのが SCM となるので、

$$\langle X, Y \rangle := \text{tr}(XY) \quad \text{但し、} X, Y \in Sym(n)$$

は、 $\langle X, Y \rangle = x_1y_1 + 2x_2y_2 + x_3y_3$ となり、これは線形代数の内積の定義を満たしている。

□**正定値**とは、二次斉次関数 $x^T X x > 0$ を満たすことで、これは「 X の固有値が2つとも正である」ことや $x_1, x_3 > 0$, $\det X > 0$ と同値である。これは二次形式であることが重要で、NLPの表層レベルから抽出 Samplingされたデータは一次形式なので、これから自己相関行列を創って二次形式にすることが同値関係を継承する条件になってくる。ここで正定値対称行列の集合を $PD(n)$ と表す。

□ $PD(n)$ は**ベクトル空間** $Sym(n)$ 中の次のような性質を満たす。

(a1)凸集合である：



(a2)錐である：



(a1)は、 $PD(2)$ の任意の2点を結ぶ線分が $PD(2)$ に含まれることを意味している。

(a2)は、 $Sym(2)$ の原点(2×2 零行列)からある方向へ伸びた半直線を含むことを意味している。

これを凸錐¹¹と呼び、 $PD(n)$ は $Sym(n)$ の開集合である。

この境界を含めた半正定値対称行列の集合を $PD^{\sim}(n)$ と表すと、

$$x_1, x_2 \geq 0, \det X \geq 0 \quad \text{を満たす行列が SCM である。}$$

□ここで**双対な対称行列**を定義する。 $X \in PD(n), Y \in PD^{\sim}(n)$ に対し、

$$PD^*(n) := \{X \in Sym(n) | \langle X, Y \rangle > 0, \forall Y \geq 0, Y \neq 0\}$$

という内積が正であることを使って $Y > 0$ と鋭角の関係にある対称行列の集合を表し、この対称行列の集合が開集合で凸錐なら双対な関係になっている。すなわち、

$$PD(n) = PD^*(n) \quad \text{という自己双対}^{12} \text{の関係になっている。}$$

従って、SCM は内積を保つ線形変換を有する直交行列である。

□次に n 次実正則行列 T が定める行列の**合同変換**¹³を

⁹ Semantic Conceptual Matrix : SCM

¹⁰ Positive semidefinite symmetric matrix : PSSM

¹¹ Convex cone

¹² 複素空間上のユニタリー行列で、実数空間上なら直交行列にあたる

 が成り立つ。

これは次の性質を満たす。

(b1) $\tau_T(\cdot)$ は $\text{Sym}(n)$ の正則な線形変換であり、 $\text{PD}(n)$ を保存する。

$$\forall P \in \text{PD}(n), \tau_T(P) \in \text{PD}(n)$$

(b2) (推移性) 任意の $P_1, P_2 \in \text{PD}(n)$ に対して、ある T が存在して、

$$\tau_T(P_1) = P_2 \text{ が成り立つ。}$$

$\text{PD}(n)$ に、このような変換 $\tau_T(\cdot)$ が存在する性質は「等質性¹⁴」と呼ばれる。

命題の証明は $n=2$ の場合、任意の $P \in \text{PD}(2)$ は正則行列¹⁵ L により次のように分解できる。

$$\text{PD}(2) \ni P = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} = LL^T, \quad L = \begin{bmatrix} l_1 & l_2 \\ l_3 & l_4 \end{bmatrix}$$



これより、 $\tau_T(P) = TLL^TT^T$ の二次形式は任意の非零ベクトル $x \in R^2$ に対し、

$$x^T TLL^TT^T x = \|L^T T^T x\|^2 > 0$$

なので、(b1) が確認できる。また、



と分解しておく、 $T = L_2(L_1)^{-1}$ が (b2) で所望の T である。

このような2つの条件は厳しく、対称錐¹⁶ は本質的に5種類しかないことが証明されており、 n 次正定値対称行列の空間 $\text{PD}(n)$ はその一つの例になっている。

□そして、 $\text{PD}(n)$ が**双対平坦**な幾何構造を持っていることが重要なことである。

0 を含む適当な区間 I の $t=0$ で P を通る $\text{PD}(n)$ 内の滑らかな曲線 $P(t), t \in I$ の時、

$$X := \lim_{t \rightarrow 0} \frac{P(t) - P}{t} = \frac{d}{dt} P(t) |_{t=0} = \dot{P}(0)$$

を曲線 $P(t)$ の P での接ベクトルという。これは $P(t)$ の P での変化方向とその大きさ(速度ベクトル)を表す。この接ベクトルの集合を接空間と呼び、 $T_P \text{PD}(n)$ と表す。 $T_P \text{PD}(n)$ はベクトル空間である。

$n=2$ のときなら、滑らかな関数 $x_i(t), t \in I, i = 1, 2, 3$ に対し、



とすると、接ベクトルは

$$X = \begin{pmatrix} \dot{x}_1(0) & \dot{x}_2(0) \\ \dot{x}_2(0) & \dot{x}_3(0) \end{pmatrix} \in T_P \text{PD}(2) \text{ になる。}$$

各 P に対して $T_P \text{PD}(n)$ の元(接ベクトル)は対称行列として扱う。但し、 $P_1 \neq P_2$ のとき、 $T_{P_1} \text{PD}(n)$ と $T_{P_2} \text{PD}(n)$ は異なる接空間である。

¹³ T による P の共役行列で相似変換のこと

¹⁴ 共役が正定値対称行列内やその内にある他の行列になる閉じた変換規則

¹⁵ 逆行列が成立する行列

¹⁶ 2次錐、実正定値対称行列、複素、四元数、八元数(正定値エルミート行列)の5種類しかない

各接空間 $T_P PD(n)$ に内積¹⁷を



... *

と定める。 $g_P(\cdot, \cdot)$ は点 P のリーマン計量である。十分小さな t に対しては、 $t=0$ でのテーラー展開 Taylor expansion を考えると接ベクトルの定義から

$$P(t) = P + tX + o(t) \quad \text{Landau's symbol}$$

である。従って、2点 P と $P(t)$ との距離は $t > 0$ が十分小さい場合はほぼ

$$t\sqrt{g_P(X, Y)} = t\sqrt{\text{tr}(P^{-1}XP^{-1}X)}$$

になり、距離が各点 P に依存している。

□歪み空間の曲率や捩率などを明示化する為に各接空間の平行移動をする。

2点 P_1 と P_2 に対し、 $T_{P_1} PD(n)$ から $T_{P_2} PD(n)$ への2つの正則な線形変換を

$$\begin{aligned} \Pi : X &\rightarrow \Pi X := X && \text{恒等変換} \\ \Pi^* : X &\rightarrow \Pi^* X := P_2 P_1^{-1} X P_1^{-1} P_2 && \text{共役変換} \end{aligned}$$

と定義し、これが平行移動である。

微分幾何学では、何をもって平行とするかは、測地線¹⁸の問題と同等であり、任意に定義できる。

上記の2つの平行移動は「双対」な平行移動である。

接ベクトル場とは、各点 $P \in PD(n)$ に1つの接ベクトル $X \in T_P PD(n)$ を滑らかに対応させたものである。

各接空間の元(接ベクトル)は対称行列なので、接ベクトル場は $\text{Sym}(n)$ に値を持つ $PD(n)$ 上の関数

$X(P)$ となる。この $X(P)$ の変化率を表すベクトル場は



□ここで空間の歪みや曲がり方を示すテンソル量を定義する。

上記の平行移動のペア (Π, Π^*) は、

(c1) $g_{P_1}(X, Y) = g_{P_1}(\Pi X, \Pi^* Y)$ を満たす

(c2) Π と Π^* から得られる捩率と曲率は0になる。

という特徴をもつ。

(c1)は、平行移動の双対的な働きを示し、始点と終点のリーマン計量(内積)の値が保存されることを意味している。



性質を $PD(n)$ の「双対平坦性」という。

接続 ∇ と ∇^* は、測地線の定義にもなるので、 ∇ -測地線は、

¹⁷ $\langle x, y \rangle = \|x\| \cdot \|y\| \cos\theta$ なので接ベクトルの直交性や長さを定める

¹⁸ 曲面の最短距離をいい、任意の定義が存在する

$$P(t) = P + tX \in PD(n), \quad P \in PD(n), \quad X \in \text{Sym}(n)$$

の形をしており、接ベクトルは任意の t で $\dot{P}(t) = X$ である。

∇^* -測地線は、

[Redacted]

∇^* -測地線は、 $(P^*(t))^{-1} = P + tY$ と同じ形をしているので、 Π^* は逆行列でみて直線に見えるものを、まっすぐの基準と見なすという平行移動である。

リーマン幾何において用いられるレヴィ-チビタ接続¹⁹ $\hat{\nabla}$ は、

[Redacted]

レヴィ-チビタ接続は自己双対的といえる。

□埋め込み曲率²⁰と自己平行性²¹とは、**双対平坦性**²²をもつ $PD(n)$ のある部分集合が局所的にどれくらい曲がっているかを接続で求めるものである。3次元集合 $PD(2)$ に属する2次元曲面 M 、すなわち部分多様体上の2種類の接ベクトル場 X, Y に対し、 $\nabla_X Y$ という新しいベクトル場が M 上で定義される。この部分多様体 M が点 $P \in M$ でどの程度曲がっているかを Π が定める平行性を用いて次のように考える。点 P で $\nabla_X Y$ は定義から $PD(2)$ の接ベクトル空間には属するが、 M の接ベクトルとは限らない。

$T_P M$ は $T_P PD(2)$ の部分空間なので、 $T_P PD(2)$ の内積であるリーマン計量 g_P を用いて、 $\nabla_X Y$ は g_P に関し

[Redacted]

Z_2 はその直交補空間²³の元である。従って、すべての方向 X, Y について $Z_2 = 0$ なら、点 P で M は Π の平行性に鑑みて平坦である。 Π^* についても同様である。

$\nabla_X Y$ と $\nabla_X^* Y$ の直交補空間の方の成分 Z_2 を埋め込み曲率(第二基本形式)と呼び、それぞれ

$H_M(X, Y), H_M^*(X, Y)$ と表す。 M の任意の接ベクトル場 X, Y に対して M 上で $H_M(X, Y) = 0, H_M^*(X, Y) = 0$ となるとき、 M はそれぞれ ∇ -自己平行、 ∇^* -自己平行という。特に $\nabla - (\nabla^*)$ -測地線とは、 M が1次元、すなわち曲線である場合の $\nabla - (\nabla^*)$ -自己平行な部分多様体のことである。

□依存関係を明示的に反映して記述する統計モデルの**グラフィカルモデル**では、変数間の依存関係を変数に対応する節点と節点の間を結ぶ枝の集合からグラフという数学的概念で表現する。これが因果推論の基本モデルである。因果推論は無向/有向性ベクトルの連結で表現する。ふきや理論は有向性

¹⁹ Levi-Civita connection : リーマン計量を保つ接束上の捩れない計量接続 (アフライン接続) のこと

²⁰ Embedding curvature

²¹ Autoparallelism

²² Dually flatness

²³ Tangent 空間の双対で直交な空間

ベクトルを非可換性関係子と言われる関数で連結をしていくので、強化学習の Q 学習に似た新理論であるが、多層化された状態が推移していくことで連結が動的に変化するところが相違する。ここでは単の為にガウシアングラフィカルモデル²⁴を考える。意図と単語ベクトルと特徴ベクトルのデータで考える。データは標本分散共分散行列および相関行列で、[単語:単語ベクトル:特徴ベクトル]の順で行列が表現されている。



この行列では、単語ベクトルと特徴ベクトルの相関関係(2,3)が高いが、これは間接的は反映であるとみられる。意図が一定なら単語ベクトルと特徴ベクトルは独立になる。

$$P(\text{単語ベクトル, 特徴ベクトル} \mid \text{意図}) = P(\text{単語ベクトル} \mid \text{意図})P(\text{特徴ベクトル} \mid \text{意図})$$

となる。そこでこれらの逆行列を求める。右の行列は偏相関行列²⁵と呼ぶ。



は無相関になったので、上記の条件付き確率分布が証明された。

分散共分散行列 $\Sigma = (\sigma_{ij})$ の逆行列を $D = (d_{ij})$ と記す。D の上三角要素 $i \leq j$ なる d_{ij} をすべて定めれば単純な**因果関係**が求まる。D の非ゼロパターンを指定した正規分布モデルを GGM という。D の非ゼロ要素の添字集合を $I = \{(i, j) \mid d_{ij} = d_{ji} \neq 0, i \leq j\}$ とし、ゼロ要素の添字集合を $\bar{I} = \{(i, j) \mid d_{ij} = d_{ji} = 0, i < j\}$ と表す。この添字集合 (I, \bar{I}) でグラフィカルモデルが定まり、上記の行列なら $I = \{(1,1), (2,2), (3,3), (1,2), (1,3)\}$ と $\bar{I} = \{(2,3)\}$ という上三角要素だけを示す。

x を平均 μ 、添字集合 (I, \bar{I}) によって定まる非ゼロパターンを持つ分散共分散逆行列 D によって指定される GGM に従う確率変数とし、その密度関数を $P(x; \mu, D)$ と記すことにすると、



ここで $(i, j) \in \bar{I}$ ならば、 $d_{ij} = d_{ji} = 0$ である。この密度関数の指数関数の中の式で x_i と x_j を共に含む項が $d_{ij}x_ix_j$ のみであることを考えると、変数間の繋がりが相互作用を直接反映しているのは、分散共分散逆行列 D の非ゼロパターンであるということは重要なファクターである。GGM は、各変数を節点として、非ゼロの節点 i と j を結ぶ枝を対応させ、変数間の関係を表したものである。一般に、分散共分散逆行列 D を持つ多変量正規分布において、

$$\hat{d}_{ij} = \frac{d_{ij}}{\sqrt{d_{ii}d_{jj}}} \text{ を偏相関係数と呼び、上記の偏相関行列の要素を求める式になる。}$$

この裏には変数 i と j が独立であり、GGM の節点間は非対角要素の内 0 に近いものを 0 とすることが必要となる。非循環性による交絡因子²⁶をグループ分けをして、交絡の構造がこれにより構築される。

²⁴ Gaussian graphical model : GGM グラフィカルモデルの発展系は LASSO モデル (可換性)

²⁵ 符号は因果関係の「向き」である

²⁶ Confounding factor

逆行列の座標表現は、 $(d_{11}, d_{22}, d_{33}, d_{12}, d_{13}, d_{23}) = (1, 2, 3, 1, 1, 2)$ となる。

ここで、 $I = \{(1,1), (2,2), (3,3), (1,2)\}$ 、 $\bar{I} = \{(1,3), (2,3)\}$ とすると、

混合座標系では、 D は、 $(d^*(I), d(\bar{I})) = ((d_{11}^*, d_{22}^*, d_{33}^*, d_{12}^*), (d_{13}, d_{23})) = ((1, 2, 3, 1), (0, -1))$

と表現される。座標表現と混合座標系が交互になることに注意する。これは混合座標系は正定値対称行列の大域的座標系であり、混合座標系での値が与えられれば、対応する正定値対称行列は一意に確定し、

[Redacted]

[Redacted] となる。但し、 $\langle P, Y \rangle = \text{tr}(PY)$ 。

特性函数の対数函数 $\psi(P) = -\ln |P|$ は、 $PD(n)$ の情報幾何ではポテンシャル²⁸と呼ばれ、この函数の二階微分、三階微分からリーマン計量や双対接続などの幾何的な量が導かれる。従って、特性対数函数は双対平坦な構造を定める基になっている。それは多変量ガウス分布の尤度やエントロピーを計算しても、この函数が現れてくることから判る。情報幾何学では、この函数にもやはり双対な関係を持つ相手である双対ポテンシャル $\psi^*(P) = -\ln |P^{-1}|$ が存在し、 ψ と対等な働きで協調して情報幾何学を支えており、 P も P^{-1} も表としての鏡²⁹のような関係である。熱力学では ψ^* は自由エネルギーであり、最適化理論では双対問題の関係になり、

- (i) 障壁関数性: $PD(n)$ の境界で発散する狭義凸函数
- (ii) 自己整合性: 任意の非零な $X \in \text{Sym}(n)$ に対し、 $D^2\psi(P)[X, X] > 0$ で、

[Redacted]

[Redacted]

[Redacted]に対するバリア函数としてニュートン法をベースにするアルゴリズムを用いると、半正定値計画と呼ばれるクラスの凸最適化問題の求解に必要な計算量の多項式が示される。

□また、上記ポテンシャルを二つ用いて、 $PD(n)$ 上の2つの行列 P_1, P_2 に対して定義された函数 $D(P_1, P_2) = \psi(P_1) + \psi^*(P_2) + \text{tr}(P_1 P_2^*) - n$ を**ダイバージェンス**と呼ぶ。 $D(P_1, P_2)$ は、 P_1, P_2 をそれぞれの分散共分散行列とする平均0の n 変量ガウス分布 $P_1(x)$ と $P_2(x)$ のカルバック・ライブラー(KL)情報量(の2倍)と一致する。すなわち、

$$D(P_1, P_2) = 2 \int_{R^n} p_1(x) \ln \frac{p_1(x)}{p_2(x)} dx \quad \text{が成り立つ。}$$

KL 情報量は統計・情報科学の随所に現れ、2つの確率分布の隔たりを表し、同じなら0になる正值に

²⁸ $E[P] = -\sum_{w=1}^n p(w) \ln p(w) = -P \ln P$: シャノンのエントロピーと比較してみる(平均情報量)

²⁹ どちらが実像か虚像かは問題外であり、群論の鏡映な関係にある

なる非距離³⁰である。従って、PD(n)上のダイバージェンス $D(P_1, P_2)$ は、対応する平均 0 のガウス分布 $p_1(x)$ と $p_2(x)$ の違いで P_1 と P_2 の乖離度を測ったものである。次にこれらの双対なポテンシャル関数 ψ, ψ^* やダイバージェンス D が行列の各要素(座標系)との有効な関係を示す。 $i \leq j$ として、正定値行列

[Redacted]

[Redacted]

[Redacted]るので、 $\{p_{ij}^*\}$ も P の別な座標系と考えることができる。これを $\{p_{ij}\}$ の**双対座標系**と呼ぶ。

一般の n 次対称行列では、

$$P = \sum_{1 \leq i \leq j \leq n} p_{ij} E_{ij}, \quad P^{-1} = \sum_{1 \leq i \leq j \leq n} p_{ij}^* E_{ij} \quad \text{となる。}$$

これら2つの座標系 $\{p_{ij}\}$ と $\{p_{ij}^*\}$ で不足要素を互いに補い合っている座標系のことである。

$f(D)$ の偏微分①式で特に $D = P, S = 0$ とすると、次の関係式

[Redacted]

[Redacted]

この関係は、 $\{p_{ij}\}$ からポテンシャル ψ の偏微分により P の別の座標系 $\{p_{ij}^*\}$ が求まることを意味する。

式⑤と逆な関係も ψ^* から得られ、これをルジャンドル変換と呼ばれる。

上記にもあるが、行列 P の座標系 p_{ij} での直線を測地線とすると、ひとつの接続が導入されたことになる。逆行列 P^{-1} による座標系 p_{ij}^* での直線を測地線と考えるともうひとつの接続が導入されたことになる。これらが正に互いに双対な2つの接続 ∇, ∇^* の一例となっている。

「ポテンシャル関数 ψ の偏微分から」→「双対座標が生成され」→「双対接続が導入される」という流れになる。これは双対平坦構造による性質であり、双対平坦な接続が存在すれば、必ずポテンシャル関数と呼ばれる凸関数が存在する。

□双対座標系の構造を**座標曲線**に着目してあげると、任意に選んだ正定値行列 $P_0 \in R^{n \times n}$ を通る次のような曲線 $P(t) = P_0 + tE_{ij}, \quad P^*(s) = (P_0^{-1} + sE_{kl})^{-1}, \quad t, s \in R$

を「 P_0 を通る p_{ij} - 座標曲線, p_{kl}^* - 座標曲線」と呼ぶ。例えば、正定値行列 $P_0 \in R^{2 \times 2}$ なら、

[Redacted]

[Redacted]

[Redacted]

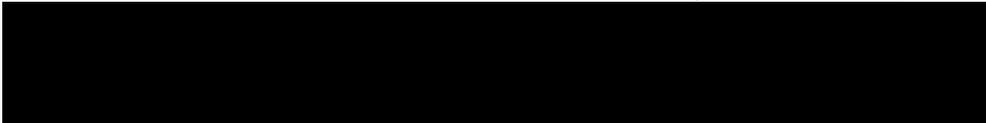
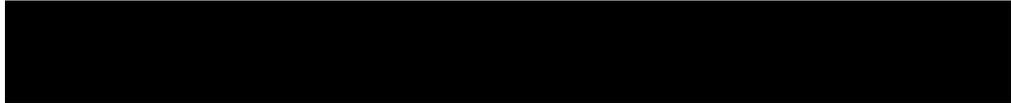
[Redacted]

が成り立つ。つまり、 P_0 を通る p_{ij}, p_{kl}^* 座標曲線は、 $((i, j) \neq (k, l))$ であれば常にリーマン計量の意味で互いに P_0 で直交していることになる。従って、例えば式④で $p_{11} = 1, p_{22} = 2, p_{12}^* = 0$ 、すなわち、

³⁰ 非可換性であるので類似度になるが、数値が小さい方が類似性が高い

$$P = \begin{pmatrix} 1 & p_{12} \\ p_{12} & 2 \end{pmatrix} > 0, \quad P^{-1} = \begin{pmatrix} p_{11}^* & 0 \\ 0 & p_{22}^* \end{pmatrix} > 0$$

と指定した行列 P がもし存在すれば、 p_{12} -座標曲線と p_{11}^* -座標曲線あるいは p_{22}^* -座標曲線は P_0 で直交する。これは上記の条件で $\Sigma = P, D = P^{-1}, \bar{I} = \{(1,2)\}, I = \{(1,1), (2,2)\}$ とした場合と同じことである。(1,1), (2,2)要素がそれぞれ $s_{11} = 1, s_{22} = 2$ の標本分散共分散行列 $S > 0$ を考えたなら、 $p_{ij} = d_{ij}^*, p_{ij}^* = d_{ij}$ という交互になるので、式②の条件は、



は互いに直交し、両者が交わる点が P_0 である。

更に標本分散共分散行列が P の形で与えられる場合、最尤推定は p_{23} の値によらず P_0 となる。

一般には、次のような構造を持ったグラフィカルモデルの共分散行列 Σ を集めた部分多様体

$M := \{\Sigma = D^{-1} \in PD(n) | \sigma_{ij}^* = d_{ij} = 0, (i, j) \in \bar{I}\}$ を考える(σ_{ij}^* は σ_{ij} の双対座標)と、

これは逆行列 Σ^{-1} が $d_{ij} = 0$ と線形にされている。尚、 D^{-1} は式③を参照する。

この M のように、逆行列でみてまっすぐ、すなわち逆行列の各要素間に線形制約がある部分多様体を



$\forall P \in M, D(S, \Sigma) + D(\Sigma, P) = D(S, P)$ が成り立つ。これとダイバージェンスは常に非負であるので、

「 Σ は M 上で S からのダイバージェンスが最小になる行列、つまり $\Sigma = \operatorname{argmin}_{P \in M} D(S, P)$ 」である。

最小化問題の最適解として特徴付けられた Σ は、 $D(S, P)$ が P の双対座標、すなわち P^{-1} に関して狭義



31 三角化グラフのことで三角行列で表現できる

32 因果関係で重要な非巡回性が明示的になる

□線形計画問題とは、線形等式・不等式系の解集合の上で線形方程式を最適化する問題である。

例 1.

最小化: $2x_1 + x_2$



これは x_{11}, x_{22} を変数として、 $\begin{pmatrix} x_{11} & 1 \\ 1 & x_{22} \end{pmatrix}$ の形の「半正定値対称行列」の中で、対角要素の和が最小であるものを求める問題である。この問題は、双曲線で区切られる右上方領域上³³で線形関数を最小化する問題になっている。下記の例題である。

最小化: $x_{11} + x_{22}$

条件: $x_{11}x_{22} \geq 1, x_{11} \geq 0$

最適解は $x_{11} = 1, x_{22} = 1$ で最適値は 2 となる。このように半正定値計画問題の実行可能領域は曲がっていて「非線形」であることを認識する。この非線形性ゆえに、半正定値計画問題は線形計画問題にはない豊かな表現力を持つ。1990 年代に入り、線形計画問題に対する下記的な多項式時間解法であった内点法³⁴が驚くべき類似性を伴って半正定値計画問題に拡張可能であることが明らかとなり、それが切っ掛けでいろいろな分野でのモデリングに活用されるようになってきた。例えば、システムの周



利」や「人数」や「割合」の推定のように当て嵌めた多項式が常に非負であることが要請される。そこで任意の点で非負の値を取る偶数次多項式を当て嵌める。多項式の次数を $2n$ とすると、非負 $2n$ 次多項式の係数を陽に表現することは一見容易ではないように思えるが、実は x を変数とする任意の非負 $2n$ 次多項式 $f(x)$ は、 a_{ij} をその i, j 要素とする $n \times n$ の半正定値対称行列 A を用いて下記のように表せる。

$$f(x) = (1 \ x \cdots x^n) A \begin{pmatrix} 1 \\ x \\ \vdots \\ x^n \end{pmatrix}, \quad A \geq 0 \text{ と表現できる。}$$

³³ 目的関数を直線とすると、それに最も近い双曲線上の点が最適解となるのが半正定値計画問題

³⁴ 同次元内の反復法で最適値を求める法のこと、次元をひとつ上げて最適値を求めるのは外的法

例えば、任意の非負 4 次多項式は半正定値対称行列 $A = (a_{ij}) \geq 0$ を用いて、

[Redacted]

て、4つの観測データ $(t, y) = (1, 3), (2, -1), (3, 0), (4, 3)$ が得られたとする。これに非負 2 次多項式を当て嵌めると、

$f(t; A) = (1 \ t) \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} 1 \\ t \end{pmatrix}$, $A \geq 0$ と表現する。積をとって正方行列となるような2つの行列 P、

Q のトレースについて成立する公式 $\text{tr}(PQ) = \text{tr}(QP)$ を用いると、

[Redacted]

条件: $s_1^2 \leq t_1; s_2^2 \leq t_2; s_3^2 \leq t_3; s_4^2 \leq t_4;$

$$s_1 = 3 - f(1; A); s_1 = -1 - f(2; A); s_1 = 0 - f(3; A); s_1 = 3 - f(4; A); \quad A \geq 0$$

と変形できる。例えば、

$s_1^2 \leq t_1$ は、 $\begin{pmatrix} t_1 & s_1 \\ s_1 & 1 \end{pmatrix} \geq 0$ と同値である。また、 $s_1 = 3 - f(1; A)$ は、式⑥で $t = 1$ と置いて、

$s_1 = 3 - \text{tr} \left[A \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right]$ となる。同様の考え方をすべての不等式や等式に当て嵌めることにより、すべての制約を半正定値条件や線形等式条件で表すことができ、結果として以下のような最適化問題に定式化できる。

[Redacted]

非線形な条件が半正定値条件のみであり、目的関数や変数間の関係はすべて線形(1次式)である。これが半正定値計画問題の典型例である。簡単な為、2次式としたが、4次式以上の場合も同様の形で定式化できる。

□半正定値計画問題を正式に導入する。 C, A_1, \dots, A_m を $n \times n$ 実対称行列とし、 b_1, \dots, b_m を実数とする。行列 $X \in \text{Sym}(n)$ を変数とする次のような最適化問題を半正定値計画問題という。

$$\begin{aligned} & \text{maximize} && C \bullet X \\ & \text{subject to} && A_i \bullet X = b_i, \quad i = 1, \dots, m \\ & && X \succeq 0 \end{aligned} \quad \dots \textcircled{8}$$

③、④の条件を満たす点をそれぞれの問題の実行可能解といい、 X や S が正定値対称行列であるものを内点実行可能解という。主問題の内点実行可能解の集合を P 、双対問題の内点実行可能解 (S, y) の S 部分の集合を D と記す。正定値対称行列集合 $\text{PD}(n)$ は自己双対錐であるので、「主問題の目的関数値は双対問題の目的関数値よりも大きいか等しい」ことが示せる。

下では、「両問題に最適解が存在して最適値は必ず等しいこと」が知られており、これを**双対定理**という。主問題の目的関数の値と双対問題の目的関数の値の差を**双対ギャップ**と呼ぶ。双対ギャップが0であるような主問題と双対問題の実行可能解は、両問題のそれぞれ最適解である。従って、双対ギャップが十分に0に近い実行可能解は、それぞれの問題の良い近似最適解となっており、そのような解を求めることが、半正値計画問題の解法の主要目的である。

例2の双対問題は、

例3.

るのは、主問題⑦の正定値対称行列の空間に $-\ln |X|$ をポテンシャル関数として導入して、情報幾何的構造を考えることで明確に判る。この構造は、元の行列の座標系についての直線を測地線と定める ∇ -接続と、その逆行列の空間での直線を測地線と定める ∇^* -接続と双対は ∇^* -接続の2つの接続、そして $-\ln \det$ 関数のヘッセ行列で定められるリーマン計量からなるものであった。このような構造に照らし合わせてみると、主問題は元の行列の最適化問題、双対問題は逆行列の空間での最適化問題と考えるのが自然である。そして元の座標系に関して「アファイン空間」として表現されるものを ∇ -自己平行部分多様体、双対座標系に関して「アファイン空間」として表現されるものを ∇^* -自己平行部分多様

体と呼ぶのであった。このとき、主問題の実行可能領域 P は、明らかに ∇ -接続に関して自己平行な部分多様体となる。双対問題のじっこ可能領域 D は、 ∇^* -接続に関して自己平行な部分多様体とあなる。元の座標系で見ると、 P は真っ直ぐで、 D は曲がっており、双対座標系で見ると、 D は真っ直ぐで、 P は曲がってみえる。例3で見ると、双対問題の実行可能領域は、

[Redacted]

[Redacted]

[Redacted]

[Redacted], $y = (-1 + \sqrt{5})/2$ でこの等式が成立し、

このとき、 $x_{11} = 1/(1 - y^2)$ となる。実は、 $P \cap D$ は一般の半正定値計画問題について1点からなることが知られており、効率的な解法に重要な役割を果たす。

□半正定値計画問題の主要解法である**内点法**を導入し、その計算複雑度が問題の情報幾何学曲率の積分で表される。半正定値計画問題を解く上で特に問題となるのは、 X や S が半正定値対称であるという非線形な条件である。この条件が問題となるのは、行列が半正定値対称行列の集合の境界の部分、すなわち正定値対称ではなくなる部分である。そこから十分に離れていれば、この条件は気にしなくても良い。そこで、次のようなことを考える。まず、

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted] 行列で、しかも先の問題の最適解の近くにあると考えられるので、⑨の近似最適解を中心として⑩の目的関数を凸2次関数で近似して、線形条件 $\text{tr}(A_i X) = b_i, i = 1, \dots, m$ の下で最適化する。いわゆるニュートン法の反復を一度行くと、重みが 1.1 の問題⑩の十分良い近似最適解が得られることが期待される。このように重みを大きくしながら線形制約のみの最適化問題を解くことを繰り返す。その結果、重みは 10^6 として、

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

の近似最適解である。この最適化問題は、⑦の目的関数 $\text{tr}(CX)$ と、半正定値対称行列の境界で無限

大のいく障壁函数となる $-\ln |X|$ の重み付き和の最小化問題であるが、重みが十分に小さければ、元の半正定値計画問題の十分に良い近似となっている。この重み、すなわち 1.1 や 10^6 を連続的に変えていくと、重みを変えた問題の最適解は常に正定値対称行列で、重みが無限大に近づくにつれて半正定値計画問題の最適解に近づいていく曲線を描く。この曲線を**中心曲線**と呼ぶ。すなわち、主問題(P)に対する中心曲線は t をパラメータとして、下記の最適化問題の解が成す曲線である。

$$\begin{aligned} & \text{maximize } Cx - \sum_{i=1}^m \lambda_i y_i \\ & \text{subject to } S = C - \sum_{i=1}^m \lambda_i A_i y_i, \quad S \geq 0 \end{aligned} \quad \dots \textcircled{12}$$

の最適解が成す曲線として導入される。この問題の最適解を $(y(t), S(t))$ と記すことにする。

⑪の最適性条件は、等式条件 $\text{tr}(A_i X) = b_i$ に対する未定乗数を λ_i として、ラグランジュ関数を $\text{tr}(CX) - \sum_{i=1}^m \lambda_i (\text{tr}(A_i X) - b_i)$ と書くと、通常の手続きにより、

$$tC - X^{-1} - \sum_{i=1}^m \lambda_i A_i y_i = 0, \quad \text{tr}(A_i X) = b_i, \quad i = 1, \dots, m \quad \dots \textcircled{13}$$

と書ける。

⑫の最適性条件は y_i で微分して 0 とおいて、

$$tb_i - \text{tr}(A_i S^{-1}) = 0, \quad S = C - \sum_i \lambda_i A_i y_i \quad \dots \textcircled{14}$$

となる。

更に、⑬と⑭より、 $X(t), (S(t), y(t))$ は、

$$\text{条件 } XS = tI, \quad \text{tr}(A_i X) = b_i, \quad i = 1, \dots, m, \quad S = C - \sum_i \lambda_i A_i y_i, \quad X \geq 0, \quad S \geq 0 \quad \dots \textcircled{15}$$

を満たす点として定式化可能である。

$t = 1$ のときに、⑬は、

$$\begin{aligned} & \text{maximize } Cx \\ & \text{subject to } X \geq 0, \quad \text{tr}(A_i X) = b_i, \quad i = 1, \dots, m \end{aligned}$$

、双対問題の $t = 1$ のときの中心曲線上の点となる。グラフィカルモデルの最尤推定問題も、ある半正定値計画問題の $t = 1$ の中心曲線上の点を求めることに帰着する。

内点法では、 t を少しづつ更新して大きくしつつ、その近似解をニュートン法で求めることにより、最適解に近づいていく。⑪に基づいて中心曲線を追跡するものを主内点法、⑫に基づいて中心曲線を追跡するものを主双対内点法と呼ぶ。実用的には主双対内点法が使われる。主双対内点法の反復回数が情報幾何的量として表現できることを示す。 $\beta \in (0, 1)$ とし、 $N(\beta)$ を中心曲線の近傍とする。 $\beta = 0$ では中心曲線そのものとなり、 β が大きくなるにつれて拡がる。

□主双対内点法(予測子・修正子法)

1. (X, S, y) がパラメータ t の中心曲線上にあるものとする

$$\begin{aligned} & \text{maximize } Cx - \sum_{i=1}^m \lambda_i y_i \\ & \text{subject to } S = C - \sum_{i=1}^m \lambda_i A_i y_i, \quad S \geq 0 \end{aligned}$$

3. ニュートン法を用いて、パラメータ $t + u$ の中心曲線上の点 $(X(t + u), S(t + u), y(t + u))$ に戻る。

(修正ステップ)

4. ステップ 1 に戻る。

予測子・修正子法では、適切に β を選べば、1反復あたり、 $(1 - \frac{0.1}{\sqrt{n}})$ より小さい比率で双対ギャップが縮小して0に近づいていく。内点法の大きな特長は、この収束率が解から遠いところでも一様に保証されることである。このようなアルゴリズムを**多項式時間アルゴリズム**という。数千変数の線形計画問題と半正定値計画問題を近傍の大きさ β を変えて主双対内点法で解き、反復が進むに連れて、最適解までの距離の尺度である双対ギャップがどのように減少するかが判る。 β が小さい方がより多くの反復回数

を必要とするものの、グラフの横軸のスケールを変えると同じような形状になる。解析⑥、⑦によると、

_____可能領域 D や P の埋め込み曲

率であり、ノルムは $\|X\|_P^2 = g_P(X, X)$, $P \in PD(n)$, $X \in T_P PD(n)$ と各点 P においてリーマン計量 g_P より定められる。内点法の反復回数 $\#(t_1, t_2, \beta)$ は積分 I_{PD} を $\sqrt{\beta}$ で割ったものとなり、近傍を小さくすると、より精密に中心曲線を追跡するためにより多くの反復回数を必要とする。注目すべきは、この被積分関数が P, D 上での中心曲線の埋め込み曲率で書かれており、厳密な意味で情報幾何的な意味を有することである。特に、線形計画問題の場合には、

_____ることの反映でも

あり、その一方で、線形計画問題の中心曲線の大域的性質の反映となっている。このような形で計算複雑度と微分幾何的性質が結び付くのは興味深く、これが言語の確率分布空間に結び付く。

□以上は正定値対称行列の情報幾何的始点から構造を示し、ガウシアングラフィカルモデリングと半正定値計画という2つの研究領域を意味概念の言語空間に適用して論じてきた。個々に調和しながら相互に関係し合う双対性がこの領域の切り口になっていることが重要である。意味概念位相空間が以上のような半正定値行列による表現でのアファイン微分幾何、ヘッセ幾何などの幾何学や確率論、偏微分方程式などの広い分野に跨っていることはより言語解析を複雑にしているが、量子コンピュータや5G、そしてIoTなどの技術やインフラの発展で少しずつ人間の知能へ近づいてきている。人間の知能を超える日も遠からず…あとは倫理の問題を人間がどのように制御するかである。

<追記>

SCM の単語 w_i と単語 w_j との交点 w_{ij} の座標は、

$$\text{Coordinate}(w_{ij}) = (\vec{w}_i \cdot \vec{w}_j^T) \cdot (\vec{w}_i \cdot \vec{w}_j^T)^T \quad \text{where, } \vec{w}_i, \vec{w}_j \in R^n$$

別な方法として、

$\vec{w}_i, \vec{w}_j \in R^{m \times m}$ として上記の式を適用してもよい。尚、 $n=m(m+1)/2$ である。

<文法偏>

- ・自立語は、名詞、動詞、形容詞(連体詞含む)、副詞、接続詞、感動詞。
- ・付属語は、助詞、助動詞(形容動詞活用含む)。
- ・英文法での内容語は自立語、機能語は付属語と分類されるが厳密ではない。
- ・この分類が国文法と相違する理由は、構文解析³⁵も文脈解析³⁶も意味解析³⁷の為にあること。

³⁵ Syntax : 構文解析は表記(表層)レベルの分割だけで意味(深層)レベルを考慮していない為に文法が相違

³⁶ Context : 主に照応解析(代名詞代入、ゼロ代名詞補完)で、文間や文特有意味は意味解析で処理

³⁷ Semantic : 慣用句や文に依存される特有意味並びに文間の接続関係