

内点法・情報幾何・最適化モデリング

土谷 隆[†]

(受付 2013 年 1 月 11 日; 改訂 6 月 18 日; 採択 6 月 21 日)

要 旨

本論文では、凸最適化に関するモデリング・数理・アルゴリズムに関連した 3 つの話題を紹介する。第一の話題は内点法と情報幾何である。内点法の情報幾何の理論の概要を説明し、十分に大きな問題に対する数値実験を通じて、線形計画問題・半正定値計画問題両方の問題において、内点法の反復回数が微分幾何学的量であることを検証する。第二の話題はデータ同化に現れる大規模ガウシアングラフィカルモデル推定への内点法の適用である。推定された分散・共分散行列で一番大きいものは 34300×34300 であり、このモデルは 101310 個のパラメータを持つ。また、推定されたモデルのうちでパラメータ数最大のもは、行列の大きさが 8585×8585 で、150381 個のパラメータを持つ。これらを含むいくつかのモデルを最尤推定し、AIC や BIC によりモデル選択を行った。最後の話題は古代メソポタミア都市ヌジの社会構造および人口の推定である。我々は、ヌジから出土した 2000 枚弱の粘土板に記された人名のデータベースに基づき、凸 2 次計画問題を用いて文書の作成された順番および登場する人物の生誕年および死亡年を推定し、古代社会の社会構造や人口動態について検討する。

キーワード：内点法，情報幾何，グラフィカルモデル，データ同化，メソポタミア，ヌジ，人口推定。

1. はじめに

早いもので内点法が登場してから 30 年近くが経った。1984 年に Karmarkar によって提案された線形計画問題に対する内点法は、1990 年代から 2000 年代初頭にかけて、凸 2 次計画問題や半正定値計画問題、対称錐計画問題などの凸最適化問題へと拡張され、これをきっかけとして、この十年の間に凸最適化は制御、機械学習、信号処理、パターン認識、統計科学等の数理諸分野でモデリングのための強力な道具立てとなりつつある。この分野で現在広く使われている主双対内点法は、田邊と小島・水野・吉瀬によって、1987 年 2 月に統計数理研究所で開催された研究集会「線形計画問題の新解法」において世界で初めて発表されたものであり、計算とアルゴリズムの分野で日本人によって印された大きな業績の一つであるといえよう(田邊, 1987; Tanabe, 1988; 小島 他, 1987; Kojima et al., 1989; 刀根, 1987)。

数理科学の他の分野同様、最適化には、“モデリング・数理・アルゴリズム”の 3 つの側面がある。本論では、著者と共同研究者のこれまでの研究より、数理とアルゴリズムの側面を扱ったものとして、凸最適化に対する内点法の情報幾何(Ohara and Tsuchiya, 2007; 土谷・小原, 2008; Kakiyama et al., 2009, 2011, 2013)、そして凸最適化によるモデリングの例として内点法による大規模グラフィカルモデル推定(Ueno and Tsuchiya, 2009)、古代メソポタミア都市の人口動態

[†] 政策研究大学院大学：〒106-8677 東京都港区六本木 7-22-1

推定(上田, 2010; 上田 他, 2011)について紹介する.

2. 内点法と情報幾何

情報幾何(甘利・長岡, 1994; Amari and Nagaoka, 2007)は統計科学や機械学習, 信号処理, 制御などの分野に対する微分幾何学的枠組みである. 数学的には凸性の中で重要な役割を果たす. 近年, 筆者らにより, 凸最適化の情報幾何の理論が展開され(Ohara and Tsuchiya, 2007; 土谷・小原, 2008; Kakihara et al., 2009, 2011, 2013), 内点法によって凸最適化問題を解く計算の手間を, 厳密に, 問題の構造に由来する微分幾何学的な曲率積分で表すことができるようになったので, この理論について紹介する.

2.1 凸錐上の線形計画問題

\mathbf{E} を n 次元実ベクトル空間, そして $\Omega \subset \mathbf{E}$ をその上の内点を持ち直線を含まない開凸錐とする. \mathbf{E}^* を \mathbf{E} 上の線形関数の作る空間とする. \mathbf{E} と \mathbf{E}^* の元が作る(双対)積を $\langle \cdot, \cdot \rangle$ と記すことにする. 凸錐 Ω の開双対錐を $\Omega^* \equiv \{s \in \mathbf{E}^* \mid \langle s, x \rangle > 0, \forall x \in \text{cl}(\Omega) \setminus \{0\}\}$ とする. 2つの凸錐 Ω と Ω^* を用いて定義される, 次の最適化問題

$$(2.1) \quad \min \langle c, x \rangle, \quad \text{s.t. } x \in (d + \mathbf{T}) \cap \text{cl}(\Omega)$$

および

$$(2.2) \quad \min \langle d, s \rangle, \quad \text{s.t. } s \in (c + \mathbf{T}^*) \cap \text{cl}(\Omega^*),$$

を考える. ここで, $c \in \mathbf{E}^*$, $d \in \mathbf{E}$, $\mathbf{T} \subset \mathbf{E}$ は $n - m$ 次元線形部分空間, $\mathbf{T}^* \subset \mathbf{E}^*$ は m 次元双対線形部分空間であり, \mathbf{T} と \mathbf{T}^* とは互いに「直交」するものとする; すなわち, 任意の $x \in \mathbf{T}$ と $s \in \mathbf{T}^*$ に対し, $\langle x, s \rangle = 0$ が成立する. 最適化問題(2.1)と(2.2)は互いに双対であり, ごく緩い条件の下で等価である. これらの問題を凸錐上の線形計画問題という. Ω を第一象限とすると線形計画問題, 半正定値対称行列の錐とすると半正定値計画問題, 2次錐の直積とすると, 2次錐計画問題が得られる. これらの場合, Ω は対称錐と呼ばれる特別の形式の錐であり, 多項式時間内点法の一般論が構築できることが知られている(Nesterov and Nemirovski, 1994). そして, さまざまな特長を持った実用的なソフトウェアパッケージが公開されている.

本節以降, 主問題(2.1)の内点許容解の集合 $(d + \mathbf{T}) \cap \Omega$ を \mathcal{P} , 双対問題(2.2)の内点許容解の集合 $(c + \mathbf{T}^*) \cap \Omega^*$ を \mathcal{D} と記す. そして \mathcal{P} と \mathcal{D} が共に空集合ではないと仮定する. 双対理論より, この時, 両問題は $\langle \hat{x}, \hat{s} \rangle = 0$ となる最適解 $\hat{x} \in \text{cl}(\mathcal{P})$, $\hat{s} \in \text{cl}(\mathcal{D})$ を有する.

問題(2.1)においては, $x \in \Omega$ が一番難しい条件である. 内点法では, 解法の過程においてこの条件を自動的に満たすために, 障壁関数 $\psi(x)$ を用いる. 障壁関数は, 領域 Ω の境界に近づくにつれて値が無限大に発散する凸関数である. そして, 多項式時間で問題を解くためには, p -正規障壁関数(Nesterov and Nemirovski, 1994)と呼ばれる, 性質の良い凸関数を用いる(p は Ω の形状より定まるパラメータである). なお, 線形計画問題の場合には, $\psi(x) = -\sum_{i=1}^n \log x_i$, 半正定値計画問題の場合には $\psi(x) = -\log \det(x)$ である. パラメータ p は, 線形計画問題では非負変数の数(次元), 半正定値計画問題の場合には, 最適化が行われる対称行列の行(列)数である.

p -正規障壁関数 ψ を \mathcal{P} 上で最小化する点を解析的中心という. 解析的中心の計算は, ψ を 2次近似した凸 2 次関数の最小化を反復する方法, すなわちニュートン法によって, 効率よく行えることが知られている. これが, 内点法で凸最適化問題を効率的に解く上での一つの肝となる性質である.

内点法では, 解析的中心から最適解に至る中心曲線という曲線(Nesterov and Nemirovski,

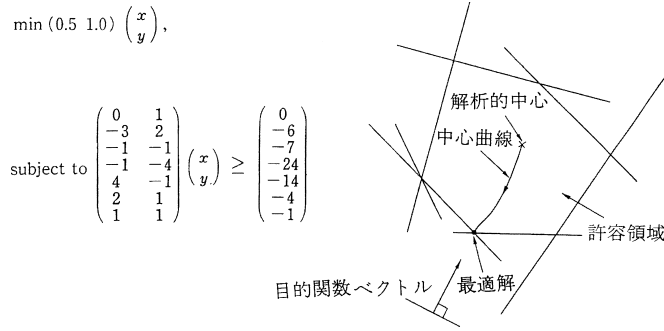


図 1. 線形計画問題と内点法の基礎概念.

1994) を辿って最適解に到達する. 中心曲線 γ_P は p -正規障壁関数を利用して定義される, $t > 0$ をパラメータとする以下の最適化問題の最適解 $x_P(t)$ の集合である:

$$(2.3) \quad \min t\langle c, x \rangle + \psi(x), \quad x \in P.$$

$x_P(t)$ における目的関数と最適値との差は, p/t で上から押さえられる. したがって, t を増やしながら (2.3) を近似的に解き, 中心曲線を近似的 (数值的) に辿ることで問題 (2.1) を解くことができる. 障壁関数の性質より, 問題 (2.3) はアフィン空間上の無制約最適化となり, 条件 $x \in \Omega$ を陽に取り扱わずに済む. 問題 (2.3) を解くために, 目的関数を 2 次関数で近似し近似関数を (P 上で) 最小化する方向に適当なステップ幅を進む, ニュートン法を用いる. 図 1 に線形計画問題と中心曲線の例を示す.

そしてこの方法により, 主問題 (2.1) に対し, パラメータ t_1 の中心曲線上の点 $x_P(t_1)$ の近傍からパラメータ $t_2 > t_1$ の中心曲線上の点 $x_P(t_2)$ の近傍までを, $O(\sqrt{p} \log(t_2/t_1))$ 回の反復で辿ることができる. これが, 計算の立場からの内点法の理論的効率性, すなわち多項式時間性を裏付ける性質である.

2.2 情報幾何

p -正規障壁関数をポテンシャル関数として用いた情報幾何を紹介する. 情報幾何では, ポテンシャル関数から定まるリーマン計量とポテンシャル関数のルジャンドル変換 (勾配写像) を用いて定義される互いに双対な 2 つのアファイン接続が重要な役割を果たす. このような構造を持つ空間を双対平坦空間と呼ぶ. Ω 上に p -正規障壁関数 ψ をポテンシャル関数として定められる双対平坦空間は以下のようなものである. まず, リーマン計量を $\psi(x)$ のヘッセ行列で導入する. そして, ルジャンドル変換を

$$s(x) = -\frac{\partial \psi(x)}{\partial x}$$

と定義する. すると, $s(\Omega) = \Omega^*$ が成立し, さらに, $s(\cdot)$ は Ω と Ω^* の間の一对一の大域的写像となっている. $s(\cdot)$ を Ω に導入されたもう一つの座標と考えることもでき, この座標系に関しては, Ω は「 Ω^* であるように見える」.

双対平坦空間では二つの特徴的なアファイン接続 ∇ と ∇^* が定まる. このうち ∇ は元の x 座標での直線が測地線となるような接続である. 一方 ∇^* は s 座標での直線が測地線となるような接続である.

$\psi(x)$ の共役関数を $\psi^*(s)$ と記す. $\psi^*(s)$ は Ω^* の p -正規障壁関数で, Ω^* 上に ψ^* をポテンシヤ

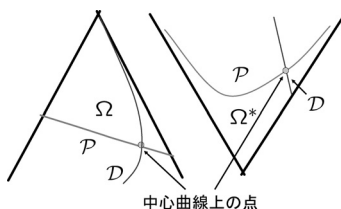


図2. x 座標系(左)と s 座標系(右)における主問題の許容領域 \mathcal{P} と双対問題の許容領域(錐の次元は2, \mathcal{P} , \mathcal{D} の次元は1).

ル関数として導入される双対平坦空間は, Ω 上に ψ に基づいて導入されたものと同型となる. 双対問題に対する中心曲線 $\gamma_{\mathcal{D}}$ は, ψ^* を用いて, (2.3)同様,

$$\min t(d, s) + \psi^*(s), \quad s \in \mathcal{D}$$

の最適解 $s_{\mathcal{D}}(t)$ が成す曲線と定義される. なお, 線形計画問題の場合には, $\psi^*(s) = -\sum_{i=1}^n \log s_i - n$, 半正定値計画問題の場合には $\psi^*(s) = -\log \det(s) + (n \text{ の関数})$ となり, 事実上 ψ と一致する. これは, これらの場合, Ω が自己双対性 $\Omega = \Omega^*$ を有することの反映の一つである.

次いで自己平行部分多様体を定義する. 自己平行部分多様体は, 平らな部分多様体として, 接続に応じて定義される. ∇ (∇^*)-自己平行部分多様体は, x 座標系 (s 座標系)でアフィン空間(線形空間を平行移動したもの)の開集合としてかけるものである.

2.3 中心曲線の情報幾何的特徴付けと予測子・修正子法

ルジャンドル変換を用いると Ω と Ω^* を同一視できる. このことを利用して, 我々は主問題と双対問題を同じ空間に埋め込んで考える. すると, 問題の形より, 主問題の許容領域 \mathcal{P} は, $(n-m)$ 次元 ∇ -自己平行部分多様体となり, 双対問題の許容領域 \mathcal{D} は, m 次元 ∇^* -自己平行部分多様体となる. 通常 \mathcal{P} は接続 ∇^* に関しては曲がっており, \mathcal{D} は接続 ∇ については曲がっていることに注意したい(図2). \mathcal{P} と \mathcal{D} は1点で交わり, 交点 $\mathcal{P} \cap \mathcal{D}$ は, $t=1$ での中心曲線上の点となる.

この事実を認め, 少々想像を逞しくすれば, $t=t'$ での中心曲線上の点は $\mathcal{P} \cap t'\mathcal{D}$ として表されることが分かるであろう. ここで, $t'\mathcal{D}$ は \mathcal{D} を s 座標において t' 倍した集合である. ここで, \mathcal{D} を同次化した, $m+1$ 次元 ∇^* -自己平行部分多様体 $\text{Hom}(\mathcal{D})$ を以下のように定義する.

$$\text{Hom}(\mathcal{D}) = \{s' \in \Omega^* \mid s' = ts, t > 0, s \in \mathcal{D}\}$$

一次元多様体として, 中心曲線 $\gamma_{\mathcal{P}}$ を $\gamma_{\mathcal{P}} = \{\gamma_{\mathcal{P}}(t) \mid t \in (0, \infty)\}$ と表す. ここで $x_{\mathcal{P}}(t) = x(\gamma_{\mathcal{P}}(t))$ である. 上で述べた同次化された双対問題の許容領域 $\text{Hom}(\mathcal{D})$ を用いて中心曲線は $\gamma_{\mathcal{P}} = \mathcal{P} \cap \text{Hom}(\mathcal{D})$ と表現できる.

この定式化にもとづいて, s 座標を用いて中心曲線を $\text{Hom}(\mathcal{D})$ 上で辿る, 以下のような予測子・修正子法を考える. $\mathcal{N}(\beta)$ で, 中心曲線 $\gamma_{\mathcal{P}}$ の近傍を表す. $\beta \in [0, 1)$ で, β が0で中心曲線に一致し, β が大きくなるにつれて広がる.

予測子・修正子法 (図3参照)

1. $\gamma_{\mathcal{P}}(t)$ をパラメータ t の中心曲線 $\gamma_{\mathcal{P}}$ 上の点とし, $s(t) = s(\gamma_{\mathcal{P}}(t))$, $\dot{s}(t)$ を $s(t)$ での接線方向とする.

2. (予測子) $s_L(t + \Delta t) = s(t) + \Delta t \dot{s}$ とし, $\Delta t_{\max} > 0$ を $s_L(t + \Delta t) \in \mathcal{N}(\beta)$ となる最大の Δt とする. ($s_L(t + \Delta t_{\max}) \in (t + \Delta t_{\max})\mathcal{D}$ となる.)

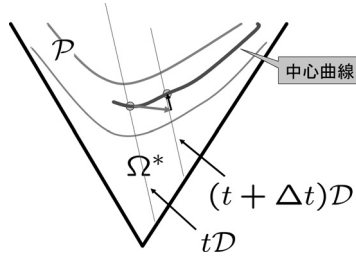


図 3. 予測子・修正子法(錐の次元は 3, \mathcal{P} の次元は 2 (\mathcal{P} は曲面), D の次元は 1 (D は(半)直線)).

3. (修正子) $s_L(t + \Delta t_{\max})$ から $s(t + \Delta t_{\max})$ に移動する.
4. $t := t + \Delta t$ としてステップ 1 に戻る.

このアイデアに基づいて、実際に計算機上に実装可能な $O(\sqrt{p} \log(1/\varepsilon))$ 反復の多項式時間アルゴリズムを構成できる.

2.4 埋め込み曲率と内点法の反復回数の評価

Ω の部分多様体の埋め込み曲率を定義する. 部分多様体 \mathcal{M} 上の点 x を考える. x における \mathcal{M} 上の接空間を $T_x(\mathcal{M})$ とする. リーマン計量の意味で $T_x(\mathcal{M})$ の直交補空間への射影を Π_x^\perp とする. X, Y を $T_x(\mathcal{M})$ とし, 接続 ∇ に関する X の Y についての共変微分 $\nabla_Y X$ を考える. すると, 通常 $\nabla_Y X \notin T_x(\mathcal{M})$ となる. この時, 埋め込み曲率 $H_{\mathcal{M}}(X, Y)$ を, $\Pi_x^\perp \nabla_Y X$ として定義する. 接続 ∇^* に関する埋め込み曲率を同様に定義し, $H_{\mathcal{M}}^*(\cdot, \cdot)$ と記す.

この時, 次の定理が成立する:

定理 1. $0 < t_1 < t_2$ とする. s_1 を $\gamma_{\mathcal{P}}(t_1)$ の近傍の点とする. $K(s_1, t_2, \beta)$ を s_1 から近傍 $\mathcal{N}(\beta)$ を通って予測子・修正子法で $\gamma_{\mathcal{P}}(t_2)$ の近傍の点を求めるのに必要な反復回数とすると, $K(s_1, t_2, \beta)$ は以下のように表される:

$$K(s_1, t_2, \beta) = \frac{1}{\sqrt{\beta}} I_{\mathcal{P}}(t_1, t_2) + \frac{o(1)}{\sqrt{\beta}}, \quad I_{\mathcal{P}}(t_1, t_2) \equiv \int_{t_1}^{t_2} \frac{1}{\sqrt{2}} \|H_{\mathcal{P}}^*(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}})\|^{1/2} dt.$$

この定理は, 積分

$$I_{\mathcal{P}}(t_1, t_2) \equiv \int_{t_1}^{t_2} \frac{1}{\sqrt{2}} \|H_{\mathcal{P}}^*(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}})\|^{1/2} dt$$

が, 内点法を用いて主問題を解く上での困難さを表すことを示している. 同様にして,

$$I_{\mathcal{D}}(t_1, t_2) \equiv \int_{t_1}^{t_2} \frac{1}{\sqrt{2}} \|H_{\mathcal{D}}(\dot{\gamma}_{\mathcal{D}}, \dot{\gamma}_{\mathcal{D}})\|^{1/2} dt$$

は, 内点法を用いて双対問題を解く上での困難さを表すことも示すことができる.

さらに, (古典的) 線形計画問題, 半正定値計画問題, さらにこれらを含む対称錐計画問題に対する主双対内点法の反復回数について, 次の微分幾何学的表現が得られる (Kakihara et al., 2009, 2011, 2013; Ohara and Tsuchiya, 2007).

定理 2. $0 < t_1 < t_2$ とする. s_1 を $\gamma_{\mathcal{P}}(t_1)$ の近傍の点とする. 古典的線形計画問題, 半正定値計画問題において, $K(s_1, t_2, \beta)$ を s_1 から近傍 $\mathcal{N}(\beta)$ を通って主双対予測子・修正子法で $\gamma_{\mathcal{P}}(t_2)$

の近傍の点を求めるのに必要な反復回数とすると、 $K(s_1, t_2, \beta)$ は以下のように表される：

$$K(s_1, t_2, \beta) = \frac{1}{\sqrt{\beta}} I_{PD}(t_1, t_2) + \frac{o(1)}{\sqrt{\beta}}, \quad I_{PD}(t_1, t_2) \equiv \int_{t_1}^{t_2} h_{PD}(t)^{1/2} dt.$$

ここで、

$$h_{PD}(t)^2 = \left(\frac{1}{2} \|H_{\mathcal{P}}^*(\dot{x}_{\mathcal{P}}(t), \dot{x}_{\mathcal{P}}(t))\| \right)^2 + \left(\frac{1}{2} \|H_{\mathcal{D}}(\dot{s}_{\mathcal{D}}(t), \dot{s}_{\mathcal{D}}(t))\| \right)^2.$$

(近傍 $\mathcal{N}(\beta)$ は、主双対内点法についての中心曲線の近傍で、 $\beta \in [0, 1)$ で、 $\beta = 0$ の時には中心曲線に一致し、 β が大きくなるにつれて近傍が広がる.)

古典的線形計画の場合、中心曲線の“全曲率”を、問題を記述するのに要する入力ビットサイズと次元の多項式で評価できる。すなわち次の定理が成立する (Kakihara et al., 2009, 2011; Ohara and Tsuchiya, 2007)。

定理 3. 古典的線形計画の場合、全曲率 $I_{\mathcal{P}}(0, \infty)$, $I_{\mathcal{D}}(0, \infty)$, $I_{PD}(0, \infty)$ が存在して有界であり、上から

$$\min[I_{\mathcal{P}}(0, \infty), I_{\mathcal{D}}(0, \infty)] \leq I_{PD}(0, \infty) = O(n^{3.5} L_A)$$

と評価できる。ここで、 L_A は線形計画問題の係数行列の入力ビットサイズである。特に A が 0-1 行列からなる組み合わせ線形計画問題の場合、問題の次元のみの多項式で以下のように評価できる。

$$\min[I_{\mathcal{P}}(0, \infty), I_{\mathcal{D}}(0, \infty)] \leq I_{PD}(0, \infty) = O(n^{4.5} m).$$

上で紹介した定理 1, 2 における反復回数の評価は漸近的なものであるが、これらの定理より、 $\beta \times$ (内点法の反復回数) は β に依存せずに、ほぼ一定値となり、中心曲線上の積分値 $I_{\mathcal{P}}$, $I_{\mathcal{D}}$, I_{PD} となるのがわかる。さらに、 $\beta = 1$ とした場合の反復回数は、積分値そのものとなる。以下、この関係が大規模問題について成立することを数値実験で確かめる (Kakihara et al., 2009, 2011)。

図 4 は線形計画問題の例である。実験に用いた問題 DFL001 は、12230 変数、6072 制約式で、線形計画問題の標準ベンチマーク問題集 Netlib の中のものである。図 4(a) は近傍の大きさ β をさまざまに変えて主双対内点法で問題を解いた結果である。横軸は反復回数、縦軸が \log (正規化された双対ギャップ) である。正規化された双対ギャップは中心曲線上のどこに反復点が存在するかを表すもので、最適値に近づくにつれて 0 に近づく。 β の値が小さいと、より反復回数が増えることが観察されるが、 β に依らず、正規化された双対ギャップの減り方には同じ傾向が見られる。そこで、(a) の横軸を $\sqrt{\beta}$ 倍したものを新たに横軸にとると、異なる β のグラフがほぼ重なることがわかる。この事実は、積分 I_{PD} が驚くべき正確さで反復回数を近似していることを示している。半正定値計画問題でも同様のことが観察される (図 5)。実験に用いた問題 EqualG51 は、行列サイズ 1000×1000 、制約式数 1001 の問題である。図 5 の見方は図 4 と同様であり、反復回数が積分でよく近似されていることが見てとれる。つまり内点法の反復回数は、中心曲線の曲率を反映した微分幾何学的量の積分そのものであると見ることができるのである。

ここで興味深いのは、線形計画問題の積分値が 100 をしばしば超えるのに対し、半正定値計画問題の積分値は 40 程度に留まっていることである。これは、幾何学的視点からは、半正定値計画問題は必ずしも線形計画問題よりも難しくはなく、むしろ易しいという側面もあることを示唆している。すなわち、半正定値計画問題の求解が困難なのは、本質的非線形性よりも数値的悪条件が理由である可能性が高い。

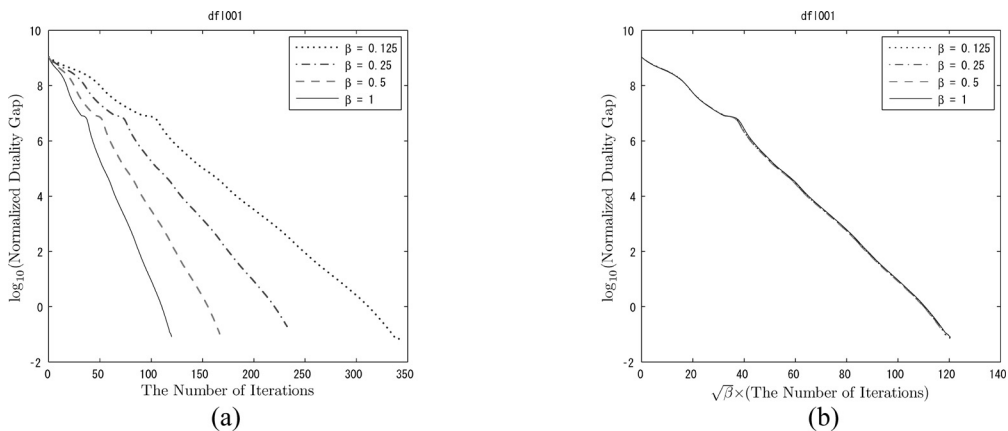


図 4. 線形計画の場合：DFL001 (12230 変数 6072 制約式；最適値： 1.12664×10^7) .

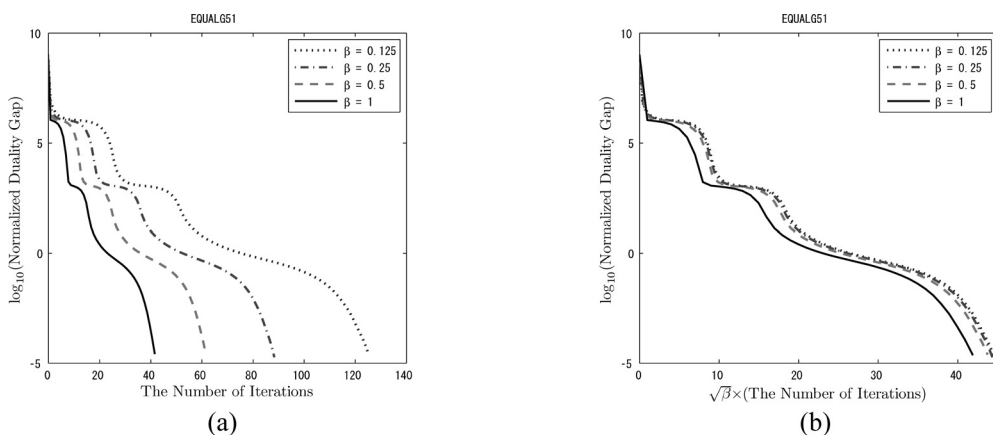


図 5. 半正定値計画の場合：EqualG51 (行列サイズ：1000 × 1000, 1001 制約式；最適値： 4.005601×10^3) .

この現象には、錐の間の組合せ的構造の複雑さが関係している可能性がある。つまり線形計画問題は単純ではあるが多数の錐の直積であるのに対し、半正定値計画問題は、それに比べると少数の錐の直積であり、より“単純な”構造をしている(場合が多い)。これはあくまで一つの仮説であり、その解明は今後の課題である。

3. 大規模グラフィカルモデル推定

3.1 グラフィカルガウシアンモデル

グラフィカルガウシアンモデル (例えば Lauritzen, 1996) は、 n 個の連続変数 $x = (x_1, \dots, x_n)$ があつたときに、それらが、分散共分散行列の逆行列 $D = \{d_{ij}\}$ の非ゼロパターンが指定されているような多変量正規分布にしたがうと仮定するモデルであり、 x の密度関数は以下のよう

に与えられる.

$$(3.1) \quad N(x; \mu, D) = \frac{1}{\sqrt{(2\pi)^n \det D^{-1}}} \exp \left[-\frac{(x - \mu)^T D (x - \mu)}{2} \right],$$

このモデルの分散共分散行列は D^{-1} である. D の (非対角要素の) 非ゼロパターンと, 各変量を頂点とし, D の非対角非ゼロ要素に対応する変量の対を枝とする無向グラフとを自然に対応させることができる. このグラフを $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ と記すことにしよう. ここで, \mathcal{V} と \mathcal{E} はそれぞれ頂点と枝の集合である. 変量 i と j を結ぶ無向グラフの各枝を (i, j) と表すことにする ($i < j$) の場合のみを考える). なお, 正定値性より, D の対角要素は常に正である.

$\mathcal{I} = \{(1, 1), \dots, (n, n)\}$ とすると, D の非ゼロ要素の集合は, $\mathcal{J} \equiv \mathcal{I} \cup \mathcal{E}$ となる. E_{ij} を, (i, j) 要素と (j, i) 要素が 1 であとは 0 であるような $n \times n$ 対称行列とすると, D を次のように表現することができる.

$$(3.2) \quad D = \sum_{(i, j) \in \mathcal{J}} y_{ij} E_{ij}, \quad D \succ 0.$$

ここで, y_{ij} は実数, $D \succ 0$ は D が正定対称行列であることを表す. (3.2) を満たす D の集合を \mathcal{S} と記す. グラフィカルガウシアンモデルは (3.1) と (3.2) によって表現されるモデルである. $\mathcal{G}(\mathcal{V}, \mathcal{E})$ を (無向) 独立グラフと呼ぶ. 独立グラフは対ごとの条件付き独立性を視覚的に表現したものである. つまり, $(i, j) \in \mathcal{P}$ が \mathcal{J} に含まれない時, そしてその時に限り, x_i と x_j は, 残りの変数を固定するという条件の下で独立となる.

標本平均が \bar{x} , 標本分散共分散行列が \hat{S} であるような m 個のサンプルが与えられたとすると, グラフ \mathcal{G} によって規定されるグラフィカルガウシアンモデルの対数尤度関数は次のように書ける:

$$(3.3) \quad l(x; \mu, D) = -\frac{mn \log 2\pi}{2} - \frac{m}{2} (\bar{x} - \mu)^T D (\bar{x} - \mu) + \frac{m}{2} [-\text{Tr}(\hat{S}D) + \log \det D].$$

ここで, D は (3.2) で与えられた条件を満たす. (3.3) を最大化する μ, D がモデルの最尤推定である. 平均の最尤推定が $\mu = \bar{x}$ で与えられることは上の式の形から簡単に分かる. すると, D の最尤推定は,

$$(3.4) \quad \min_D \text{Tr}(\hat{S}D) - \log \det D, \text{ s.t. } D \in \mathcal{S}.$$

の最適解として与えられる.

このように, D の最尤推定量を求める問題は, 線形制約の下での $-\log \det$ 関数と線形関数 and の最小化問題の最適化問題となる. したがって, ガウシアングラフィカルモデルの最尤推定は, 次の半正定値計画問題の中心曲線上の点を求める問題そのものに帰着する.

$$\min_D \text{Tr}(\hat{S}D), \text{ s.t. } D \in \mathcal{S}.$$

この半正定値計画問題の中心曲線は, ν をパラメータとする次の凸計画問題

$$\min_D \text{Tr}(\hat{S}D) - \nu \log \det D, \text{ s.t. } D \in \mathcal{S},$$

の最適解として与えられ, したがって, $\nu = 1$ に対する中心曲線上の D が最尤推定になる. そこで, グラフィカルガウシアンモデルの最尤推定を計算するために, 半正定値計画問題に対する内点法を利用することができる.

3.2 データ同化にグラフィカルガウシアンモデルを活用する

さて, 現在, 海洋学や気象学の分野を中心として, 観測データに偏微分方程式モデルを当てはめ, 観測不可能な場所での物理量を推定することを目的とするデータ同化という分野の研究

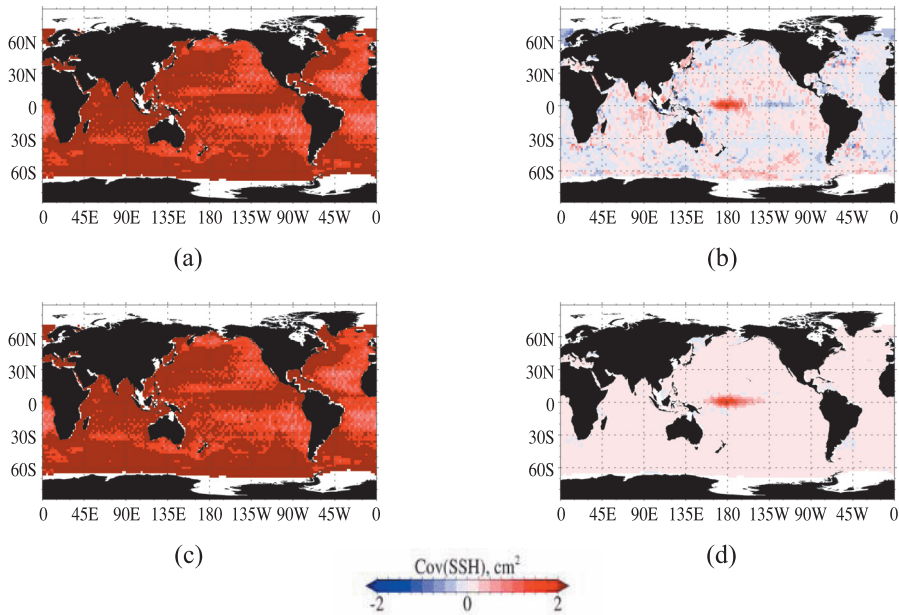


図 6. 海面高度の分散・共分散. (a) データによる各点の分散, (b) データによる中心点(緯度 0 度経度 180 度)と各点の共分散, (c) 推定された正規分布による各点の分散, (d) 推定された正規分布による中心点と各点との共分散 (正規分布の次元: 8585, パラメータ数 119682, 近傍の大きさ 28).

が進められている (淡路 他, 2009; 中村 他, 2005). データ同化でよく用いられるのが Ensemble Kalman Filter と呼ばれる方法であるが, モデルの同定の過程で, 人工衛星により取得される, メッシュ状に設定された数千から数万の観測点での海面高度データに対し, 正規分布を当てはめる必要が生じる. ただし, データ数は 300 程度, 正規分布の次元はメッシュの点数で数千から数万となる.

この正規分布をグラフィカルモデルにより推定する. 分布 D の要素の内, 各グリッド点のある近傍に対応する要素のみが非ゼロで残りの要素はゼロであるという構造を仮定し, 内点法によって (3.4) を解いて最尤推定を行う. 近傍の大きさをを変えて情報量規準によって適切な近傍の大きさを定める. 詳細は Ueno and Tsuchiya (2009) に譲るが, これは, 球面上を緯度経度 2 度ごとのメッシュに分割した場合は 8585 次元, そして緯度経度 1 度ごとのメッシュに分割した場合は 34300 次元の大規模正規分布を推定する問題となる. そして, 現在, スパコンを用いて, D の次元が 8585×8585 , 150381 パラメータのガウシアングラフィカルモデル, D の次元が 34300×34300 , 101310 パラメータのガウシアングラフィカルモデルの最尤推定を行うことが可能である. これらのモデルはブロック構造を持たず, おそらく現時点では最大規模のグラフィカルガウシアンモデル推定およびパラメトリックモデル選択であろう. 図 6 にモデル選択後の推定結果を示す.

4. 古代メソポタミア社会の人口動態の復元

次に取り上げるのは, 古代メソポタミア社会の人口動態の復元という問題である. この問題

に対して、凸 2 次計画法を用いた接近法を試みた(上田, 2010; 上田 他, 2011).

イラク・キルクーク北方に紀元前 15 世紀にヌジという都市が存在した. ヌジ遺跡からはさまざまな社会契約が記された粘土板が 4000 枚以上発掘されており, 古代メソポタミア社会の生活に関するさまざまな手がかりとなってきた. 各粘土板に記されている人名を抜き出して索引としたヌジ人名史料と呼ばれる史料が作成されている. その特徴は以下の通りである.

1. 登場する人名の多くは, “A の息子 B”, “A の父親 B”等, 断片的ながらも家族関係(主として親子・兄弟関係)がわかる形で記されている.
2. 各登場人物について, 登場する文書とその文書において登場する行が記されている.(登場する文書のみの場合もある.)

1 と 2 の情報から家系図を作成することができる. 一番大きな家系は 6 世代 16 人からなる家系である. この家系はヌジ社会において中心的な役割を示していることが知られている. 一族の中でも一番多く文書に登場する人物は, ミタンニ国と結びついた行政官であったとされる, ティヒプティラである.

さて, 同じ文書に記されている人物は, 同じ時期に生存していたと考えられる. また, 親子関係にある人物においては, 親の方が先に生まれているなどの条件が成り立つ. そこで, 各登場人物について, ヌジ人名史料や家系図から得られる情報と矛盾がないように, 生誕年と死亡年を定めることを試みる. これにより, 各年に何人程度の人物が同時に生存していたか, あるいは文書はどの時期に作成されたかについての情報を得ることができる. 以下ではこの問題を凸 2 次計画問題に定式化する. この際に以下の仮定を置く:

- (A) 人物 i の“寿命”は μ_i である.
- (B) 子が生まれた時, 父親は 15 歳以上, 母親は 20 歳以上である.
- (C) 文書に登場する時点で人物は 10 歳以上である.

文書に記載されている人物の数は 10244 人である. また, 文書数は 1822 である. 各人に対して寿命をあらかじめ想定することは大胆なように思われるが, 寿命を想定することは, その社会の歴史について, 一つのシナリオを仮定することに対応する. いくつもの妥当なシナリオに対して推定を行い, 同じ結果が得られれば, その結果が真である確度は高いといえよう.

変数は, 各人物 i の生誕年 b_i , 死亡年 d_i である. 当然, 各人物について,

- (D) 死亡年は生誕年よりも後である, $b_i \leq d_i$

が成立する. また, 文書 k の成立年を P_k とすると,

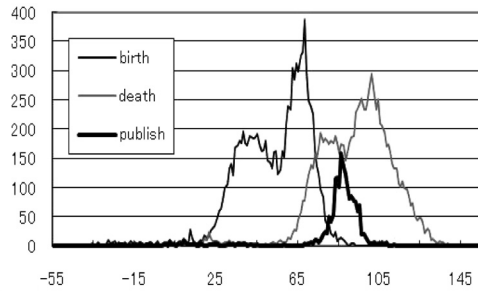
- (E) 文書 P_k に登場する人物 i の死亡年について, $P_k \leq d_i$

が成立する. 以上 (A) から (E) の設定(制約)の下で最小化すべき目的関数を

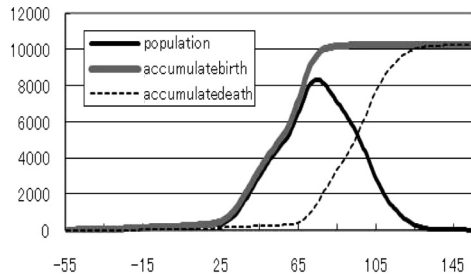
$$\sum_i (d_i - b_i - \mu_i)^2$$

として, 矛盾なく各人物の生存期間を時間軸上に配置することを試みた. もし, 最初に与えた寿命通りに矛盾なく生誕年, 死亡年を定めることが可能であれば, 目的関数値は 0 となるが, 実際には, ティヒプティラの家系の人物などの場合には, 最初に与えた寿命よりも長い寿命としないとすべての条件を満たすことができない.

このようにして, 文書や家系図と矛盾しないように, 各人物の生誕年と死亡年, そして文書の成立年を定める問題を, 約 22301 変数, 39311 制約式の凸 2 次計画問題として定式化した. 問題の求解にあたっては数理システムの NUOPT を使用した. シナリオとしては, 各人物の寿



(a)



(b)

図 7. 各人物の寿命を [20, 60] の一様分布で与えた場合の一例. (a) 各年毎の文書に登場する人物の生誕数, 死亡数および文書の作成数, (b) 文書に登場する人物の生誕, 死亡の累積および人口の推移. (文書 JEN523 が作成された年を 100 年として時間軸を定めている; このシナリオでは, ティヒプティラは 52.91 年から 107.91 年まで生きたことになっている.)

命を [20,60] の(様々な)一様分布に従うとしたもの, 一様に寿命を 40 としたものを想定した. その結果, どのシナリオでも

- (1) 文書に現れる人物が最大同時に生存した数は, 約 8000 人から 9000 人程度となる.
- (2) 文書は一時期(10-20 年の間)に大量に作成された.

などの事実が判明した(図 7). ヌジ社会は 100 年程度存続していたと見られており, その中で文書が一時期に大量に作成されていたという事実は, 社会構造, 権力構造の変化, 特に, 財産がいくつかの特定の家系に集約されていった過程を窺わせるものであり, 興味深い.

女性は文書の中には多く現れないもののほぼ同人数程度いた筈であること, そして 10 歳以下の子供が含まれていないことなどより, ヌジには (最低)約 20000 人程度の人間が住んでいたであろうと考えられる. これを手がかりとしてももう少し考察を進めてみる.

当時の都市遺跡では, 1 エーカー(4047m²)に 100 人程度が住んでいたと考えられることより, 約 800000m² の居住領域が必要である. ヌジ遺跡はその面積が 38100m² であり, したがってそこには約 1000 人程度が住んでいたと考えられる. このような集落が 20 個程度あってヌジ社会を構成していたと思われる. また, 当時の小麦の収量を 10 アールあたり 300kg とし, これで 2 人を 1 年間養えると考えれば, 20000 人を養うには, 約 10km² の畑が必要となる. した

がって、少なくとも、ヌジでは 10 km^2 の広さの土地に、1000 人程度の大きさの 20 程度の集落が存在していたと考えることができる。

ヌジ人名史料で索引化されているのは、4000 枚といわれる粘土板の内 2000 枚弱であることや、未だに発掘されていない粘土板もあることを考慮すると、これらの数値はあくまでも人口や社会の大きさの下限を与えるものと考えることが妥当であろうが、凸最適化モデリングによって初めて得られる古代社会についての具体的な知見の一つである。

5. おわりに

凸最適化に纏わる“モデリング・数理・アルゴリズム”について行ってきた研究を紹介した。最適化は勝れて横断的な学問分野である。その特性を生かして今後もさまざまな分野と交流しつつ研究を進めていければと考えている。いうまでもなく、これらは最適化と計算を軸として展開した統計数理の研究である。学問としての統計数理は統計数理研究所が培ってきた伝統、世の中からの社会的・技術的要請、各研究者の知的好奇心によって方向づけられ、間断ない研究への情熱と努力によって進歩していくものであろう。本稿は、1986 年より 2010 年まで 24 年間に渡って統計数理研究所に在籍した著者がささやかながら培ってきた統計数理の側面をまとめたものでもあることを申し添えたい。

なお、本稿の第 2 節の内容は、小原敦美福井大学電気電子工学科教授、柿原聡博士(政策研究大学院大学)、第 3 節の内容は上野玄太統計数理研究所准教授、第 4 節の内容は伊藤栄明統計数理研究所名誉教授、牧野久実鎌倉女子大准教授、上田澄江統計数理研究所元助手との共同研究の成果である。ここに記して深甚たる感謝の意を表したい。

最後に、著者に研究の機会を与え、いかなるときも大らかに暖かく支援してくれた統計数理研究所に心より感謝する。

参 考 文 献

- 甘利俊一, 長岡浩司 (1994). 『情報幾何の方法』, 岩波書店, 東京.
- Amari, S.-I. and Nagaoka, K. (2007). *Methods of Information Geometry*, AMS, Providence.
- 淡路敏之・蒲地政文・池田元美・石川洋一 編 (2009). 『データ同化: 観測・実験とモデルを融合するイノベーション』, 京都大学学術出版会, 京都.
- Kakihara, S., Ohara, A. and Tsuchiya, T. (2009). Information geometry and primal-dual interior algorithms, Research Memo., No. 1120, The Institute of Statistical Mathematics, Tokyo.
- Kakihara, S., Ohara, A. and Tsuchiya, T. (2011). Curvature integrals and iteration complexities in SDP and symmetric cone programs, Manuscript.
- Kakihara, S., Ohara, A. and Tsuchiya, T. (2013). Information geometry and interior-point algorithms in semidefinite programs and symmetric cone programs, *Journal of Optimization Theory and Applications*, **157**, 749–780.
- 小島政和, 水野真治, 吉瀬章子 (1987). 多項式オーダーの主双対内点法, 「線形計画問題の新解法」(刀根薫 編), 統計数理研究所共同研究リポート, No. 5, 13–24.
- Kojima, M., Mizuno, S. and Yoshise, A. (1989). A primal-dual interior point algorithm for linear programming, *Progress in Mathematical Programming* (ed. N. Megiddo), 29–47, Springer-Verlag, New York.
- Lauritzen, S. L. (1996). *Graphical Models*, Oxford Science Publications, Oxford, England.
- 中村和幸, 上野玄太, 樋口知之 (2005). データ同化: その概念と計算アルゴリズム, 統計数理, **53**(2), 211–229.

- Nesterov, Yu. and Nemirovski, A. (1994). *Interior-point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia.
- Ohara, A. and Tsuchiya, T. (2007). An information geometric approach to polynomial-time interior-point algorithms: Complexity bound via curvature integral, Research Memo., No. 1055, The Institute of Statistical Mathematics, Tokyo.
- 田邊國士 (1987). Complementarity-enforcing centered Newton method for mathematical programming, 「線形計画問題の新解法」(刀根薫 編), 統計数理研究所共同研究リポート, No. 5, 118–144.
- Tanabe, Kunio (1988). Centered Newton method for mathematical programming, *Lecture Notes in Control and Information Sciences*, Vol. 113, 197–206, Springer, Berlin Heidelberg.
- 土谷 隆, 小原敦美 (2008). 内点法と情報幾何—計算の複雑さへの微分幾何学的アプローチ, 数学セミナー, 3月号, 38–43.
- 刀根 薫(編) (1987). 線形計画問題の新解法, 統計数理研究所共同研究リポート, No. 5.
- 上田澄江 (2010). 人文社会データへの統計数理的アプローチ—古代社会の家系図の復元と人口推定および語順規則による言語の分類—, 総合研究大学院大学博士論文.
- 上田澄江, 牧野久実, 伊藤栄明, 土谷 隆 (2011). 古代社会の復元と人口推定, 統計数理研究所共同研究リポート, No. 267, 128–148.
- Ueno, G. and Tsuchiya, T. (2009). Covariance regularization in inverse space, *Quarterly Journal of the Royal Meteorological Society*, (135), Iss. 642, 1133–1156.

Interior-point Algorithms, Information Geometry and Optimization Modeling

Takashi Tsuchiya

National Graduate Institute for Policy Studies

This paper deals with three topics related to algorithm, mathematics and modeling aspects of convex optimization. First it deals with the information geometric framework of interior-point algorithms. The iterative complexity of interior-point algorithms is expressed in terms of an information geometric integral. Through experiments on fairly large instances, the number of iterations of the interior-point algorithm is shown to be a differential geometrical integral over the central trajectory, both in classical linear programs and in semidefinite programs. The second topic is an application of the interior-point method to estimation of a large-scale normal distribution in data assimilation. The size of the matrix is up to 34300×34300 without any block structure, and the number of parameters in the model, which is compared with AIC or BIC, is up to 150381. In the last topic, we estimate the population of a village in ancient Mesopotamia called Nuzi using convex quadratic programming. The estimation is based on personal records written in Cuneiform on clay tablets excavated from the ruins.