最小二乗法を巡って

土谷　隆

最小二乗問題は最も基本的な最適化問題の一つである．本稿ではまず，重み付き最小二乗法やその極限としての層別最小二乗法，最小二乗問題の条件数などを紹介する．そして線形計画問題に対する内点法，特に旧ソ連で発見されたアフィンスケーリング法や主双対内点法，計算複雑度が係数行列にしか依存しない多項式時間解法である Vavasis–Ye アルゴリズムにおける最小二乗問題の役割について述べる．さらに，リニアモーターカー最適磁気シールド設計や凸最適化の情報幾何などさまざまな分野との関係について，著者の過去の研究とのかかわりをたどりながら議論する．

キーワード：最小二乗問題，重み付き最小二乗問題，層別最小二乗問題，線形計画問題，アフィンスケーリング法，Vavasis–Ye アルゴリズム，リニアモーターカー，情報幾何

1. はじめに

久しぶりにオペレーションズ・リサーチに書くことになった．前回は「モデリング雑感」ということでモデリングについて書いた [20]．8 年経って今読み返してみると，自分の見方がまた変わってきていることにも気がつく．しかし，これについてはまた機会があれば改めて書くことにして，今回はもう少し技術的なことを取り上げたい．最小二乗法の一風変わった側面と線形計画問題，そして自分の研究とのかかわり，である．これは「モデリング雑感」の「双対編」ともいえるかもしれない．しばらくお付き合いいただければ幸いである．

2. 重み付き最小二乗法と条件数

最小二乗問題

$$\min\|c-A^Ty\|^2$$

は最も基本的な最適化問題の一つである．この問題の解は、$y=(AA^T)^{-1}Ac$ で与えられる（以下，適宜行列の正則性などは仮定する）．ベクトル c のノルムと解 x のノルムの比の上限は，$\|(AA^T)^{-1}A\|$ となる．残差空間では，$\|A^Ty\|\le\|c\|$ となることはよく知られている．

各式に重みが付いている最小二乗問題

$$\min\|D(c-A^Ty)\|^2$$

を考えよう．ここで D は，その対角要素が正の任意の対角行列である．この問題の解は

$$y=(AD^2A^T)^{-1}AD^2c,$$

で与えられ，残差空間では，

$$A^Ty=A^T(AD^2A^T)^{-1}AD^2c$$

となる．重み付き最小二乗問題においてベクトル c と解 y や A^Ty のノルムの比は「D に依存しない形」で有限の値で抑えられるであろうか？興味深いことに，この比は D に依存せずに有界であることが知られている．すなわち，$\mathcal{D}$ を各対角要素が正の対角行列全体の集合として次の値 $\chi_A,\bar{\chi}_A$ が存在する．

$$\chi_A=\sup_{D\in\mathcal{D}}\|(AD^2A^T)^{-1}AD^2\|,$$
$$\bar{\chi}_A=\sup_{D\in\mathcal{D}}\|A^T(AD^2A^T)^{-1}AD^2\|,$$

そして，これらの値は以下の組合せ的な特徴づけができる：

$$\chi_A=\max\{\|A_B^{-1}\| \mid A_B \text{ は } A \text{ の任意の基底行列.}\},$$
$$\bar{\chi}_A=\max\{\|A_B^{-1}A\| \mid A_B \text{ は } A \text{ の任意の基底行列.}\}$$

この事実は，これらの定数と線形計画問題との密接な関係を示唆する．

χ_A や $\bar{\chi}_A$ の存在は Dikin によって [3] において最初に示された．これらを最小二乗問題の条件数と呼ぶ．その後の研究で条件数を計算することが NP 困難な問題であることや，A が整数行列であれば，その入力ビット長を L_A として，$2^{O(L_A)}$ で評価できることなどが示されている．

さて，重みの比が無限大となる極限の場合を考えて

つちや　たかし
政策研究大学院大学
〒 106–8677 港区六本木 7–22–1
E-mail: tsuchiya@grips.ac.jp

みよう．このとき重み付き最小二乗解にも極限が存在し以下のようになる．簡単のために，重みの比の一カ所を無限大にした極限を考える．重みを正のパラメータ w でパラメトライズして，

$$D(w) = \mathrm{diag}\big(w\hat{D}_1 \quad \hat{D}_2\big)$$

とする．$\hat{D}_1$，$\hat{D}_2$ は対角要素が正の定数対角行列である．w を無限大に近づけると重み付き最小二乗解 y^* は，

$$\begin{aligned}
y^* =& y_1^* + y_2^*,\\
y_1^* =& (\min\ \|\hat{D}_1(A_1^T y - c_1)\|^2 \text{ の最適解}),\\
y_2^* =& (\min\ \|\hat{D}_2(A_2^T(y_1^* + y_2) - c_2)\|^2,\\
& \quad \text{subject to } A_1^T y_2 = 0 \ \text{ の最適解})
\end{aligned}$$

へと収束していく．この事実は直観的には，まず重みの大きい成分の残差を最小化して，続いて残りの成分を最小化する，という考え方で理解できるであろう．もちろん，この考え方は，2 層だけに限らずより多くの層に一般化することができる．これを，層別最小二乗法と呼ぶ．あとでより詳しく述べるが，層別最小二乗法は Vavasis と Ye によって導入された [25]．層別最小二乗法は，重み付き最小二乗法の極限として得られるので，重み付き最小二乗法と同様の評価

$$\|y^*\| \leq \chi_A \|c\|, \quad \|A^T y^*\| \leq \bar{\chi}_A \|c\|$$

が成立する．また，層別最小二乗解と対応する重み付き最小二乗解の差は，2 層の場合 $O(\chi_A \bar{\chi}_A^4 / w^2)$ で評価でき，層の数がより多い場合にも同様の結果が成立する [7, 21]．既存研究についてはここではこれ以上触れないが，興味のある方は [7, 21] をご覧いただければと思う．

3. 内点法と重み付き最小二乗法: ソビエト連邦における展開

以下，線形計画問題

$$(P)\ \min\ c^T x, \ \text{ subject to } Ax = b, \quad x \geq 0$$

とその双対

$$(D)\ \max\ b^T y, \ \text{ subject to } c - A^T y = s, \quad s \geq 0$$

を考える．ここで $A \in R^{m\times n}$, $x, s, c \in R^n$, $y, b \in R^m$ である．

Karmarkar によって射影変換を用いた多項式時間内点法が 1984 年に発見されたことはよく知られているが，それ以前 1967 年にソ連の Dikin によって，Karmarkar 法を簡単にした一変種であるアフィンスケーリング法がすでに提案されていた．この方法は重み付き最小二乗法と密接な関係がある (残念ながら Dikin 博士は 2008 年に亡くなられた)．

当時社会主義国であったソ連では耕作地の地代を定める，といった問題が，線形計画問題を用いて検討されていた．そして，最適解でない点においても，何らかの形でシャドウプライスを求める必要があった．そのために，シャドウプライスを相補性条件をできるだけ満たすように「重み付き最小二乗法で」推定することが Kantorovich によって提案された [4]．これにさらに改善を施したものが Dikin による推定法で，主問題の実行可能内点解 $x > 0$ において，双対問題の変数を相補性条件ができるだけ満たされるように重み付き最小二乗問題

$$\min\nolimits_{(y,s)}\ \|Xs\|^2, \quad \text{subject to } c - A^T y = s$$

を解くものである．ここで X は x を対角要素とする対角行列であり，解として，

$$\begin{aligned}
y =& A^T(AX^2A^T)^{-1}X^2 c,\\
s =& (I - A^T(AX^2A^T)^{-1}X^2)c
\end{aligned}$$

が得られる [4]．この量を「双対推定」と呼ぶ．ここで好都合なのは，s をスケーリングした量 $-X^2 s$ が，条件 $A(-X^2 s) = 0$, $-c^T X^2 s \leq 0$ を満たしており，主問題の目的関数値をより小さくするために探索方向として用いることができることである．双対推定を計算しながら，この探索方向に進んで目的関数値を改善していく，これがアフィンスケーリング法である [2〜4]．

このアルゴリズムが必ず最適解に到達するかどうか，という問題は 1990 年代初頭は未解決であった．筆者は局所 Karmarkar ポテンシャル関数という関数を導入し，当時大学院生であった村松正和氏とも共同研究を行って，大域的収束性の証明を与えた [18, 23]．解析にあたっては，多面体の境界近くでの探索方向の振舞いを明らかにする必要がある．多面体の境界近くでは X の値は大きくばらつく．そのような状況での探索方向の振舞いを明らかにする重み付き最小二乗法の解析が必要となるが，Sherman–Morrison–Woodbury 公式（逆行列公式）などを用いることで可能となる．

4. 情報幾何

情報幾何 [1] は，統計や学習理論などを微分幾何学的に扱う枠組みの一つである．情報幾何では凸関数を

ポテンシャル関数として導入される双対平坦空間が重要な舞台設定であり，2 種類の測地線が重要な役割を果たす．これがリーマン幾何と大きく異なる点である．

卒業論文の研究室が甘利俊一先生の研究室で，当時は統計幾何と呼ばれていた情報幾何が卒論のテーマであった．伊理正夫先生の研究室で修士を修了したのちに 1986 年に統計数理研究所に入所した．ほどなくして内点法の研究を行うようになり，田邊國士先生の指導を受けて，先生とともに，多面体の内部に対数障壁関数をポテンシャル関数とした双対平坦空間の構造が自然と入り，元の問題での直線を情報幾何の一方の測地線とすると，アフィンスケーリング法の連続版の軌跡がもう片一方の測地線となっていることを指摘する解説記事を書いた [17]．さらに内点法の反復回数と微分幾何学的曲率などの関係が見いだせたら大変面白いであろう，とは思いつつも，その後，その問題に再び取り組むまでには 20 年近くの時間が経過する．主双対内点法も登場したばかりで，今では当たり前となっている，中心曲線などの性質もあまり明らかになっていなかったころの話である．

5. Vavasis–Ye の内点法

1995 年夏，当時，統計数理研究所の同僚であった水野眞治氏（現：東京工業大学）とともに，IBM のアルマーデン研究所に 2 カ月滞在する機会が与えらえた．何を研究するか，議論しているときに，共同研究者の N. Megiddo 博士にこのような論文があるので読んでみないか，と薦められたのが，Vavasis と Ye の「層別最小二乗法による，計算複雑度が行列 A にしか依存しない多項式時間内点法」という論文であった [25]．このアルゴリズムの肝となるのが先に述べた層別最小二乗法のアイディアである．

計算複雑度が係数行列のみにしか依存しない，線形計画問題に対する多項式時間アルゴリズムは，線形計画問題に対する強多項式時間アルゴリズムの存在という有名な未解決問題に直接関連する関連した面白い研究対象である．このようなアルゴリズムはÉ. Tardos によって最初に提案されたが，Vavasis と Ye は，内点法の枠組みで同様のことができることを示したのである．

よく知られているように，主双対内点法では，ν をパラメータとした次の方程式

$$Ax = b, \quad c - A^T y = s, \quad x \circ s = \nu \mathbf{1}$$

の解として定義される中心曲線を辿って最適解に向かう．ここで $\circ$ は要素ごとの積，$\mathbf{1}$ はその要素がすべて 1 のベクトルである．典型的な主・双対内点法は水野・Todd・Ye の予測子・修正子法である（以下 MTY–予測子・修正子法と記す）．予測子は，上記の方程式で $\nu = 0$ の点，すなわち最適解を目指す Newton 方向として導入されることが多いが，先に述べた，双対推定の考え方で導入することもできる．

現在いる内点許容解を (x, s, y) として，重みを $d_i = \sqrt{x_i/s_i}$ として，双対推定を行うことを考える（中心曲線上では，$d_i = x_i/\sqrt{\mu} = \sqrt{\mu}/s_i$ となり，重みとしては，x_i や $1/s_i$ と同じである）．すると，重み付き最小二乗問題

$$\min_{y'} \|D(s - A^T y')\|^2$$

の最適解が現在いる点での双対推定であり，$y - y'$ を探索方向と思うと主双対内点法の予測子としての探索方向が得られる．主問題のほうの推定は，

$$\min_{x'} \ \|D^{-1}x'\|^2, \ \text{subject to } Ax' = b$$

であり，探索方向も同様にして求められる．

さて，Vavasis と Ye はアルゴリズムを加速するために，以下のような変形版を考えた．まず，重み D を小さい順に並べ，隣同士の比が非常に大きくなるところをギャップと呼び，ギャップが大きいところで，重みの比を強制的に無限大にした極限をとる（ギャップのあるところより大きな重みについてはそれらの間の比は変化しないよう，すべて同じ比率で大きくする）．これは先に述べた層別最小二乗法にほかならない．Vavasis と Ye は，層別最小二乗方向を，通常の予測士・修正子法における通常の予測子に代わる，もう一つの予測子として用いて主双対内点法を加速し，b や c に異存しない計算複雑度を達成している．特に，$O(n^{3.5} \log \bar{\chi}_A)$ 解の反復で有限回の反復で最適解に到達して停止する．ここで出てくるのが，先に述べた行列 A の条件数 $\bar{\chi}_A$ である．

アルゴリズムとその解析は難解であるものの，あたかも「中心曲線の挙動から多面体の境界の構造について手さぐりで探っている」ようなスリリングな感じのする，たいへん興味深い論文である．Karmarkar の論文もそうであるが，Vavasis と Ye の論文もたいへん独創的で，どのようにしてこのような結果が得られるのかまったく想像できない．アフィンスケーリング法の解析時に重み付き最小二乗法については解析した経験

もあるので何かできるかもしれない，という思いもあり，一時期かなりの労力を費やしてこの方法の研究に取り組んだ [9〜11].

Vavasis と Ye の内点法の重要なアイディア，何故，層別最小二乗法が必要になるか，について，そして，どのようにして，$\bar{\chi}_A$ が使われるか，ここで説明を試みる．簡単のために，双対問題について議論する．$\mathcal{S} = \{s | s = c - A^T y,\ y \in R^m\}$ とする．$s \in \mathcal{S}$ として，その中で，最適解でアクティブと推定される変数 s_1 と残りの変数 s_2 に現在変数がわかれているとしよう（通常小さい変数をアクティブと推定する）．層別最小二乗法の考え方で，まず s_1 部分について，対応する重み行列 D_1 を用いて重み付き最小二乗法を行い，そして，

$$\varepsilon = \min_y \|D_1(s_1 - A_1^T y)\|$$

とする．もし $\varepsilon > 0$ ならば，$s_1 - A_1^T y = 0$ が解を持たない．このことは，一見，退化した最適解で頂点のように見えるものが，実は，近づいてよく見ると，複雑な多面体的な構造を持つことを意味する（図 1）．また，同時に，スラック s_1 の中に，最適解で 0 でないものが必ず一つは存在することを意味する．ここで，アフィン空間 $\mathcal{S}$ が原点からどの位離れているかを評価したい．定義より，任意の $s' \in \mathcal{S}$ について

$$\|c_1 - A_1^T y^*\| \leq (1 + \bar{\chi}_A)\|s_1\|$$

が成立する．つまり，直観的には，s_1 のノルムを，今いる点のノルムから，

$$\delta = \frac{\|c_1 - A_1^T y^*\|}{1 + \bar{\chi}_A}$$

位まで減らせれば，次の微細構造（図 1 の「頂点」部分）が見えてくる，という仕掛けになっているのである．$\bar{\chi}_A$ が有限でなく ∞ であれば，重み付き最小二乗解を使って δ の評価ができないことに注意しよう．

そしてさらにうまいことに，層別最小二乗法による探索方向を用いることによって，微細構造の近くまで 1 反復で行くことができるのである．これが，通常の主双対内点法ではできないことなのである．それは，層別最小二乗法による探索方向の s_1 成分は s_1 部分のみで決まるのに対し，主双対内点法の探索方向の s_1 成分は，主として s_1 部分で決まるが，s_2 部分の影響も受けるためにその分だけ減速してしまうからである．

このような形で $\bar{\chi}_A$ が内点法の性能評価に使われるのは興味深い．Vavasis–Ye のアルゴリズムに関するより詳しい日本語の解説として [22] を挙げておく．

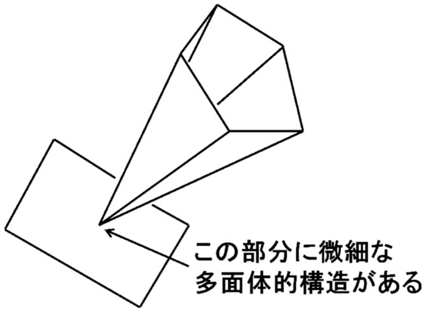


図 1　悪条件問題

さて，ここで，一つ筆者が興味を持っている問題を述べよう．MTY–予測子・修正子法に有限回で最適解を発見して停止するための手続きを組み込むと，ごく自然な形で，問題のスケーリングに対する不変性を持った，2 層の層別最小二乗法が現れる．この 2 層の層別最小二乗法による探索方向を普通の予測子と併せて用いるようにする，MTY–予測子・修正子法の改訂版を考えることができる．このアルゴリズムの計算複雑度は Vavasis–Ye のアルゴリズムと同様 A にしか依存しないのではないかと考えられる．現在のところ「もし (D) の実行可能領域が有界ならば，チューリングマシンモデルの下では，このアルゴリズムの反復計算複雑度は $O(n^{3.5}L_A + n^2 L_{A,c})$ となり，b のビットサイズには依存しない」ということがわかっている．ここで $L_{A,c}$ は A と c のビットサイズである．この結果は北原知就氏との共同研究を通じて得られたものである [8]. この改訂 MTY–予測子・修正子法はきわめて自然なアルゴリズムなので，その計算複雑度が，完全に A のみに依存することを証明することは面白く挑戦的な研究課題の一つだと思う．より詳しくは論文 [8] を参照されたい．

6. リニアモーターカーと層別最小二乗法

ローザンヌで 1997 年に開催された ISMP で Alizadeh の 2 次錐計画法の講演を聞いたことがきっかけとなり，2 次錐計画法に対する主双対内点法の一般化ができるといいと思い Faraut と Koranyi の本などを見ながら，多項式時間主双対パス追跡法を 2 次錐計画問題に拡張した [19].

ここで，統計数理研究所で毎年開催されている研究集会「最適化：モデリングとアルゴリズム」において，伊理研究室の後輩である鉄道総合技術研究所の笹川卓氏が，リニアモーターカーの磁気シールドの最適設計問題について講演されていたことを思い出した．

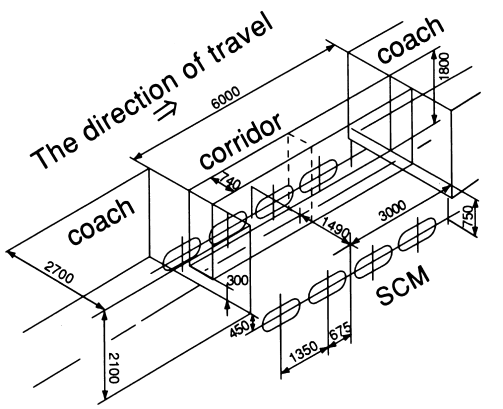


図 2　リニアモーターカーと超電導磁石 (SCM) の配置

リニアモーターカーは磁気で浮上し走行するために，走行時には車体に積んだ超伝導磁石から強い磁場が発生する．車内の乗客を磁場から守るためには，車内をシールドする必要がある．シールドするためには単に鉄板で囲えばよいのであるが，浮上走行することを考えると，鉄板の厚みはできるだけ軽くしたい（図 2）．笹川氏は巧妙なモデル化により，この問題を，シールド内に通る磁場のベクトル場のユークリッドノルムを車体上で積分したものを最小化することに帰着した．磁場のベクトル場には満たすべき線形制約があり，以下のような最適化問題として定式化できる．

$$\min \int \|\vec{F}\| dS, \quad \text{subject to } \mathrm{div}\vec{F} = B_n$$

B_n は車体に取り付けらえた超電導磁石が作る外部磁場で，超電導磁石の配置と車体の形状から決まるものである（図 2）．そして，$\vec{F}$ がシールド（鉄板）内を流れる磁束で，流量の保存則にあたるものが，線形制約として課される．各点での磁束の流れ $\vec{F}$ を実現するためには，物理的考察から，$\|\vec{F}\|$ に比例するだけの厚みが必要となる．厚みを積分すると体積となり，それが目的関数である．彼は，この問題を有限要素法で離散化し，発見的解法を考案してスパコンでこの問題を解き，近似解を求めていた [14]．

この問題は 2 次錐計画問題に帰着するということで，笹川氏と共同研究を始めた．そして，開発したばかりの主双対内点法を実装して，高速に正確にパソコンで解くことができた [14〜16, 24]．おそらくは，当時としては大規模な数千変数レベルの 2 次錐計画問題を解いた世界最初の例ではないかと思われる．

さて，電磁気では透磁率という概念がある．透磁率は，磁束密度と磁場の強さとの比である．金属ではその値は大きく，笹川モデルでは，その値 μ を ∞ として最適化問題を導出している [15]．笹川氏にとってはごく自然な近似であるこのモデルが，筆者にはどうにも理解できない．ところが，あるときに，二人で議論しているうちに，このモデルを理解するのに，層別最小二乗法が使えるのではないか，という着想が得られた．笹川モデルの線形制約は，外部磁場の総エネルギーと，シールド内部の磁場の総エネルギーと，車内空間の磁場の総エネルギーを最小化する，という変分原理で導出できる．変分原理をとるべき総エネルギーを求めると，

$$\begin{aligned}&(\text{変分をとるべき総エネルギー})\\ &\quad = (\text{外部磁場の総エネルギーに}\\ &\qquad\quad \text{対応する凸 2 次関数})\\ &\qquad + \frac{1}{\mu}(\text{シールド内磁場の総エネルギーに}\\ &\qquad\quad \text{対応する凸 2 次関数})\\ &\qquad + \frac{1}{\mu^2}(\text{車内空間磁場の総エネルギーに}\\ &\qquad\quad \text{対応する凸 2 次関数})\end{aligned}$$

という構造をしている．右辺の各項は，エネルギーで 2 次関数であるから，まさに，重み付き最小二乗法である．透磁率を無限大にする極限をとることで層別最小二乗法としての線形制約が現れ，それが，笹川モデルの線形制約となるのである！[14, 24] 私は，層別最小二乗法がこんな意外なところに現れることに驚いた．このことは，層別最小二乗法のモデリングにおける有用性を示唆するものであろう．

余談になるが，2 次錐計画問題に対する主双対内点法は，宇宙のダークマターの分布を推定するための統計的手法に活用されている [26]．一般相対論や宇宙論に大学入学時に憧れていた自分としては，意外なところで引用されてちょっとうれしく感じたものである．

7. 中心曲線の構造と内点法の反復回数

さて，内点法と情報幾何の関係の話に戻ろう．ここに最小二乗問題の条件数 $\bar{\chi}_A$ が再び登場する．Vavasis と Ye は，彼らの論文で中心曲線は高々$O(n^2)$ 個の「長い直線部分」と「曲線部分」に分解できると述べている．そして，まさにこの「長い直線部分」こそが，彼らの層別最小二乗方向を用いて加速できるところなのだと．これは幾何学的な言明として魅力的なものであるが，Vavasis と Ye のアルゴリズムは問題のスケーリング（すなわち問題をどのような単位系で記述するか）に対して不変ではない．その意味では，幾何学的な言明にはなっていない．この結果はきちんと「幾何的な

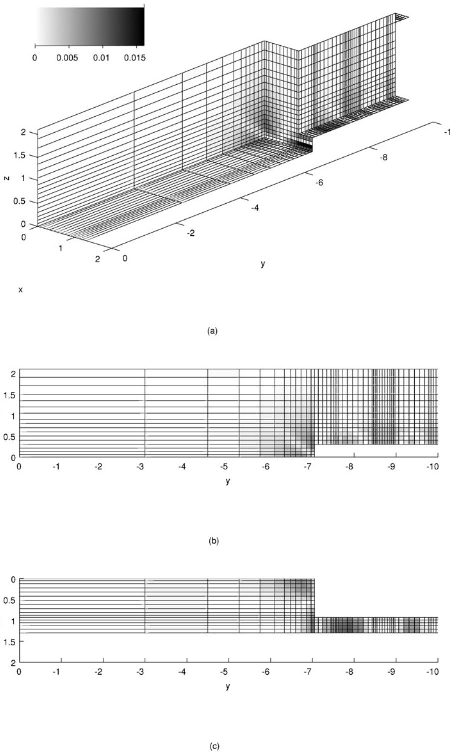


図 3　最適磁気シールドの厚み（単位：m）

言明」として述べることができるのであろうか？美しい結果なので，幾何学的不変性とどこかでつながっているはずである．

幾何学と結びつけるには，その前段階として，まず，内点法の反復回数を表現する中心曲線上の積分を求める必要がある．実は，広さ β の中心曲線近傍を用いてパラメータ ν_1 の中心曲線上の点から ν_2 まで辿るときの MTY–予測子・修正子アルゴリズムの反復回数を $\#(\nu_1, \nu_2, \beta)$ とすると，

$$\#(\nu_1,\nu_2,\beta)\sim\frac{I_{PD}(\nu_1,\nu_2)}{\sqrt{\beta}},$$

$$I_{PD}(\nu_1,\nu_2)=\int_{\nu_2}^{\nu_1}\sqrt{\frac{\|\dot{x}\circ\dot{s}\|}{\nu}}d\nu$$

が成立する．ここで $\sim$ は近似的に成立という意味である．この積分については，2003 年ミネアポリスの IMA に行ったときに考えついたことを覚えているが，後に，Sonnevend，Stoer と Zhao によって 1991 年にすでに提案されていたことが判明した．少々がっかりしたものの，Vavasis と Ye のアルゴリズムについての共同研究を行っていたジョージア工科大の Monteiro 氏と一緒に，Vavasis と Ye アルゴリズムの計算複雑度と中心曲線の積分を結び付ける研究を行い，$I_{PD}(0,\infty)$ が広義積分の意味で存在し，これが，最小二乗法の条件数 $\bar{\chi}_A$ を用いて，

$$I_{PD}(0,\infty)=O(n^{3.5}\log\bar{\chi}_A)$$

と有限値で抑えられることを証明した [12, 22]．関連する研究発表を Oberwolfach の数学研究所で行ったのも懐かしい思い出である．

8.　凸最適化の情報幾何へ

さて，先に述べた線形計画問題に対する情報幾何の枠組みは，のちに小原敦美氏によって，半正定値計画問題に拡張された．そのような経緯もあり，2000 年前後から，小原氏と凸最適化の情報幾何と内点法についての共同研究がスタートし，その後柿原聡氏の協力も得て，一定の成果を得ることができた [5, 6, 13].

線形計画問題の場合，情報幾何構造は以下のようになる．(P) の空間の第一象限の内部を Ω と記すことにして，Ω に対する対数障壁関数 $\psi(x)=-\sum_{i=1}^{n}\log x_i$ を考える．これが情報幾何の双対平坦空間を定めるポテンシャルとなる．

ルジャンドル変換 $s(x)=-\partial\psi(x)/\partial x$ を導入すると，$s(x)$ は Ω における非線形大域座標系となり，$s(\Omega)=\Omega$ となる．x を ∇ 接続に対するアファイン座標系，$s(x)$ を ∇^* 接続に対するアファイン座標系として，双対平坦空間の構造が導入される（ここで ∇，∇^* は勾配を意味する数学記号と異なる情報幾何の用語である）．この空間のリーマン計量は $\partial^2\psi/\partial x^2$ で与えられる．主問題の実行可能領域の内部 $\mathcal{P}$ は Ω の部分多様体だが，双対問題の実行可能領域の内部 $\mathcal{D}$ も $s(x)$ を用いると Ω の部分多様体と考えることができる．x 座標系では $\mathcal{P}$ は平坦だが $\mathcal{D}$ は曲がっており，s 座標系では $\mathcal{D}$ は平坦だが $\mathcal{P}$ は曲がっている（図 4）．この立場では，中心曲線のパラメトライゼーションとして $t=1/\nu$ を導入して，中心曲線を

$$Ax=b,\quad s=c-A^Ty,\quad x\circ s=\frac{1}{t}\mathbf{1}$$

と表すのが自然である．このとき，主中心曲線を $(x(t),s(t),y(t))$ と記すと，詳細は省くが，$h_P(t)$ を主実行可能領域への主中心曲線 $x(t)$ の ∇^* 接続に関する埋め込み曲率の（リーマン計量の意味での）ノルム，$h_D(t)$ を，双対実行可能領域への双対中心曲線の ∇ 接続の埋め込み曲率のノルムとすると，ある種の主内点法の反復回数や双対内点法の反復回数が，用いる近傍の大きさを β として，中心曲線上の積分

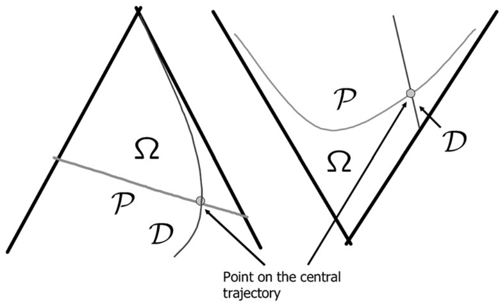


図 4　最適化の情報幾何の概念図（両者は同じ問題を x と s による座標系で表現したものである）

$$I_P(\nu_1,\nu_2)=\frac{1}{\sqrt{2}}\int_{1/\nu_1}^{1/\nu_2}\sqrt{h_P(t)}dt,$$
$$I_D(\nu_1,\nu_2)=\frac{1}{\sqrt{2}}\int_{1/\nu_1}^{1/\nu_2}\sqrt{h_P(t)}dt,$$

を用いて，$I_P/\sqrt{\beta}$, $I_D/\sqrt{\beta}$ と書けることを見いだした．これは，主中心曲線，双対中心曲線の幾何学的性質を表す量である．一方，主双対内点法の反復回数を表す積分を ν から t にパラメタを変換して，

$$\int_{1/\nu_1}^{1/\nu_2}\sqrt{h_{PD}}(t)dt$$

と表す．すると，下記のような「ピタゴラスの定理（関係）」が成立するのである！

$$h_{PD}^2 = h_P^2 + h_D^2 \qquad (1)$$

この事実は，主双対内点法の反復回数を表す積分が，情報幾何的な量として書けることを意味する．これより，内点法の反復回数と密接に結び付いた，中心曲線の情報幾何的積分 $I_P(0,\infty)$, $I_D(0,\infty)$, $I_{PD}(0,\infty)$ について，

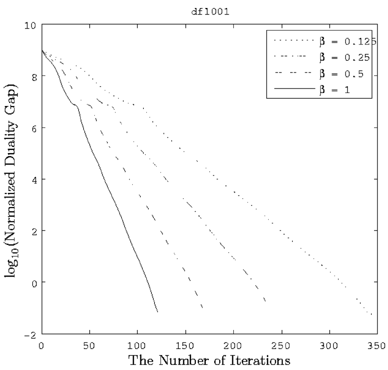
$$\max[I_P(0,\infty), I_D(0,\infty), I_{PD}(0,\infty)]$$
$$= O(n^{3.5}\log\bar{\chi}_A)$$

という評価が成立し，特に，ネットワークフロー問題については，$\log\bar{\chi}_A$ が $L_A = mn$ で抑えられることより，

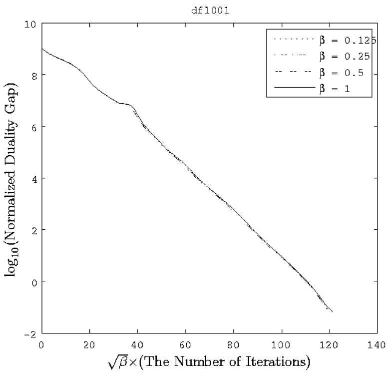
$$\max[I_P(0,\infty), I_D(0,\infty), I_{PD}(0,\infty)] = O(n^{4.5}m)$$

となる．これは，中心曲線の大域的幾何学的性質ともいえる評価である．ここまで来て，最小二乗問題の条件数と，情報幾何が結び付くのである．

線形計画問題の場合に，情報幾何による主内点法／双対内点法の反復回数を表現する積分がどうなるかを計算し，それと主双対内点法の反復回数を表現する積分とどういう関係があるかを計算し，ピタゴラスの定理が出てきたときには，こんなきれいな関係が成立するのか，と本当に驚き，感動したものである（大体こういう評価の計算はうまくいかないことがほとんどである）．これは，研究がある意味では，鋭い思いつきのみではない，ということを表していると思う．結果が出るまではわからないものの，適切な問題設定をして，愚直に計算すれば，面白い結果が得られることもある，ということを物語っている．



(a) 反復回数と正規化双対ギャップ ν



(b) (反復回数 /$\sqrt{\beta}$) と正規化双対ギャップ ν

図 5　DFL001 に対する主双対内点法の挙動

この積分が反復回数を非常によく近似していることを実用レベルの問題で示そう．DFL001 は有名な Netlib 問題である．制約式の数が 6072，変数の数が 12230 である．図 5(a) をご覧いただきたい．β を変えると，反復回数と ν の値の関係は，似た形をとり，横軸を適当にスケーリングすれば，重なりそうな感じがある．関係 (1) からは，反復回数を $\sqrt{\beta}$ で割ったものは，積分値そのものになるので，重なることが期待される．実際，図 5(b) に反復回数を $\sqrt{\beta}$ で割ったものをプロットすると重なることがわかる．積分そのものは $\beta = 1$ としたときの反復回数であり実験的にも「内点法の反復回数は情報幾何的積分量である」ことが示されている

といってよいであろう．この結果は半正定値計画問題や対称錐計画問題に拡張されている．

冒頭で紹介した最小二乗法の条件数が，このように離れた，一見無関係なところに表れることは，不思議である．研究を進めていくにつれてこういうところにつながっていったと考えるとその時間距離は 10 年とか 20 年ということになろうか．

9. これからどこへ？

最小二乗法と線形計画問題について，少し変わった視点から，自分の研究の跡をたどる形で好きなように述べてきた．「最小二乗法」から始まった話が「ピタゴラスの定理 (1)」にたどりついた！というのがこの話のオチである．線形計画問題にまつわる話題は尽きない．半正定値計画問題や対称錐計画問題の研究が進展した後，圧縮センシングやスパースモデリングへの線形計画の活用という，フィールズメダリストも加わっての驚くべき展開が待ち受けていた．機械学習や統計科学の研究者が最適化にかつてない高い関心を持ち，賑やかな学際的な展開が続いている．そのなかで，層別最小二乗法や χ_A や $\bar{\chi}_A$ などがモデリングに活用されるときがくるかもしれない．思わぬところでいろんなことがつながり，研究は面白いものだと思う．まだまだいろんな出会いがあることを期待して筆をおきたい．

参考文献

[1] 甘利俊一，長岡浩司：情報幾何の方法．岩波書店，1993.

[2] I. Dikin, Iterative solution of linear and quadratic programming problems, *Soviet Mathematics Doklady*, **8**, 674–675, 1967.

[3] I. Dikin, On the convergence of an iterative process (in Russian), *Upravlyaemye Sistemi*, **12**, 54–60, 1974.

[4] I. Dikin and V. Zorkaltsev, *Iterative Solution of Mathematical Programming Problems* (in Russian), Academia Nauk, USSR, 1980. （およびその英訳原稿）.

[5] S. Kakihara, A. Ohara and T. Tsuchiya, Information geometry and interior-Point algorithms in semidefinite programs and symmetric cone programs, *Journal of Optimization Theory and Applications*, **157**, 749–780, 2013.

[6] S. Kakihara, A. Ohara and T. Tsuchiya, Curvature integrals and iteration complexities in SDP and symmetric cone programs, *Computational Optimization and Applications*, published online, DOI 10.1007/s10589-013-9608-x, 2013.

[7] T. Kitahara and T. Tsuchiya, Proximity of weighted and layered least squares solutions, *SIAM Journal on Matrix Analysis and Applications*, **31**, 1172–1186, 2009.

[8] T. Kitahara and T. Tsuchiya, A simple variant of the Mizuno–Todd–Ye predictor-corrector algorithm and its objective-function-free complexity, *SIAM Journal on Optimization*, **23**, 1890–1903, 2013.

[9] N. Megiddo, S. Mizuno and T. Tsuchiya, A modified layered-step interior-point algorithm for linear programming, *Mathematical Programming*, **82**, 339–355, 1998.

[10] R. D. C. Monteiro and T. Tsuchiya, A variant of the Vavasis–Ye layered-step interior-point algorithm for linear programming. *SIAM Journal on Optimization*, **13**, 1054–1079, 2003.

[11] R. D. C. Monteiro and T. Tsuchiya, A new iteration-complexity bound for the MTY predictor-corrector algorithm. *SIAM Journal on Optimization*, **15**, 319–347, 2004.

[12] R. D. C. Monteiro and T. Tsuchiya, A strong bound on the integral of the central path curvature and its relationship with the iteration complexity of primal-dual path-following LP algorithms, *Mathematical Programming*, **115**, 105–149.

[13] A. Ohara and T. Tsuchiya, An information geometric approach to polynomial-time interior-point algorithms: complexity bound via curvature integral, *Research Memorandum* No. 1055, The Institute of Statistical Mathematics, December 2007.

[14] 笹川卓，2 次錐計画問題を用いた直流磁気シールドの最適化．博士論文，東京大学大学院情報理工学研究科，2005.

[15] 笹川卓，モデリング：直流磁気シールドの開発を例に取って．オペレーションズ・リサーチ，**52**, 216–220, 2007.

[16] T. Sasakawa and T. Tsuchiya, Optimal magnetic shield design with second-order cone programming, *SIAM Journal on Scientific Computing*, **24**(6), 1930–1950, 2003.

[17] 田辺国士, 土谷隆，線形計画法の新しい幾何学．数理科学, **303**, 32–37, 1988.

[18] T. Tsuchiya, Global convergence of the affine scaling methods for degenerate linear programming problems, *Mathematical Programming Series B*, **52**, 377–404, 1991.

[19] T. Tsuchiya, A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming, *Optimization Methods and Software*, **11/12**, 141–182, 1999.

[20] 土谷隆，モデリング雑感，オペレーションズ・リサーチ，**50**，560–563, 2005.

[21] 土谷隆，層別最小二乗法 ―重み付き最小二乗法の極限―，統計数理，**61**，391–404, 2005.

[22] 土谷隆，層別最小二乗法と交差を用いた線形計画問題に対する内点法の解析と中心曲線の幾何学的性質．日本 OR 学会第 17 回 RAMP シンポジウム予稿集，pp. 197–211, 2005.

[23] T. Tsuchiya and M. Muramatsu, Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems, *SIAM Journal on Optimization*, **5**, 525–551, 1995.

[24] 土谷隆，笹川卓，2 次錐計画問題による磁気シールドのロバスト最適化，統計数理，**53**, 297–315, 2005.

[25] S. Vavasis and Y. Ye, A primal-dual accelerated interior point method whose running time depends only on A, *Mathematical Programming*, **74**, 79–120, 1996.

[26] X. Wang, M. Walker, J. Pal, M. Woodloofe and M. Mateo, Model-independent estimates of dark matter distributions, *Journal of American Statistical Association*, **103**, 1070–1084, 2008.