

統計数理研究所における最適化研究

土谷 隆^{1,2}

(受付 2019 年 5 月 7 日；改訂 9 月 10 日；採択 9 月 20 日)

要 旨

最適化は統計科学における重要な方法論である。最適化アルゴリズムの発展は新しいモデルの実用化を可能とし、新しいモデルの追及は、最適化アルゴリズムの進展をもたらす。統計科学の研究所としての統計数理研究所では、さまざまな形で最適化の研究が行われてきた。本論説では、統計数理研究所における最適化法の研究の展開について概説する。

キーワード：無限次元最適化、最急降下法、Kaczmarz 法、非線形最適化、内点法、情報幾何。

1. はじめに

最適化は統計科学における重要な方法論である。統計科学は、データの背後にそれを生み出す確率的機構、すなわち統計モデルを想定することで、現象の予測や説明を目指す。その過程では、最小二乗法や最尤法に見られるように、適切な定量的規準を最適化してモデルのパラメータを推定することが行われる。最適化手法が深化し発展することで、より多くの統計モデルが実用化される。遡ってみれば、線形計画法の原型は、 L_1 回帰として現れたともいわれている。また、ネイマン・ピアソンの基本定理にも見られるように、推定量や検定量の最適性についての解析が、古くからの統計学と最適化の接点としてあった。そこでは、より抽象的な、無限次元の最適化問題が扱われ、最適性条件等が問題となることが多かった。このように、最適化法は、さまざまな形で統計科学に古くからかわってきた。

第 2 次大戦後の電子計算機の登場は、現代科学技術の性格を大きく変えることとなった。実際に計算を行い現実問題を解くことができる可能性が拓けたことを契機とし、多くの数学的・数理的分野が「モデリングとアルゴリズム」を意識した新しい方向に力強く発展していった。統計学についても然り、また最適化についても然りである。特に最適化においては、1947 年、Dantzig による線形計画法の実用化とともに、数理計画法という新しい学問が勃興し、最適化手法の開発とそれを意識したモデリングと数理に重点を置いた分野として研究が活発に進められるようになってきた。

数理計画の分野では、連続最適化と離散最適化の 2 つの分野が互いに影響を及ぼしあいながら研究が展開された。1950 年代から 70 年代にかけては、連続最適化では、数理的には、双対理論の構築や最適性条件の解析などの研究が進展し、アルゴリズム研究においては、無制約最適化問題に対する最急降下法、ニュートン法の解析や、準ニュートン法などの開発、制約付最適化問題については、罰金関数法や障壁関数法、乗数法などの開発が行われた。離散的最適化

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

² 統計数理研究所 客員：〒190-8562 東京都立川市緑町 10-3

では、ネットワーク最適化問題の解法や整数計画問題に対する分枝限定法などが開発されてきた。また、モデリングの分野では、線形計画問題がさまざまな分野に適用され、問題解決のための有力な道具として活用されるようになった。

1970 年代の後半に計算複雑度の理論が勃興し、最適化問題を解くための計算量が離散的最適化、連続的最適化の両分野で重要な研究課題となった。そして、線形計画問題に対する最初の多項式時間解法である楕円体法が Khachiyan によって 1979 年により提案され、さらに、実用的多項式時間解法である Karmarkar 法の 1984 年の登場により、そして分野全体を巻き込んだ大きな変革をもたらした内点法の研究へと進んでいった。

計算機の急速な発展と相俟って、凸 2 次計画問題や半正定値計画問題、2 次錐計画問題などが新たな最適化問題として解けるようになった。これは、線形計画法が数理モデリングの世界で果たしてきた役割をさらに広げ、最適化モデリングの可能性を大きく拡大するものであった。多くの実際問題が「最適化問題を解くことに帰着される」のではなく「線形計画問題、凸 2 次計画問題、半正定値計画問題や 2 次錐計画問題という『最適化モデル』に帰着することにより」解かれるようになったのである。凸 2 次計画法が実用化されたのちに、サポートベクターマシンが現れ、また、圧縮センシングが線形計画問題の思いがけない応用として登場し、半正定値計画法は制御理論の世界を大きく変え、さらに、統計学、信号処理や種々の分野で統計的モデリングに活用されている。

そして、2000 年代に入り、ビッグデータの時代の到来とともに、機械学習における基本的道具立てとしての連続的最適化が改めて注目され、超大規模問題に対する最適化手法が精力的に研究され、また、離散的最適化においては、数千から数万変数の整数計画法の問題が日常的に解かれるようになり、様々な分野で活用されている。

このような数理計画分野の研究の歴史の中で、統計数理研究所においては、1955 年の改組において第 3 研究部が「OR および統計解析理論を研究する部署」として位置づけられ(統計数理研究所, 1994)、最適化とモデリングの学問としての数理計画法の研究が統計数理研究の一環として行われ、国際的にも通用する重要な研究成果を発信してきた。これらの成果の中から、本論説では、石井による、無限次元線形計画問題とチェビシェフ限界に関する研究、赤池による最急降下法の解析、田邊による Kaczmarz 法の解析、非線形最適化に関する微分幾何学的接近法、主双対内点法、そして、筆者を含む研究所の最適化研究者が国内外の共同研究者とともに行ってきた、内点法や最適化の情報幾何についての研究成果を中心として紹介する。なお、本論に登場する研究者の中で、統計数理研究所員については、以下で初めて登場した時に在職時期を付記している。

本論説には、「最適化」という言葉と「数理計画」という言葉が現れる。本論に進む前に、二つの言葉の意味するところの微妙な違いについて触れておきたい。最適化は数理の世界で古くからある言葉であるが、数理計画は Dantzig による線形計画法の登場後、使われるようになった言葉である。数理計画においては数理やアルゴリズムのみならず、モデリングが重要な研究対象となる。実際、数理計画の草創期には、線形計画問題が解けることを前提として、いかにして現実の問題を線形計画問題に落とし込むかというモデリングの手法が一つの重要な研究課題であった。より俯瞰的に見れば、豊かな構造を持つ基本的モデルがあり、実際の問題をそのモデルに帰着できる形にモデリングすることは、戦後勃興した数理科学諸分野において共通に見られた研究の流れであった。このように、数理計画法においては「最適化の数理やアルゴリズムを意識した」モデリングの研究が不可欠なものであり、数理計画法は最適化に関する「モデリング・数理・アルゴリズム」の 3 本の柱からなる学問分野である、と考えられる。このことは、Dantzig の古典的名著や Nesterov による定評ある教科書の序文でも強調されており、また、統計数理研究所で長年開催されてきた最適化の研究集会「最適化：モデリングとアルゴリズム」の

名称にはそのような気持ちが込められている。また、近年盛んに研究されている、機械学習分野における「最適化研究」も、これら3つの分野に跨っていることが見て取れる。

2. チェビシェフ型のバウンドと無限次元線形計画法

モーメントが与えられた時にそのモーメントを実現するような確率分布が存在するかどうかはモーメント問題と呼ばれ、古典的かつ重要な問題である。石井恵一(1955年-1964年在職)は1950年代後半から1960年代にかけて、この問題に連なる一連の数学的問題に取り組んだ(Isii, 1960, 1962, 1964)。その典型的な成果を以下に紹介する(Isii, 1962)。

Ω を全体集合、 $\mathcal{U}$ をその上の σ 加法族とし、可測空間 $(\Omega, \mathcal{U})$ 上での非負測度の族 $\mathcal{B}$ が与えられているものとする。 $f_0(\omega), \dots, f_m(\omega)$ を Ω から R^{m+1} へのボレル可測写像で $\mathcal{B}$ の各要素について積分可能とする。また、 $g(\omega)$ を $\mathcal{B}$ の各要素について積分可能な実関数とする。

この設定の下で、次の条件を満たす $(\Omega, \mathcal{U})$ 上の非負測度 $P \in \mathcal{B}$ を考える。即ち $M_0, \dots, M_m$ を与えられた実数として、

$$(2.1) \quad E_P(f_i) \equiv \int f_i(\omega) dP(\omega) = M_i,$$

がすべての $i = 0, \dots, m$ について成立すると仮定する。この条件を満たす P の内、積分

$$\int g(\omega) dP(\omega)$$

の値をできるだけ大きくすることを考える。すなわち、

$$\mathcal{B}(M) = \{P | P \in \mathcal{B}, E_P(f_i) = M_i, \quad i = 0, \dots, m\}$$

として、次の最適化問題を考える：

$$(2.2) \quad \sup \int g(\omega) dP(\omega), \quad \text{subject to } P \in \mathcal{B}(M).$$

この問題(2.2)は、無限次元線形計画問題と捉えることができる。例えば、確率変数の平均や分散が与えられた時に、ある特定の領域にどの位まで大きな確率を割り当てることが可能であるか？といった、確率分布についての minimax 型の限界を与えるような問題(チェビシェフ型不等式)が(2.2)に帰着できる。

石井の得た主結果は下記の通りである。

定理 1. もし、 $\mathcal{B}$ が凸ならば、集合

$$\mathcal{M} = \left\{ \int f dP(\omega) \mid P \in \mathcal{B} \right\} \quad (f = (f_0, \dots, f_m)^T)$$

は R^{m+1} 上の凸集合である。もし、 $(M_0, \dots, M_m)$ が $\mathcal{M}$ の内点であれば、(2.2)の最適値は、

$$(2.3) \quad \inf_{a_0, \dots, a_m} \left\{ \sum_{i=0}^m a_i M_i + \sup_{P \in \mathcal{B}} \int \left(g - \sum_i a_i f_i \right) dP \right\}$$

と書くことができる。さらに、もし、最適値が有限であれば、実際に最適値を達成する $a_0, \dots, a_m$ が存在する。(2.2)には最適解(最適な測度 P)が存在するとは限らないが、もし、それが存在するならば、(2.3)を達成するものである。

石井は Isii (1964) でこの結果を拡張し、さらに、目的関数を非線形化することで、Cramér-Rao

の不等式を導出している。石井の成果はモーメント問題の古典的な結果として (Lasserre, 2010; Wolkowicz et al., 2000) 等良く知られている専門書に引用されている。また、最近でも, Minimax Probability Machine (Lanckriet et al., 2002) の構築など機械学習の文脈でも活用されており, 息の長い基礎的な成果である。

3. 最急降下法について

最急降下法は, 微分可能な関数 $f(x)$ を最小化するために,

$$x^+ = x - t\nabla f(x)$$

という反復を繰り返す基本的最適化手法である。この手法について, 赤池弘次 (1952 年-1994 年在職) は, Akaike (1959) において, 現在では最適化分野では古典的な結果として広く知られ, 標準的な教科書 (Fletcher, 2000; Noceda and Wright, 2000) 等にも掲載されている, 以下の結果を示した。

「目的関数 f が (そのヘッセ行列の固有値がすべて異なるような) 狭義凸 2 次関数の場合, 正確な直線探索を行うと, 最急降下法によって生成される反復列 x^k は, 漸近的に, f のヘッセ行列の最大固有値と最小固有値に対応する 2 つの固有ベクトルからなる 2 次元空間の方向から最適解に近づく。」

より正確には, 赤池は, 正則な係数行列を持つ連立一次方程式 $Ax = b$ を解く問題を, 正定値対称行列 P を用いて,

$$f(x) = (Ax - b)^T P (Ax - b)$$

の最小化問題と捉え, この問題に最急降下法を適用した時の反復過程を解析している。この時, 正確な直線探索を与える t は, $t = 2(Ax - b)^T Pg / g^T Pg$, $g = A^T P (Ax - b)$ と書ける。

赤池は, ヘッセ行列の固有ベクトルと関連づけて適切な座標を取り直して最急降下法の反復を解析した。新たな座標系で見た誤差ベクトル $(\alpha_1, \dots, \alpha_n)$ をもとに生成される, $(\alpha_1^2, \dots, \alpha_n^2) / \sum \alpha_i^2$ の各要素を生起確率とする多項分布を考え, 各反復でこの多項分布のパラメータが受ける変化を解析することで, 上記の結果を得ている。

赤池は, のちに, 本研究を回想する中で「この問題の取り扱いを通じて, 筆者は数値計算法全般に対する興味と, 関数の局地近傍の形状の二次曲面による近似の実用性の感覚を得た。これが後に筆者を統計モデルの評価に関する情報量規準の導入に誘うことになる。」(赤池, 1999) と記している。一見関係なさそうに見える, 最急降下法と AIC を関連づける形で赤池の研究が進められているのは印象的である。また, この論文の参考文献を見ると, Chernoff や Forsythe, Kantorovich ら統計学, 数値解析, 最適化の一流研究者の論文 (のみ) が参照されており, 統計学・最適化・数値解析各分野の研究者が互いの問題に大きな関心を寄せていた, 数理科学が細分化される前の学際的な雰囲気が感じとれる。また, 数値実験が富士通 FACOM-128 というリレー計算機で行われていたことも記されており, その点でも当時の雰囲気を感じさせるものとして興味深い。

4. Kaczmarz 法の収束性

最適化と数値解析の境界領域となるが, 田邊國士 (1967-2005 年在職) は, 論文 (Tanabe, 1971) において, 連立一次方程式に対する Kaczmarz 法の解析を行った。Kaczmarz 法は, 連立一次方程式に対する緩和法の一種であり, 元来正則な連立一次方程式系のための解法であったが, 田邊は変数と方程式の数が異なるような場合まで含めた形で拡張して解析し, この方法に関す

る古典的結果を導びいた。田邊は複素数の場合について解析を行っているが、ここでは簡単のため、以下、係数行列や右辺は実数とする。

A を $m \times n$ 行列、 b を m 次元ベクトルとして、 n 変数の m 次元連立一次方程式

$$Ax = b$$

を考える。連立一次方程式 $Ax = b$ の解法として、次の反復を考えるのが Kaczmarz 法である。

まず、内部反復として、各方程式が定義する超平面への射影を逐次行う。 x^k が与えられたとしよう。内部反復は下記の通りとなる。

$$(1) x := x^k.$$

$$(2) i = 1, \dots, m \text{ について,}$$

$$x^+ := \operatorname{argmin}_{a_i^T y - b = 0} \|y - x\|_2; x := x^+;$$

内部反復が終了したあとの x^+ を x^{k+1} として、同様にして反復を続ける。この時、適当な行列 Q, R を用いて、

$$x^{k+1} = Qx^k + Rb$$

と書ける。

Kaczmarz 法は、元々は、行列 A が正方で正則な場合の連立一次方程式の解法として提案され、解析されたが、田邊は、この方法の収束性を変数と方程式の次元が異なる場合を含め、ごく一般的な場合に解析し、以下の結果を得た。

「初期点 x^0 から出発した Kaczmarz 法の反復列は、次の点に収束する：

$$x^\infty = P_K x^0 + Gb.$$

ここで、 $G = (I - Q)^{-1}R$ である。 G は行列 A のムーア・ペンローズ一般化逆行列となっている。 P_K は、 A の核 $K = \{u | Au = 0\}$ への直交射影である。」

この結果は、方程式 $Ax = b$ が解を持つ場合には、任意の x^0 について x^∞ は解となり、さらに、 $x^0 = 0$ ととると、方程式が解を持たなくても、点列が最小二乗解に収束することを示している。

Kaczmarz 法は緩和法と呼ばれる方法の一種であり、この考え方は、凸集合の共通部分を求めるなどの手法に一般化されている。田邊の解析した方法は、コンピュータトモグラフィの画像再構成法にも活用され、現在では、機械学習の文脈で再び脚光を集めている。コンピュータトモグラフィがノーベル賞を受賞した時期には、多くの別刷り送付の依頼が寄せられた、とのことである。

5. 非線形計画問題に対する微分幾何的接近法

1960 年代、連続的最適化の分野で、制約付き最適化問題に対する典型的接近法は、罰金関数法や障壁関数法と呼ばれるものであった。これは、目的関数に対して、罰金関数や障壁関数と呼ばれる、制約条件が満たされない点においては無限大かあるいは極端に大きな値をとるような補助関数を足したものを考え、この関数、つまり目的関数 + 補助関数、を無制約最適化することで、制約付き最適化問題を解くものである。罰金関数や障壁関数は、制約条件を破ることに對して罰金を科す補助の関数で、この関数を導入することで、制約付き最適化問題を無制約最適化問題に直すことができる。現在でもよく使われる方法であるが、最適解の近くで数値的

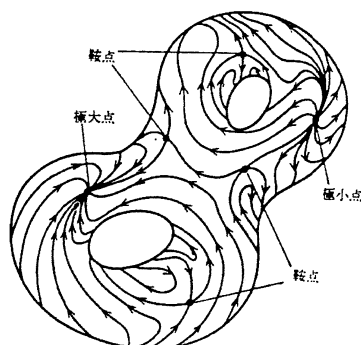


図 1. 射影勾配流の概念図(田邊, 1976, 図 4 を引用).

に悪条件となることが欠点とされていた. そのために, 1970 年代に入り, 拡大ラグランジュ関数などを用いて悪条件性を緩和した, 乗数法などの解法が提案され, 研究が進められた.

田邊は, 同時期に, 非線形計画問題に対し, 微分幾何学的(力学系的)アプローチを試みていた. 等号制約付き最適化問題

$$\max f(x), \text{ subject to } g_i(x) = 0, i = 1, \dots, p, x \in R^m$$

を考える. (不等号制約 $h(x) \leq 0$ は, 新しい変数 s を導入して, $h(x) + s^2 = 0$ と書き直せる.) 以下, ∇ は x に関する勾配を表すものとする.

目的関数 f の勾配を制約領域の接平面に射影したベクトル場を射影勾配と呼ぶことにする. 制約条件を満たす領域上での f の射影勾配流(射影勾配の積分曲線)は概念的には図 1 のようになり, 最適解は射影勾配流の停留点となる. 射影勾配流を定義する微分方程式を数値的に解くことで, 最適解を求めることができる. 田邊はこの微分方程式の漸近的な挙動等について解析した(Tanabe, 1974). これは, Rosen の射影勾配法の連続版である.

射影勾配流自身は, 制約領域を満たさない点においても, $g_i(x)$ の右辺を適切に(定数分)変更することで同様のものが定義できる. しかし, そのような初期値から出発して射影勾配流を数値的に積分しても制約条件を満たさないので, 最適解を求めることはできない. 一方, 一般には, 制約領域を満たさない点から出発しても, 制約領域に近づくことができるように, 制約条件が満たすべき方程式系 $g_i(x) = 0, i = 1, \dots, p$ 上の点を求める「ニュートン法」のベクトル場を定義することができる. この方程式系は変数の方が方程式の数よりも多い過少決定系であるが, 適当な正則条件の下で, 一般逆行列を用いることで, ニュートン法のベクトル場を定義できる. したがって, (いわばシフトした)射影勾配と制約領域を求めるニュートン法のベクトル場を加えたベクトル場を定義し, それを数値的に積分して最適解に至る解法を考えることができる(Yamashita, 1980). 田邊は(Tanabe, 1980)等でこれらのベクトル場が定義する力学系の漸近的な性質を詳しく解析した.

さらに, 田邊は,

$$\min f(x), \text{ subject to } g_i(x) = 0, i = 1, \dots, p, x \in R^m$$

に対するカルシュ・キューン・タッカー条件(一次の最適性条件)

$$\nabla f + J(x)^T \lambda = 0, \quad g(x) = 0, \quad \left(J_{ij}(x) = \frac{\partial g_i}{\partial x_j} \right)$$

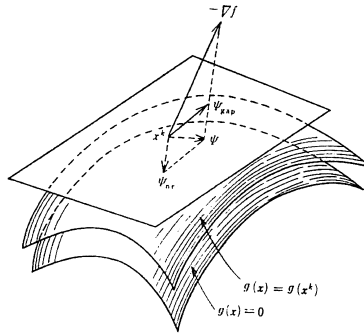


図 2. FIGAP 法概念図(田辺, 1981, 図 4 を引用; 図中 Ψ は探索ベクトル, Ψ_{nr} は制約条件を満たすためのニュートン方向, Ψ_{gap} は目的関数を最適化する方向で $\Psi = \Psi_{nr} + \Psi_{gap}$ となっている).

に対するニュートン法の反復式

$$(5.1) \quad \begin{pmatrix} \nabla^2 L(x, \lambda) & J^T(x) \\ J(x) & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \lambda^+ \end{pmatrix} = - \begin{pmatrix} \nabla f \\ g(x) \end{pmatrix}, \quad (L(x, \lambda) = f(x) + \lambda^T g(x))$$

において, (5.1) の左辺係数行列の $\nabla^2 L(x, \lambda)$ をさまざまな行列に置き換えることで, 非線形最適化アルゴリズム設計の統一的指針を得ることができることを見出し, この考え方に基づいて FIGAP (Feasibility-Improving-Gradient-Accute-Projection) 法を提案した (図 2) (Tanabe, 1981; 田辺, 1981).

これらのアルゴリズムの背後にある考え方は, 罰金法や障壁関数法のように問題を無制約最適化に落とし込むのではなく, カルシュ・キューン・タッカー条件を満たす点が漸近安定点となるようなベクトル場を最急降下法やニュートン法などを考慮しつつ適切に定義し, そのベクトル場の積分曲線を離散的に追跡することにより最適化アルゴリズムを構築するというものである. 田辺は折に触れてニュートン法の重要性を強調している. この思想は, ある意味でその後登場する内点法の考え方に通じるものである.

6. 内点法を巡って

線形計画法は, 1947 年, Dantzig によって創始されて以来, 現在にいたるまで, 数理計画法の中心にありつづけている. 線形計画法の解法としては Dantzig による単体法が定番であったが, 1970 年代に計算複雑度の理論が勃興し, 線形計画問題に対して所謂多項式時間解法が存在するか否かが一つの大きな未解決問題となった. この問題は, Khachiyan により, 初めての多項式時間解法である楕円体法の登場によって肯定的に解かれた. しかし, 楕円体法は理論的には数理計画法全体に大きく貢献したものの, 実用解法ではなかった. 1984 年に射影変換と最急降下法を繰り返して問題を解く射影変換法が多項式性を有する実用解法として Karmarkar によって提案され, 数理計画全体に大きな影響をおよぼす潮流を 20 年余りに渡り作り出した. Karmarkar に端を発する一連のアルゴリズムは, 実行可能領域の内部から最適解に近づいていく解法なので, 内点法と呼ばれる. 統計数理研究所では, 内点法についての研究が活発に行われた. 以下, 成果を紹介する.

次の線形計画問題 (P)

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

と双対問題 (D)

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

を考えよう. x, s の次元を n とし, A の行ベクトルは一次独立とする.

主問題 (P) の実行可能解 x における目的関数は, 双対問題の実行可能解 (s, y) における目的関数の値よりも大きいとか等しい. これは, 以下のようにして簡単にわかる.

$$c^T x - b^T y = x^T c - x^T A^T y = x^T s \geq 0.$$

上式の左辺の量を双対ギャップという. 双対ギャップが 0 の場合には, その時の x と (s, y) は, それぞれの問題の最適解となる. $x^T s = 0$ は, $x \geq 0, s \geq 0$ の条件の下では, $x_i s_i = 0$ と等価であるから, 主問題と双対問題を解くことは, 次のような双線形等式・不等式系を解くことと等価である.

$$(6.1) \quad x \circ s = 0, \quad Ax = b, \quad s = c - A^T y, \quad x \geq 0, \quad s \geq 0.$$

ここで, $x \circ s$ はその各要素が x と s の各要素の積であるベクトルを表す. $x > 0, s > 0$ であるような実行可能解を内点実行可能解という. 以下, 主問題と双対問題に内点実行可能解があることを想定する.

内点法の研究は, Karmarkar による射影変換法に端を発する. (P) に対する射影変換法は, 実行可能領域上で, Karmarkar ポテンシャル関数

$$f(x) = (n+1) \log(c^T x - c_{\text{opt}}) - \sum_{i=1}^{n+1} \log x_i = \log \frac{(c^T x - c_{\text{opt}})^{n+1}}{\prod x_i}$$

が減少する方向に進むことを繰り返すことで (P) を解く反復解法である. ここで c_{opt} は (P) の最適値である. 実行可能領域上で $f(x)$ が十分に小さくなると, $c^T x$ は c_{opt} に十分に近い値となる. 反復の各ステップでポテンシャルの減少方向を求めるのに射影変換と最急降下法を用いることで, 多項式時間アルゴリズムを構築できる. Karmarkar ポテンシャル関数は, $-\sum \log x_i$ を含む点に特色がある. すなわち, この項があることにより, (最適解以外の) 実行可能領域の境界に近づくとき $f(x)$ はその値が無限に発散するため, $f(x)$ は, 実行可能領域の内側を通過して効率よく最適解に近づくためのガイドの役割を果たすことができる. この関数

$$\psi(x) = -\sum \log x_i,$$

を対数障壁関数という. 対数障壁関数は第一象限の境界に近づくにつれてその値が無限大に発散する凸関数である. 制約領域において, 対数障壁関数を最小化する点を解析的中心と呼ぶ. 対数障壁関数は内点法の理論において重要な役割を果たす.

正のパラメータ ν を導入し, 対数障壁関数を用いて, 次のような最適化問題を考える.

$$(6.2) \quad \min c^T x + \nu \psi(x) \text{ subject to } Ax = b, \quad x \geq 0.$$

この問題の最適解は ν に依存する. これを $x(\nu)$ と記す. この問題は, $\psi(x)$ の項が不等式制約領域の境界に近づくにつれて $(x_i \rightarrow 0)$ で無限大に発散するため, 事実上不等式条件を無視できる最適化問題である. ν を小さくすれば, より境界に近いところに $x(\nu)$ は移動し, 直観的には $\nu \rightarrow 0$ で (P) の最適解に収束する. $x(\nu)$ の満たすべき条件は, 任意の $A\Delta x = 0$ なる Δx について,

$$\nabla(c^T x + \nu \psi(x))^T \Delta x = (c - \nu \text{diag}(x)^{-1} \mathbf{1})^T \Delta x = 0$$

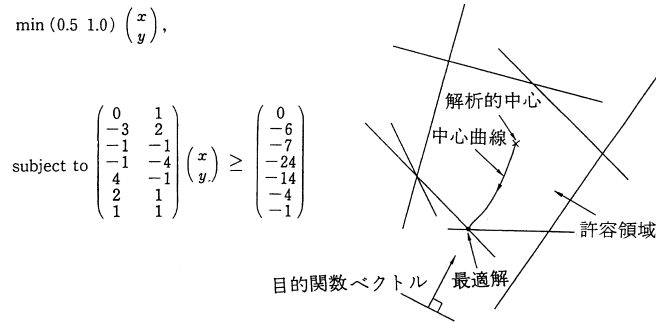


図 3. 内点法の概念図.

が成立することであるが、これは、

$$(6.3) \quad c - \nu \text{diag}(x)^{-1} \mathbf{1} = A^T y$$

がある y について成立することと同値である。ここで、ベクトル u に対し、 $\text{diag}(u)$ は、 u の各成分をその対角要素とする対角行列である。 ν を動かした時、 $x(\nu)$ は実行可能領域の内部を通り $\nu \rightarrow 0$ で (P) の最適解に収束する曲線となる。これを主中心曲線と呼ぶ。

さて、同様のことを双対問題について考えてみると、(6.2) に対応する問題は、

$$(6.4) \quad \max b^T y - \nu \phi(s), \quad s = c - A^T y$$

となる。双対問題においては、 $\phi(s) = -\sum \log(c_i - a_i^T y)$ が制約条件の境界に近づくにつれてその値が無限大に発散する障壁関数となる。パラメータ ν の時の最適解を $(s(\nu), y(\nu))$ と記すこととする。 $(s(\nu), y(\nu))$ の満たすべき最適性条件は、

$$(6.5) \quad b = \nu A \text{diag}^{-1}(s) \mathbf{1}, \quad s = c - A^T y$$

である。 ν を (正の範囲で) 動かした時、 $(s(\nu), y(\nu))$ は実行可能領域の内部を通り $\nu \rightarrow 0$ で (D) の最適解に収束する曲線となる。これを双対中心曲線と呼ぶ。

そこで、(P) を解くために、 ν を固定しておいて、 $x(\nu)$ を近似的にニュートン法によって求め、 ν の値を減らす、という手続きを繰り返して (P) の最適解を求める方法、あるいは、同様のことを (D) について行って (D) の最適解を求める線形計画問題の解法が考えられる。これが内点法の基本的な考え方である。前者を主内点法、後者を双対内点法という。図 3 に上で説明してきた内点法の概念を示す。

内点法の登場は衝撃的であった。次第に理論的にも実用的にも優れた解法であることが判明してくるにつれ、その研究のうねりは 80 年代後半から、20 年にわたり、最適化のみならず、隣接分野にも影響を及ぼす大きな潮流となっていった。後述するように、機械学習分野でサポートベクターマシンや圧縮センシングなどの新しいモデルを生み出す原動力ともなった。本節冒頭でも述べたように、統計数理研究所では、種々の内点法の解析や情報幾何との繋がりの研究が活発に行われた。

7. 主双対内点法

初期の内点法は、これまで説明してきたように、主問題あるいは双対問題のどちらかに点列を生成するものであったが、新しい立場の内点法アルゴリズムが、Megiddo によって示唆され、

田邊と小島・水野・吉瀬の 3 人によって、1987 年 2 月、統計数理研究所の研究集会「線形計画問題の新解法」において、世界で始めて、同時に提案された(小島 他, 1987; Tanabe, 1987a). 以下、その内容を述べる.

パラメータ $\nu > 0$ に対して、主中心曲線上でパラメータ値が ν である点 $x(\nu)$ と双対中心曲線上のパラメータ値が ν である点 $(s(\nu), y(\nu))$ は、実は、下記のような方程式の解となっていることが Megiddo によって指摘された(Megiddo, 1989).

$$x \circ s = \nu \mathbf{1}, \quad Ax = b, \quad s = c - A^T y, \quad x \geq 0, \quad s \geq 0.$$

この事実は(6.3)と(6.5)が満たしている関係を上式に代入することで簡単にわかる. この方程式は(6.1)の最初の方程式の右辺を 0 から各 i について等しく ν に緩めたものであるとも考えられる.

この事実は、主問題と双対問題を対等に扱う対称な解法が存在すること、そして、 $x \circ s$ の値を揃えながら減少させることで、効率良い解法が得られることを示唆する. つまり、障壁関数を用いて実行可能領域の境界から離れることにより、効率的な最適化を実現する、というメカニズムが、相補性条件 $x \circ s = 0$ を緩め、相補性に関する残差を揃えながら減少させる、という形でも実現しうることを示唆している.

このような背景の下で、田邊は、(Tanabe, 1987a, 1988)において、前節で発展させた、非線形最適化に関する微分幾何学的方法に「中心化」という、各方程式の残差を揃えながら減少させていく、という考え方を加えているいろいろな形に一般化し「中心化ニュートン法」という一連の解法を提案した. 特に、線形計画問題と凸 2 次計画問題については、現在主双対内点法として知られているものの具体形を与えた. そして、最適解への近さを計る尺度として、主双対ポテンシャル関数

$$f_T(x, s) = \rho \log(x^T s) - \sum_i \log(x_i s_i)$$

を導入した. ただし $\rho > n$ である. この関数は、

$$f_T(x, s) = (\rho - n) \log(x^T s) + g(x, s), \quad g(x, s) = n \log(x^T s) - \sum_i \log(x_i s_i)$$

と書き直すことができる. 簡単な解析により、 $g(x, s)$ は実行可能領域上では中心曲線上でのみ最小値 $n \log n$ をとることがわかる. そこで、田邊は $g(x, s)$ の値で中心曲線の近傍を定義することを提案した. また、上の表現より、 $\rho > n$ であれば、 $f_T(x, s)$ を $-\infty$ に減少させることで、双対ギャップ $x^T s$ を 0 に近づけることができ、(P) と (D) を解くことができることもわかる. したがって、主双対内点法の枠組みで、このポテンシャル関数を減少させるというアルゴリズムが考えられる. ポテンシャル関数 $f_T(x, s)$ は、現在では、Tanabe-Todd-Ye ポテンシャル関数、と呼ばれている. 田邊はさらに、主双対内点法の挙動の解析を連続ニュートン法や力学系の立場から行った. 図 4 に中心化ニュートン法概念図を示す. 一方、田邊の発表と同時に知られた、小島・水野・吉瀬の発表は、線形計画問題に対する主双対内点法の計算複雑度を解析し、多項式性の証明を与える点に力点があった. 後年、現在広く使われている非実行可能点列内点法をニュートン法の立場から一早く提案したことも、田邊の注目すべき業績の一つである(田邊, 1989).

論文(Kojima et al., 1989)の著者の一人である水野眞治(1990 年-1999 年在職)は研究所に合流後、さらに主双対内点法の研究を推進し、Todd や Ye 等と国際共同研究を行い、水野・Todd・Ye 予測子・修正子法(Mizuno et al., 1993)、非実行可能点列内点法に対する予測子・修正子法(Mizuno, 1996)、同次自己双対内点法(Ye et al., 1994)等の良く知られている成果を得た. 土

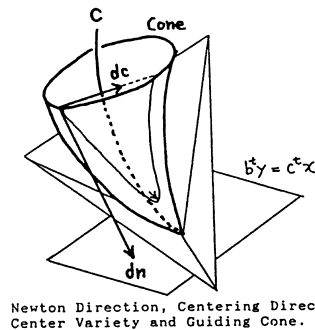


図 4. 中心化ニュートン法概念図(Tanabe, 1987a, Fig.6 を引用).

谷隆(1986 年-2010 年在職)も、総合研究大学院大学学生であった村松正和と共に、基本的な内点法であるアフィンスケーリング法の大域的収束性を示し(Tsuchiya and Muramatsu, 1995), このことをきっかけとして, Monteiro, Faybusovich, Megiddo, Terlaky, Roos 等と国際共同研究を展開した. また, El-Bakry, Tapia, Zhang と共に、一般の非線形最適化に対する主双対内点法の拡張についての論文を執筆した(El-Bakry et al., 1996). 論文(Mizuno, 1996; Mizuno et al., 1993; Ye et al., 1994; El-Bakry et al., 1996)は今でも折に触れて引用される標準的文献となっている. なお、本節で参照した田邊の中心化ニュートン法, 5 節および 9 節で紹介している最適化の微分幾何学に関する一連の研究は、共同研究リポート(田辺, 1996)にまとめられている.

8. 半正定値計画問題, 2 次錐線形計画への主双対内点法の拡張

線形計画問題に対する内点法の研究の進展が一段落した 1990 年代中盤より、線形計画問題における「変数の非負条件」を「変数がある凸錐に属している」という条件に拡張した最適化問題についての研究が進展した. その中でも典型的なものが、半正定値計画問題と 2 次錐計画問題であり、これらは以下のようにかけられる問題である. 半正定値計画問題は、正定値対称行列上の最適化問題で、主問題は

$$\min \operatorname{tr}(CX), \text{ subject to } \operatorname{tr}(A_i X) = b_i, i = 1, \dots, n, X \text{ は半正定値対称行列}$$

と書ける. この問題の双対問題は,

$$\max \sum b_i y_i, \text{ subject to } S = C - \sum_i A_i y_i, i = 1, \dots, n, S \text{ は半正定値対称行列}$$

となる. また、2 次錐計画問題は、線形計画問題における変数の非負条件を、変数がいくつかの 2 次錐の直積に属する、という条件に置き換えた、線形計画問題の拡張である. ここで 2 次錐とは、3 次元円錐の高次元版で、一般に n 次元 2 次錐は $\{(u_0, u_1) \in R \times R^{n-1} \mid \|u_1\| \leq u_0\}$ とかける集合である. 具体的には、2 次錐計画問題は下記のようにかける問題である.

$$\min c^T x, \text{ subject to } Ax = b, x \text{ はいくつかの 2 次錐の直積に属する}$$

この問題(主問題)の双対問題は,

$$\max b^T y, \text{ subject to } s = c - A^T y, s \text{ は上と同じ 2 次錐の直積に属する}$$

という問題である. 線形計画問題と同様、主問題の任意の実行可能解での目的関数値は双対問

題の任意の実行可能解での目的関数値よりも大きいとか等しく、等しければ、それぞれの問題の最適解となっている。また、緩い正則条件の下で主問題と双対問題の目的関数値が等しい最適解の存在が保証される。

半正定値計画問題や 2 次錐計画問題は、線形計画問題よりもより複雑な条件をモデリングすることができる。統計学、機械学習、制御、信号処理、最適設計、量子化学等多くの分野に応用を持ち、凸最適化の適用範囲を大きく広げた。線形計画問題の場合の主双対内点法を半正定値計画問題や 2 次錐計画問題に拡張することは、1990 年代後半の興味深い研究課題であった。土谷は、Monteiro とともに半正定値計画問題への主双対内点法について共同研究を進めるとともに (Monteiro and Tsuchiya, 1999)、ユークリッド的ジョルダン代数を用いて 2 次錐計画問題への多項式時間主双対内点法の拡張を行った (Tsuchiya, 1999; Monteiro and Tsuchiya, 2000)。また、Faybusovich とともに、2 次錐計画問題に対する主双対内点法を無限次元に拡張した (Faybusovich and Tsuchiya, 2003)。Tsuchiya (1999), Monteiro and Tsuchiya (2000) は 2 次錐計画問題に関する標準的文献として折に触れて引用されている。

9. 内点法の情報幾何

情報幾何は、甘利と長岡によってその枠組みが作られた、統計学や機械学習等の解析に有効な微分幾何である。統計数理研究所は、情報幾何と内点法の両方に関心を有する研究者がいる世界でも数少ない場であったこともあり、内点法の情報幾何の研究が活発に展開された。

情報幾何では、凸ポテンシャル関数とそのヘッセ行列として定まるリーマン計量、そして、2 つの互いに双対な測地線が基本的な役割を果たす。さて、山下によって、主内点法や双対内点法の探索方向が、対数障壁関数を最小化するためのニュートン方向と、アフィンスケーリング方向という、アフィンスケーリング法の探索方向から成っていることが指摘されていた (Yamashita, 1986)。アフィンスケーリング方向は、対数障壁関数のヘッセ行列を計量とする目的関数の最急降下方向である。アフィンスケーリング法はカーマーカー法よりも 20 年近く前にソ連の Dikin によって提案されていた世界最初の内点法である。Bayer と Lagarias は、Legendre 変換によってこれらの内点法の探索方向が直線化されることを見出した (Bayer and Lagarias, 1989)。田邊は双対内点法に現れるベクトル場の積分曲線の挙動を解析し、その立場から同様の事実を見出し、この変換を Center Flattening Transformation と名づけた (Tanabe, 1987b)。さらに、田邊と土谷は双対内点法に以下のような情報幾何的な構造があることを指摘した (田辺・土谷, 1988)。すなわち、多面体上の対数障壁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の構造を考えると、対数障壁関数のヘッセ行列によりリーマン計量が定義され、アフィンスケーリング法の連続版の軌跡は、情報幾何における双対測地線となっている。この時、主測地線は、元の空間での直線である。特に、中心曲線は、特別な双対測地線であり、これを離散的に近似的に辿ることによって、最適解を効率良く求めることができる。図 5 に図 3 の例題についてリーマン計量による“半径”1 の楕円(体)を、図 6 に双対測地線(アフィンスケーリング法の連続版の軌跡)を示す。図 5 の楕円体は Dikin 楕円体と呼ばれている。このような視点から、田邊は双対問題 (D) に対する中心化ニュートン法を提案し、関連する連続版の漸近的な挙動の解析を行った (田辺, 1990)。この線形計画問題に対する情報幾何の枠組みは、のちに小原敦美によって、半正定値計画問題に拡張された (小原, 1998)。そして、2000 年前後から、小原、土谷は柿原聡と共に凸最適化の情報幾何と内点法についての共同研究を行い、以下の成果を得た (Kakihara et al., 2013, 2014)。

線形計画問題の場合、凸最適化の情報幾何構造は以下ようになる。(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

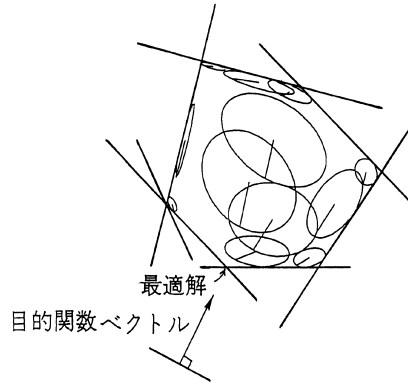


図 5. Dikin 楕円体.

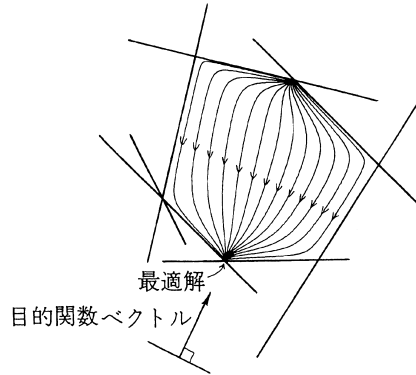


図 6. 双対測地線(連続版アフィンスケーリング法の軌跡).

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 7). 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

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

と表すのが自然となる. この時, 主双対中心曲線を $(x(t), s(t), y(t))$ と記すとすると, $h_P(t)$ 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, 中心曲線を追跡する, パス追跡型の主内点法の反復回数や双対内点法の反復回数が, 中心曲線上の積分

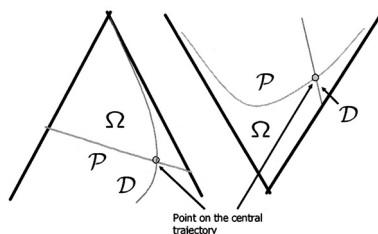


図 7. 最適化の情報幾何の概念図(両者は同じ問題を x と s による座標系で表現したものである).

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

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

を用いて, $I_P/\sqrt{\beta}$, $I_D/\sqrt{\beta}$ と書けることが分かる. ここで, β は, アルゴリズムが中心曲線を追跡する時にその中に反復列が生成される中心曲線の近傍の大きさを示すパラメータで, その値が小さいほど, より狭い近傍に留まりながら中心曲線を追跡することになり, 精密に曲線を追跡するため, 反復回数が大きくなる傾向がある. 定義からもわかるように, I_P , I_D 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, パス追跡型の主双対内点法の反復回数についても, 同様の積分表示

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

が求められ, 反復回数が $I_{PD}/\sqrt{\beta}$ とかけることが知られていた. この事実自身は (Monteiro and Tsuchiya, 2008) において情報幾何とは独立に導出されたことであるが, 一見情報幾何とは無関係に導出された量であった h_{PD} が実は, h_P と h_D を用いて定義される, 下記のような情報幾何学的量

$$(9.1) \quad h_{PD}^2 = h_P^2 + h_D^2$$

として書けるという少々驚くべき結果が (Kakihara et al., 2013) において得られた. これは「ピタゴラスの定理(関係)」を想起させる興味深い関係である. このように, 主双対内点法の反復回数を表す積分も情報幾何的な量として書くことができることが明らかとなった.

また, 論文 (Monteiro and Tsuchiya, 2008) では, L_A を (P), (D) の係数行列 A の入力ビット数として, 中心曲線全体に渡る I_{PD} の積分値が存在し, $I_{PD}(0, \infty) = O(n^{3.5} L_A)$ であることが示されている. この結果と上のピタゴラス関係を用いると, $I_P(0, \infty)$, $I_D(0, \infty)$, $I_{PD}(0, \infty)$ について,

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

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, $L_A = mn$ 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. これらは, 中心曲線の大域的な幾何学的性質を明らかにした結果である.

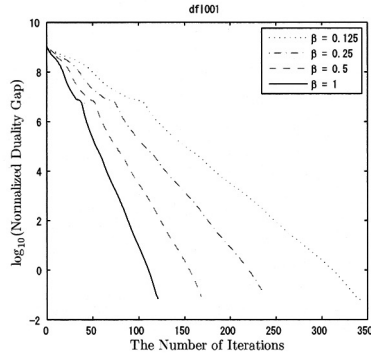
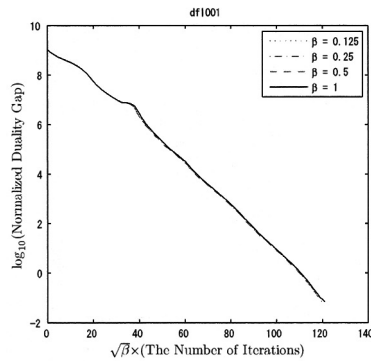
(a) 反復回数と正規化双対ギャップ ν (b) (反復回数 $\times \sqrt{\beta}$) と正規化双対ギャップ ν

図 8. 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著名なベンチマーク問題 DFL001 を例にして紹介する (Kakihara et al., 2014). DFL001 は制約式の数が 6072, 変数の数が 12230 の線形計画問題である. この問題に対して主双対内点法を行ったときの反復回数と双対ギャップの減少の関係が図 8(a) である. 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ルすれば, 重なりそうである. 反復回数を $\sqrt{\beta}$ 倍し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, 反復回数に $\sqrt{\beta}$ かけたものを横軸に取り直してグラフをプロットすると重なることがわかる (図 8(b)). 積分そのものは $\beta = 1$ とした時の反復回数であり, このように, 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. この結果は半正定値計画問題や 2 次錐計画問題にも拡張された (Kakihara et al., 2014).

10. 最適化モデリング, さまざまな研究

統計数理研究所で行われてきた最適化モデリングのひとつの典型例は, パラメータ推定問題を,

$$(10.1) \quad \min f(x, \theta) + \lambda g(\theta)$$

の形に帰着させるものである. x がデータ, θ がパラメータである. λ はハイパーパラメータ

である。直観的には、 f がデータへのモデルのあてはまりをあらわす(対数)尤度関数、 g はパラメータについての事前情報を表現する正則化項である。これは、現在では、Lasso などのスパースモデリングの形式として幅広く研究されている枠組みであるが、パラメータ数がデータよりも多いモデルも取り扱うことができ、柔軟なモデリングが可能となる点が長所である。統計数理研究所では、このタイプのモデリングが、赤池を中心に、1980 年代にベイズモデルの枠組みで精力的に研究されてきた。田邊は、共同研究者とともに、最適化を意識した立場から、地震波による地球内部構造推定や肺換気能の解析、肺気道の導通度の解析等にこの枠組みを積極的にもちいたモデリングを行い、先駆的な成果を得ている(Inoue et al., 1990; Kobayashi et al., 1987; Kawashiro et al., 1991)。特に Inoue et al. (1990) は地震波によるトモグラフィの基本的文献として今でも参照され続けている。

2 次錐計画問題に対する主双対内点法が開発された時には、リニアモーターカーの磁気シールドの最適設計問題が 2 次錐計画問題に定式化され、同手法で解かれている(Sasakawa and Tsuchiya, 2003)。これは、世界で最初の大規模な 2 次錐計画問題に対して主双対内点法を適用した例となった。解法の高速性により、多数回、ランダムに摂動を加えた問題を解きなおすことによって、設計されたシールドのロバストネスを評価することなども可能となった(土谷・笹川, 2005)。

ガウシアングラフィカルモデルの最尤推定は、半正定値計画問題に対する内点法と密接に関連した凸最適化問題に帰着する。この事実を用いて、上野玄太(2003 年より在職)と土谷は、スーパーコンピュータを用いて、海洋データ同化に現れる行列サイズ約 3 万 $\times$ 約 3 万、パラメータ数 10 万強の大規模分散共分散行列の推定を行った(Ueno and Tsuchiya, 2009)。

北川源四郎(1974 年-2011 年在職)が提案した粒子フィルタ(モンテカルロフィルタ, Kitagawa, 1996)は、統計数理研究所で開発され現在も幅広く使われている統計手法の一つである。土谷と中村和幸(当時総合研究大学院在学)は、関数の勾配を計算するための自動微分法のメモリ節約の手法が粒子フィルタのスムージングへ適用できることを指摘し、スムージングに使える粒子数を格段に増やすことができることを示した(Nakamura and Tsuchiya, 2007)。一風変わった形ではあるが、これも最適化と統計科学の交わった研究であるといえよう。

また、上田澄江(1970 年-2010 年在職)・牧野久実・伊藤栄明(1972 年-2007 年在職)と土谷は、2000 枚程度現在存在する、メソポタミア粘土板データに登場する人物の索引から、当時の有力人物の生存年代や人口動態、文書の成立した順番等を推定する問題を最適化モデリングの立場から扱った。内点法を用いて大規模凸 2 次計画問題を解くことで推定が可能となり、現在、さらに解析が進められている(Ueda et al., 2014)。

2000 年代の後半から、機械学習への最適化の適用が活発に行われてきた。この文脈では、伏木忠義(2003 年-2014 年在職)・土谷・堀内によって半正定値計画問題を用いた密度関数推定の手法が開発されている(Fushiki et al., 2006)。田邊は Kernel 法に触発され、双対理論の枠組みを活用する形で Dualized-Penalized-Logistic-Regression-Machine を提案し、松井知子(2003 年より在職)らと共にさまざまな分野への適用を試みている(田邊, 2007; Matsui and Tanabe, 2006)。また、武田朗子(2016 年-2018 年在職)や田中未来(2017 年より在職)等によって、DC(Difference of Convex functions)計画法の機械学習への適用や、非負行列分解の応用などの研究が進められている(Fujiwara et al., 2017; Mizutani and Tanaka, 2018)。

ここまで、幾つか焦点を絞って、最適化に関連する成果を中心に述べてきたが、統計数理研究所では、他にもさまざまな形で最適化に関する研究が進められてきた。例えば、半無限最適化計画問題は、有限次元のパラメータの最適化を無限個の制約の下で最適化する問題であり、凸錐上の線形計画問題を含む興味深い最適化問題のクラスであるが、この問題について、伊藤聡はしばしば国際共同研究を行い、さまざまな手法の開発に取り組んできた(Ito et al., 2000)。

本稿ではあまり取り上げられていなかったが、離散的最適化問題、特に整数計画問題は、重要な分野である。この分野については、モデリングの観点から、吉本敦(1999年-2003年在職、2008年より再度在職)によって、森林管理の最適計画などの研究が進められている(Yoshimoto et al., 2017)。また、伊藤聡(1991年より在職)らによってスポーツのリーグ戦におけるクリンチナンバーなどの計算もいち早く進められ、実用に供されている(伊藤, 2012; Ito and Shinano, 2018)。

11. 研究集会等

1944年に創立された統計数理研究所は1985年に改組され、国立大学共同利用機関となり、その後、大学共同利用機関となった。これに伴い、共同研究集会や、共同研究リポートの発行など、共同利用の体制が整えられた。先に述べた、内点法の研究が契機となって、1986年度に「線形計画問題の新解法」という研究集会が開催された。先に述べたように、この研究集会において、その後世界中で使われることとなった主双対内点法が、田邊と小島・水野・吉瀬より世界で初めて提案された。研究集会「線形計画問題の新解法」は、1988年度まで開催され、1989年度には「非線形最適化：モデリングとアルゴリズム」と名前を変えて引き継がれ、1990年度以降は「最適化：モデリングとアルゴリズム」の名前で、東日本大震災のあった2010年度を除いてずっと継続して開催されてきた。長年続いてきた最適化分野での研究集会として、若手を中心に多くの興味深い優れた研究成果が発表されてきた。各研究集会での発表は、共同研究リポートとしてまとめられている。第1節でも述べたように、最適化を意識したモデリングが重要である、ということが、本研究集会の開催趣旨の一つである。その考えを反映した形で「シリーズ：最適化モデリング」が立ち上げられ、現在刊行が進められている(赤池弘次 他, 2015; 山下浩 他, 2016; 池上敦子, 2018)。

これらの国内研究集会と同期する形で、国際研究集会が、1996年と2001年の2回、文部科学省からの予算を得て開催された。1996年に開催された国際研究集会“International Symposium on Optimization and Computation”は、湘南国際村で開催され、最適化に関する一流研究者が世界中から集い、研究成果を発表し、議論・情報交換を行った。2001年に京都の京大会館にて開催された研究集会“New Trends on Optimization and Computational Algorithms”では、最適化・機械学習・統計科学の一流研究者が集い、学際的交流を図った。これは、最適化と機械学習の交流を意識して開催された、最初期の研究集会の一つとして認識され、成果がMathematical Programming Series B, そして、Annals of the Institute of Statistical Mathematicsの特集として出版されている。また、2017年度より、統計数理研究所は九州大学のInstitute of Mathematics for Industry, ドイツのZuse Institute Berlin (ZIB)と共同して、最適化に力点を置いたワークショップISM-ZIB-IMI MODAL Workshopをもち回りで開催している。

1988年に、Mathematical Programming Societyにより3年に一度開催される、世界で最大の最適化関係の集会International Symposium on Mathematical Programmingが東京春日の中央大学で開催された。この時に、田邊は国際プログラム委員長の重責を担い、シンポジウムの成功に貢献した。28年後の2016年、Mathematical Optimization Society(旧Mathematical Programming Society)による、連続的最適化に関する最大の国際研究集会International Conference on Continuous Optimizationが、東京六本木の政策研究大学院大学で開催された。この時に、組織委員長を水野が、共同実行委員長を水野と土谷と村松が努めた。また、伊藤、武田と田中が実行委員会のメンバーとなった。統計数理研究所で若手時代に最適化の研究を行ってきた研究者が第一級の国際学会を日本に招致しそれを開催する中心となったことは、人材育成を担ってきた研究所の役割として、記憶に留められてもよいことではないかと思う。

12. 将来

最適化は、統計数理の新しい潮流と密接な関係を保ち続けている。1990 年代の後半から、サポートベクターマシン (SVM) が新しい判別機械として導入され、現在では完全に定着しているが、SVM の開発には、凸 2 次計画法が実用化されることが重要であった。主双対内点法によって大規模な凸 2 次計画問題が解けるようになって、はじめて SVM が実用化されたといえよう。SVM で重要な役割を果たしたスパース性の概念は、その後、Candes, Tao らによる圧縮センシング等、スパースモデリングの展開につながっていく。そこでは、線形計画法が重要な役割を果たし、2 次錐計画や半正定値計画も大きく寄与している。スパースモデリングは (10.1) に帰着するため、多くの最適化アルゴリズムの研究がさまざまな立場から行われた。ここに、1950 年代の赤池の論文から読み取れるような、数理計画と統計科学の実り多い交流を再び見出すことができるのである。アルゴリズムの進展が新しいモデルの実用化を促す好例である。スパースモデリングは、ブラックホールの撮像にも重要な役割を果たした。

一方、統計科学や機械学習の現場で使われるモデルが最適化研究を促すこともある。まず、機械学習における最適化問題は超大規模で、ヘッセ行列を必要とする、ニュートン法系統の解法は困難である。そのことにより、Nesterov に代表される、最急降下法を加速する手法や、確率的降下法などが提案されている。最近急激に研究が進んでいる深層学習においては、ReLU 関数の合成関数により関数が表現され、バックプロパゲーション (最急降下法) によってパラメータ最適化が行われている。ところが、ReLU 関数は原点で微分不可能であり、勾配を計算することが困難である。非凸関数であるために劣微分の理論も簡単ではない。このような問題についても、新しい光があたりつつある、といってもよいであろう。

21 世紀に入り、ビッグデータの時代を迎え、社会現象、経済現象、気象現象、災害、環境問題、マーケティング、社会保障等、科学技術の新しいフロンティアとチャレンジは、より複雑で不確実性を含んだ問題となりつつある。その中で予測とモデリングの科学としての統計科学は大きな貢献を成し得るものと考ええる。この文脈では、予測のみならず、モデルに基づいた制度設計や制御が必須となる。そこで重要な役割を果たすと考えられるのが、整数計画法と確率計画法である。これらの分野にも目配りをしながら、短期的な成果を求めず、腹を括って長期的な視点に立って総合的な立場で視野の広い人材を育成していくことが統計数理研究所における最適化の研究の展開にとっては重要であろう、と考えるし、そのような方向で研究所が発展することを望みたい。特に研究所の若手研究者諸賢におかれては、地に足のついた問題意識を持ち、統計数理研究所の伝統と文化を受け継ぎつつ、自らの力で少しずつでも着実に、統計数理の可能性を拓げる研究成果を積み上げていかれるよう、研究所で良き時代を過ごした一在職経験者として、心より願っている。

参 考 文 献

- Akaike, H. (1959). On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method, *Annals of the Institute of Statistical Mathematics*, **11**, 1-16.
- 赤池弘次 (1999). 応用数理の遊歩道 (17): 統計的思考と応用数理-形式的な確率的構造の利用, 応用数理, **9**, 169-171.
- 赤池弘次 他 (2015). 『モデリング—広い視野を求めて—』 (室田一雄, 池上敦子, 土谷隆 編), 最適化モデリング, 第 1 巻, 近代科学社, 東京.
- Bayer, D. A. and Lagarias, J. C. (1989). The nonlinear geometry of linear programming II. Legendre transform coordinates and central trajectories, *Transactions of the American mathematical society*, **314**, 527-581.

- El-Bakry, A., Tapia, R. A., Tsuchiya, T. and Zhang, Y. (1996). On the formulation and theory of the Newton interior-point method for nonlinear programming, *Journal of Optimization Theory and Applications*, **89**, 507-541.
- Faybusovich, L. and Tsuchiya, T. (2003). Primal-dual Algorithms and Infinite-dimensional Jordan Algebras, *Mathematical Programming*, **97**, 471-493.
- Fletcher, R. (2000). Practical Methods of Optimization, John Wiley & Sons, Chichester.
- Fujiwara, S., Takeda, A. and Kanamori, T. (2017). DC algorithm for extended robust support vector machine, *Neural Computation*, **29**, 1406-1438.
- Fushiki, T., Horiuchi, S. and Tsuchiya, T. (2006). A maximum likelihood approach to density estimation with semidefinite programming, *Neural Computation*, **18**, 2777-2812.
- 池上敦子 (2018). 『ナーススケジューリング』(室田一雄, 池上敦子, 土谷隆 編), 最適化モデリング, 第3巻, 近代科学社, 東京.
- Inoue, H., Fukao, Y., Tanabe, K. and Ogata, Y. (1990). Whole mantle P-wave travel time tomography, *Physics of the Earth and Planetary Interiors*, **59**(4), 294-328.
- Isii, K. (1960). The extrema of probability determined by generalized moments (I) bounded random variables, *Annals of the Institute of Statistical Mathematics*, **12**(2), 119-134.
- Isii, K. (1962). On sharpness of Tchebycheff-type inequalities, *Annals of the Institute of Statistical Mathematics*, **14**(1), 185-197.
- Isii, K. (1964). Inequalities of the types of Chebyshev and Cramér-Rao and mathematical programming, *Annals of the Institute of Statistical Mathematics*, **16**(1), 277-293.
- 伊藤聡 (2012). クリンチとエリミネーションの数理～リーグスポーツにおけるロバスト最適化, システム／制御／情報, **56**(7).
- Ito, S. and Shinano, Y. (2018). Calculation of clinch and elimination numbers for sports leagues with multiple tiebreaking criteria, *ZIB Report* 18-51, Zuse Institute Berlin.
- Ito, S., Liu, Y. and Teo, K. L. (2000). A dual parametrization method for convex semi-infinite programming, *Annals of Operations Research*, **98**, 189-213.
- 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.
- Kakihara, S., Ohara, A. and Tsuchiya, T. (2014). Curvature integrals and iteration complexities in SDP and symmetric cone programs, *Computational Optimization and Applications*, **57**, 623-665.
- Kawashiro, T., Yamasawa, F., Okada, Y., Kobayashi, H. and Tanabe, K. (1991). Uneven distribution of ventilation-perfusion ratios in lungs estimated by a modified Newton method, *Mathematical Programming*, **52**(1-3), 1-9.
- Kitagawa, G. (1996). Monte Carlo filter and smoother for Non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics*, **5**(1), 1-25.
- Kobayashi, H., Abe, T., Kawashiro, T., Tanabe, K. and Yokoyama, T. (1987). Estimation of the distribution profile of airway resistance in the lungs, *Computers and Biomedical Research*, **20**(6), 5070-5525.
- 小島正和・水野眞治・吉瀬章子 (1987). 多項式オーダーの主双対内点法. 統計数理研究所共同研究レポート 5, “線形計画問題の新解法,” 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.
- Lanckriet, G. R. G., Ghaoui, L. El, Bhattacharyya, C. and Jordan, M. I. (2002). A robust minimax approach to classification, *Journal of Machine Learning Research*, **3**, 555-582.
- Lasserre, J.-B. (2010). Moments, positive polynomials and their applications, Imperial College Press, London, UK.
- Matsui, T. and Tanabe, K. (2006). Comparative study of speaker identification methods: DPLRM,

- SVM and GMM, *IEICE Transactions on Information and Systems*, **E89-D**, 1066-1073.
- Megiddo, N. (1989). Pathways to the optimal set in linear programming. *Progress in Mathematical Programming* (ed. N. Megiddo), 131-158, Springer-Verlag, New York.
- Mizuno, S. (1996). Polynomiality of infeasible-interior-point algorithms for linear programming, *Mathematical Programming*, **67**(1-3), 109-119.
- Mizuno, S., Todd, M. J. and Ye, Y. (1993). On adaptive-step primal-dual interior-point algorithms for linear programming, *Mathematics of Operations Research*, **18**, 964-981.
- Mizutani, T. and Tanaka, M. (2018). Efficient preconditioning for noisy separable nonnegative matrix factorization problems by successive projection based low-rank approximations, *Machine Learning*, **107**, 643-673.
- Monteiro, R. D. C. and Tsuchiya, T. (1999). Polynomial convergence of a new family of primal-dual algorithms for semidefinite programming, *SIAM Journal on Optimization*, **9**, 551-577.
- Monteiro, R. D. C. and Tsuchiya, T. (2000). Polynomial convergence of primal-dual algorithms for the second-order cone program based on the MZ-Family of directions, *Mathematical Programming*, **88**, 61-83.
- Monteiro, R. D. C. and Tsuchiya, T. (2008). 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.
- Nakamura, K. and Tsuchiya, T. (2007). A recursive recomputation approach for smoothing in nonlinear state-space modeling: An attempt for reducing space complexity, *IEEE Transactions on Signal Processing*, Vol. 55, 5167-5178.
- Nesterov, Y. (2004). *Introductory Lectures on Convex Optimization*, Boston, Dordrecht, London.
- Nocedal, J. and Wright, S. (2000). *Numerical Optimization*, Springer-Verlag.
- 小原敦美 (1998). 半正定値計画問題に関する情報幾何を用いた考察, 統計数理, **46**(2), 317-334.
- Sasakawa, T. and Tsuchiya, T. (2003). Optimal magnetic shield design with second-order cone programming, *SIAM Journal on Scientific Computing*, **24**(6), 1930-1950.
- Tanabe, K. (1971). Projection method for solving a singular system of linear equations and its applications, *Numerische Mathematik*, **17**(3), 203-214.
- Tanabe, K. (1974). An algorithm for the constrained maximization in nonlinear programming, *Journal of the Operations Research Society of Japan*, **14**, 184-201.
- 田辺國士 (1976). アナログ型非線形計画法, 数理科学, **157**, 45-51.
- Tanabe, K. (1980). Geometric method in nonlinear programming, *Journal of Optimization Theory and Applications*, **30**, 181-210.
- Tanabe, K. (1981). Feasibility-improving gradient-acute-projection methods: A unified approach to nonlinear programming, *Lecture Notes in Numerical and Applied Analysis*, **3**, 57-76.
- 田辺國士 (1981). 最適化問題とニュートン法, 数理科学, **218**, 71-76.
- Tanabe, K. (1987a). Complementarity-enforcing centered Newton method for linear programming, 「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, 118-144. 統計数理研究所, 東京.
- Tanabe, K. (1987b). Center flattening transformation and a centered Newton method for linear programming, 日本オペレーションズ・リサーチ学会数理計画研究部会 1987 年 7 月部会発表資料, ((田辺國士, 1996) 統計数理研究所共同研究リポート, No. 89, 中心化ニュートン法と最適化の微分幾何, 1996 年, 5-18.), 統計数理研究所, 東京.
- Tanabe, K. (1988). Centered Newton method for mathematical programming, *System Modelling and Optimization* (eds. M. Iri and K. Yajima), Lecture Notes in Control and Information Sciences, Vol. 113, 197-206, Springer, Berlin, Heidelberg.
- 田辺國士 (1989). Centered Newton method for linear programming: Exterior point method, 統計数理, **37**(1), 203-204.
- 田辺國士 (1990). 最適化の微分幾何学, 応用数理, 創刊準備第 3 号, 39-50.
- 田辺國士 (1996). 中心化ニュートン法と最適化の微分幾何, 統計数理研究所共同研究リポート, No. 89,

- 統計数理解研究所, 東京.
- 田邊國士 (2007). 帰納推論機械 PLRM と dPLRM: 方法論, モデル, アルゴリズムおよび応用, システム／制御／情報, **51**(2), 87-95.
- 田辺国士, 土谷隆 (1988). 線形計画の新しい幾何学, 数理科学, **303**, 32-37.
- 統計数理解研究所 (1994). 統計数理解研究所 50 年のあゆみ, 統計数理解研究所, 東京.
- Tsuchiya, T. (1999). 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.
- Tsuchiya, T. and Muramatsu, M. (1995). Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems, *SIAM Journal on Optimization*, **5**(3), 525-551.
- 土谷隆, 笹川卓 (2005). 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.
- Ueda, S., Makino, K., Itoh, Y. and Tsuchiya, T. (2014). Logistic growth for the Nuzi cuneiform tablets: Analyzing family networks in ancient Mesopotamia, *Physica A*, **421**, 223-232.
- Ueno, G. and Tsuchiya, T. (2009). Covariance regularization in inverse space, *Quarterly Journal of the Royal meteorological Society*, **135**(642), 1133-1156.
- Wolkowicz, H., Saigal, R., Vandenberghe, L. (eds.) (2000). *Handbook of Semidefinite Programming*, Kluwer, Boston, Dordrecht, London.
- Yamashita, H. (1980). A differential equation approach to nonlinear programming, *Mathematical Programming*, **18**, 155-168.
- Yamashita, H. (1986). *A Polynomially and Quadratically Convergent Method for Linear Programming*, Mathematical Systems Inc., Tokyo.
- 山下浩 他 (2016). 『モデリングの諸相—OR と数理科学の交叉点—』(室田一雄, 池上敦子, 土谷隆 編), 最適化モデリング, 第 5 巻, 近代科学社, 東京.
- Ye, Y., Todd, M. J. and Mizuno, S. (1994). An $O(\sqrt{n}L)$ -iteration homogeneous and self-dual linear programming algorithm, *Mathematics of Operations Research*, **19**(1), 53-67.
- Yoshimoto, A., Asante, P., Konoshima, M. and Surovy, P. (2017). Integer programming approach to control invasive species spread based on cellular automaton model, *Natural Resource Modeling*, **30**(2), NRM12101.

A Historic Review of Research on Optimization at the Institute of Statistical Mathematics

Takashi Tsuchiya^{1,2}

¹National Graduate Institute for Policy Studies

²The Institute of Statistical Mathematics

Optimization is an important discipline in statistical science. Broadening the frontier of tractable optimization problems directly leads to innovation of statistical science. This is a main reason why optimization research has been conducted at the Institute of Statistical Mathematics. In this paper, we review historic development of optimization research at the institute, and its contribution to statistical sciences. The topics include duality theory of infinite-dimensional linear programming and its application to statistics, convergence analysis of the steepest descent algorithm and Kaczmarz method, differential-geometric approach to nonlinear programming, interior-point algorithms and information geometry for linear programming, second-order cone programming and semidefinite programming, and various applications.