

## 最尤推定の最適性と KCR 下界

金谷 健一

岡山大学工学部情報工学科

幾何学的当てはめはコンピュータビジョンの最も基本的な問題の一つである。筆者は以前これに対する精度の理論限界 (KCR 下界) を導き、最尤推定が統計的に最適であることを証明した [9, 10]。最近, Chernov ら [3] は、これが筆者の用いた仮定を弱めても成立することを証明している。本稿ではこれを紹介し、その筆者の定式化との相違や問題の背景, セミパラメトリックモデルなどの最近の話題や今後の課題を検討する。

### Optimality of Maximum Likelihood Estimation and the KCR Lower Bound

Kenichi Kanatani

Department of Information Technology, Okayama University, Okayama 700-8530 Japan

Geometric fitting is one of the most fundamental problems of computer vision. In [9, 10], the author derived a theoretical accuracy bound (KCR lower bound) for the geometric fitting problem in general and proved that maximum likelihood estimation is statistically optimal. Recently, Chernov and Lesort [3] proved a similar result, using weaker assumptions than the author's. In this paper, we compare their formulation with the author's and describe the background of the problem. We also review recent topics including semiparametric models and discuss remaining problems.

#### 1. 何が問題か

幾何学的当てはめとは、観測したデータに対して、成立すべき幾何学的拘束条件を当てはめてパラメータを計算し、データの背後にある幾何学的構造を知ることである [10]。コンピュータビジョンの非常に広い範囲の問題がこれに相当する。最も単純な例は画像上に与えられた  $N$  個の点列  $\{(x_\alpha, y_\alpha)\}$  に曲線 (直線, 円, 楕円, 多項式曲線など)

$$F(x; \mathbf{u}) = 0 \quad (1)$$

を当てはめる問題である。ただし  $x = (x, y)^\top$  と置き、曲線のパラメータをベクトル  $\mathbf{u} = (u_1, \dots, u_p)^\top$  で表した。

データ  $\{(x_\alpha, y_\alpha)\}$  に誤差があるとき、 $x_\alpha = (x_\alpha, y_\alpha)^\top$  と置くと、 $F(x_\alpha; \mathbf{u}) = 0$  が  $\alpha = 1, \dots, N$  のすべてに対して成立する解  $\mathbf{u}$  は存在しない。これに対する素朴な対処法は

$$J_{LS} = \sum_{\alpha=1}^N F(x_\alpha; \mathbf{u})^2 \rightarrow \min \quad (2)$$

となる  $\mathbf{u}$  を定める最小二乗法 (または代数距離最小化) である。しかし、この解には一般に統計的な偏差が存在し、理論的にも実験的にもよい精度を与えないことが知られている。

理論的にも実験的にもよりよい精度を与える方法は、データ  $\{x_\alpha\}$  が曲線  $F(x; \mathbf{u}) = 0$  の“真の位置”

†700-8530 岡山市津島中 3-1-1, TEL/FAX: (086)251-8173  
E-mail: kanatani@suri.it.okayama-u.ac.jp

$\{\bar{x}_\alpha\}$  から等方、独立な正規分布に従う誤差によってずれたものであるとみなし、真の位置  $\{\bar{x}_\alpha\}$  とパラメータ  $\mathbf{u}$  を同時に尤度が最大になるように推定する、すなわち、拘束条件

$$F(\bar{x}_\alpha; \mathbf{u}) = 0, \quad \alpha = 1, \dots, N \quad (3)$$

のもとで

$$J_{ML} = \sum_{\alpha=1}^N \|x_\alpha - \bar{x}_\alpha\|^2 \rightarrow \min \quad (4)$$

とする最尤推定 (または幾何学距離最小化) である。ラグランジュ乗数を導入すれば拘束条件 (3) は除去できる。誤差が小さいときは線形近似を用いると、式 (4) は次のように書ける (導出を付録 A に示す)。

$$J_{ML} = \sum_{\alpha=1}^N \frac{F(x_\alpha; \mathbf{u})^2}{\|\nabla_x F_\alpha\|^2} \rightarrow \min \quad (5)$$

ただし、 $\nabla_x F_\alpha$  は式 (1) の関数  $F(x; \mathbf{u})$  の  $x$  に関する勾配であり、 $F_\alpha$  は微分を  $x = x_\alpha$  において評価することを表す。この方法の有効性は多くの問題で実証され、現在広く用いられている [10]。

この考え方は曲線当てはめに限らず、コンピュータビジョンの多くの問題に拡張できる。例えば複数画像間の特徴点の対応が得られていれば、各点の“軌跡”は画像の直積空間 (直積画像) の 1 点とみなせる。それにカメラの撮像モデルから導かれる幾何学的拘束条件 (エピ極線拘束条件, 3 重焦点拘束条

件，4重焦点拘束条件，アフィン拘束条件など）を当てはめれば，カメラの運動や物体の3次元形状を計算することができる[7]．誤差も一様等方正規分布を仮定する必要はなく，共分散行列を導入して誤差モデルを一般化することもできる<sup>1</sup>．

筆者は，変数変換によって $F(x; u)$ が $u$ の線形式に書き直せる問題<sup>2</sup>の最尤推定解を組織的に計算するくりこみ法を提案した[8]．その後，これはいろいろな形に発展し，Leedanら[14]はHEIV法，Chojnackiら[4]はFNS法を提案している<sup>3</sup>．

しかし，未だに議論となるのは，式(5)が本当に最適なのか，これよりよい方法は本当に存在しないのか，という疑問である．

## 2. 方法をどう評価するか

この問題が困難なのは，方法の“よさ”をどう測るかが明らかでないためである．解 $\hat{u}$ の精度は真の値 $u$ との差のノルム $\|\hat{u} - u\|$ で測るとしよう．これにも異議が多く， $u$ がとりそうな値に関する我々の信念や期待を表す事前分布について期待値をとるべきだとか<sup>4</sup>，その値 $\hat{u}$ を用いて行う応用の誤差（例えば3次元復元に使うなら復元形状の誤差）で評価すべきであるなど，さまざまな異論がある．

しかし，最も素朴な $\|\hat{u} - u\|$ を採用しても問題は解決しない．なぜなら，誤差はランダムであるから，どんな方法でもたまたま $\hat{u}$ が $u$ に一致することもあるからである．そこでその二乗平均 $E[\|\hat{u} - u\|^2]$ を用いることが考えられる． $E[\cdot]$ は各データ点の誤差の分布に関する期待値である．二乗平均をとるのは，通常は解析が最も簡単になるからである．これに対しても， $\max \|\hat{u} - u\|$ がよいとか $E[\|\hat{u} - u\|]$ であるべきだなど種々の異論がある．しかし，最も扱いやすい二乗平均を用いても，解析が極めて複雑になり，異なる方法を比較するのは困難である．

これに対して，統計学では試行回数 $n$ が大きいときの漸近近似を用いて式を単純化して推定手法を比較することがよく行われる．そこで，幾何学的当てはめに対してもデータ点の個数 $N$ が大きいときの漸近近似を用いる研究が多い．しかし，データ点の個数 $N$ は“試行回数”なのであるうか．

<sup>1</sup>さらにデータ $x$ やパラメータ $u$ に何らかの（例えば単位ベクトルであるなどの）制約があったり，式(1)の形の拘束条件が複数あり，しかも重複したり冗長であったりしてもよい．これらは一般逆行列や射影作用素を導入して解析することができる[10]．

<sup>2</sup>直線や楕円当てはめ，射影変換や基礎行列の計算などのコンピュータビジョンの非常に多くの応用がこれに含まれる．

<sup>3</sup>その歴史的な経緯や最近の発展は文献[12]に詳しい．

<sup>4</sup>いわゆるベイズ推定と呼ばれる方法論[20]である．

## 3. データをどう増やすか

統計学の基本的な課題は，ある確定的なメカニズムで生じる結果にランダムな誤差が加わるとき，その観測データから背後のメカニズムを推定することである．これを一度の観測で行うのは不可能であるが，誤差はランダムであるという性質に着目すれば，多数回の観測を繰り返すほどランダム誤差の影響が相殺され，確定的なメカニズムが発現すると期待される．このことから，観測回数 $n$ を増やすとどれだけ精度が向上するかという， $n$ に対する精度の向上率で手法を評価することが行われるようになった．しかし，幾何学的当てはめにおいて，データ点の個数 $N$ を“試行回数”と同一視するとさまざまな矛盾が生じる[13]．

第1は，統計的推定では“原理的に”（すなわち，観測にコストがかかり，現実的な制約があるという事実を除けば）その現象を何度でも観測できる．しかし，コンピュータビジョンでは入力1枚の画像，または1系列の動画画像から得られる1組のデータであり，それに異なった誤差が加わったデータは（シミュレーション以外では）得ることができない．したがって観測回数は常に $n = 1$ である．

第2は，統計的推定では未知数は背後にある確定的なメカニズムのパラメータであるが，幾何学的当てはめではデータの真の値も未知である．したがってデータが増えればそれだけ推定すべき未知数も増え<sup>5</sup>，しかも，それらの推定はいくらデータ点を増やしても改善できない．例えば，曲線当てはめでは，真の曲線が推定できても，その曲線上の位置は確定できない．

第3は，データ点の個数 $N$ を単に“増やす”というだけでは不完全であり，“どのように増やすか”まで考慮しなければならないことである．例えば直線当てはめでは，ある点の近傍にいくらデータ点を増やしても当てはめの精度は向上しない．しかし，その直線に沿って限りなく遠方まで一様に分布するように増やすと当てはめの精度は著しく向上する．そこで従来から，データ点の真の位置の分布を仮定，または推定し，真の位置をその分布に関して“周辺化”する解析（セミパラメトリックモデルと呼ばれる）がさまざまな形でなされている[2, 16, 17, 18]．

## 4. 最尤推定は最適ではないのか

例えば，オプティカルフローデータのように極めて多数のデータがあるときは最尤推定が最適ではな

<sup>5</sup>そのような増える未知数をかく乱母数，それ以外の未知数を構造母数と呼んで区別することがある．

いことは遠藤ら [5, 6] が指摘していた．そして最近，セミパラメトリックモデルによって最尤推定より優れる方法が存在することが岡谷・出口 [18] や太田・栗原 [17] によって指摘されている．しかし，いずれの改善手法の非常に複雑であり，それが有効であるためにはデータが非常に多く，かつ問題が特殊な形をしているなど，多くの限定がある．

一方，式 (5) の最尤推定は現実の多くの問題で効果を発揮しており，通常の場合にはこれに勝る方法は考えられていない．ということは，最尤推定が通常の場合には何らかの意味で最適なのではなからうか．そうだとすれば，どういう意味で最適なのであるうか．

これに対して筆者は一つの回答を与えた [9, 10]．このようなことが統計学で検討されていないことが筆者には不思議であったが，その後の統計学者との議論からその理由が判明した．それは，前述のように統計学とは多数回の観測によってランダムな誤差を克服するもの，推定の評価は観測回数  $n \rightarrow \infty$  の漸近近似で行なうもの，というパラダイムが支配していることで，コンピュータビジョンに現れるような幾何学的推定に関心が持たれない（知らない）ということである．

以下，筆者の定式化述べ，これに対して最近発表された Chernov ら [3] による拡張を紹介する．そして，筆者の定式化との相違や今後の課題を検討する．

## 5. KCR 下界

筆者の定式化は，データ数  $N$  が大きいときの漸近近似ではなく，誤差  $\varepsilon$  が微小なときの漸近近似を考えるというものである．これは統計学が想定するフィールドワークのような統計調査とは異なり，画素レベルの誤差を扱うコンピュータビジョンでは妥当と考えられる．

データ  $\{x_\alpha\}$  からパラメータ  $u$  を推定することは，その推定値  $\hat{u}$  をデータ  $\{x_\alpha\}$  の関数

$$\hat{u} = \hat{u}(x_1, \dots, x_N) \quad (6)$$

として表すことである．この関数  $\hat{u}$  を  $u$  の推定量と呼ぶ．この推定量  $\hat{u}$  の共分散行列

$$V[\hat{u}] = E[(\hat{u} - u)(\hat{u} - u)^\top] \quad (7)$$

を考える<sup>6</sup>．各データ  $x_\alpha$  はその真の値  $\bar{x}_\alpha$  の各成分に独立に期待値 0，標準偏差  $\varepsilon$  の正規分布に従う誤差が加わる，すなわち

$$x_\alpha = \bar{x}_\alpha + \Delta x_\alpha, \quad \Delta x_\alpha \sim N(\mathbf{0}, \varepsilon^2 \mathbf{I}) \quad (8)$$

<sup>6</sup>そのトレース  $\text{tr}V[\hat{u}] = E\|\hat{u} - u\|^2$  が二乗平均誤差である．

とし， $\varepsilon$  をノイズレベルと呼ぶ．以下の議論は一般的な確率分布<sup>7</sup>に従う誤差に対しても行なえるが [9, 10]，簡単のためにここでは正規分布に従う一様等分散の場合で考える．

推定量  $\hat{u}$  の誤差を  $\Delta u$  と置く．

$$\hat{u} = u + \Delta u \quad (9)$$

筆者は式 (8), (9) を式 (5) に代入し，誤差が小さい極限を考えて微小量  $\Delta x_\alpha$ ,  $\Delta u$  に関してテイラー展開して，これを最小とする  $\Delta u$  を評価した．そして，最尤推定量  $\hat{u}_{\text{ML}}$  の共分散行列  $V[\hat{u}_{\text{ML}}]$  が次のように展開できることを示した [9, 10]．

$$V[\hat{u}_{\text{ML}}] = \varepsilon^2 \left( \sum_{\alpha=1}^N \frac{(\nabla_u \bar{F}_\alpha)(\nabla_u \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} \right)^{-1} + O(\varepsilon^4) \quad (10)$$

ただし， $\nabla_u \bar{F}_\alpha$  は式 (1) の関数  $F(x; u)$  の  $u$  に関する勾配であり， $\bar{F}_\alpha$  は微分を  $x = \bar{x}_\alpha$  において評価することを意味する．導出を付録 B に示す．

一方，筆者は式 (10) の右辺第 1 項が任意の不偏推定量  $\hat{u}$  に対して次の意味で下界になっていることを示した [9, 10]．

$$V[\hat{u}] \succ \varepsilon^2 \left( \sum_{\alpha=1}^N \frac{(\nabla_u \bar{F}_\alpha)(\nabla_u \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} \right)^{-1} \quad (11)$$

ただし， $\succ$  は左辺から右辺を引いたものが半正値対称行列であることを表す．導出を付録 C に示す．

以上より，最尤推定量  $\hat{u}_{\text{ML}}$  の共分散行列はノイズレベル  $\varepsilon$  の第 1 近似において（すなわち  $O(\varepsilon^4)$  を除いて）精度が下限に到達しており，この意味で最尤推定は最適といえる．Chernov ら [3] は式 (11) の右辺を KCR（金谷・クラメル・ラオ）下界と呼んでいる．

## 6. CR 下界

それでは上記の KCR 下界は通常の CR（クラメル・ラオ）下界とどう違うのであろうか．これは問題の違いである．前述のように，統計的推定とは多数回の観測によって背後にあるメカニズムを推定することである．形式化すると，ある未知パラメータ  $\theta$  をもつ確率密度  $p(x; \theta)$  に従って発生する確率変数  $X$  の独立な実現値  $x_1, \dots, x_n$  を観測し，これから未知パラメータ  $\theta$  を推定する問題となる．最尤推定とは尤度

$$L = \prod_{i=1}^n p(x_i; \theta) \quad (12)$$

<sup>7</sup>例えば指数分布族と呼ばれるクラスに属する分布であればよい [9, 10]．

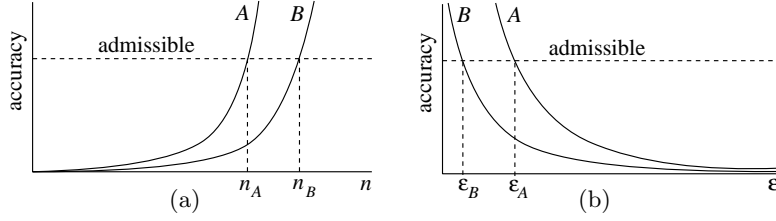


図 1: (a) 通常の統計的推論では観測回数  $n \rightarrow \infty$  で急速に精度が向上することが望ましい．なぜなら，より少ない観測回数で許容精度を達成できるからである．(b) 幾何学的当てはめではノイズレベル  $\varepsilon \rightarrow 0$  で急速に精度が向上することが望ましい．なぜなら，より大きな不確定性があっても許容精度が達成できるからである

を最大にする  $\theta$  の値  $\hat{\theta}_{ML}$  を求めることである．ここで  $n \rightarrow \infty$  の極限を考えると，独立な確率変数のサンプル平均はその期待値で近似されるという大数の法則と，その分布が正規分布に近づくという中心極限定理が成り立つ．このことから  $p(x; \theta)$  が変則でない限り最尤推定量  $\hat{\theta}_{ML}$  の共分散行列が次のように  $1/n$  に関して展開できる．

$$V[\hat{\theta}_{ML}] = \frac{1}{n} J^{-1} + O\left(\frac{1}{n^2}\right) \quad (13)$$

ここに  $J$  は次のように定義されるフィッシャー情報行列である．

$$J = E\left[\left(\nabla_{\theta} \log p(x; \theta)\right)\left(\nabla_{\theta} \log p(x; \theta)\right)^{\top}\right] \quad (14)$$

ただし  $E[\cdot]$  は確率密度  $p(x; \theta)$  に関する期待値である．式 (13) の右辺第 1 項が CR (クラメル・ラオ) 下界と呼ばれ，任意の不偏推定量  $\hat{\theta}$  に関して次の不等式が証明される．

$$V[\hat{\theta}] \succ \frac{1}{n} J^{-1} \quad (15)$$

以上より，最尤推定量  $\hat{\theta}_{ML}$  は観測回数  $n \rightarrow \infty$  の極限で第 1 近似において (すなわち  $1/n$  の高次の項を除いて) CR 下界に到達しており，この意味で最適であるといえる．これが最尤推定の漸近的有効性と呼ばれる性質である．

## 7. 解釈の双対性

以上のように，KCR 下界と CR 下界は全く異なる概念である．しかし，何らかの類似性も感じられる．

統計的推定で  $n$  に関する漸近解析が行われるのは，観測回数  $n \rightarrow \infty$  で急速に精度が向上する推定法はそうでない方法に比べてより少ない観測回数で許容精度を達成できるからである (図 1 (a))．現実問題の観測コストを考えればこれは妥当である．

一方，幾何学的当てはめに  $\varepsilon$  に関する漸近解析が適しているのは，ノイズレベル  $\varepsilon \rightarrow 0$  で急速に精度が向上する推定法はそうでない方法に比べてより不

確定な画像データに対しても許容精度が達成できるからである (図 1(b))．画像処理に不可欠の不確定性を考えればこれは妥当である．

しかし，次のように考えてみる．幾何学的当てはめでは各データ点の真の位置に対して誤差の加わった位置は，その画像を何度観測しても常に同じである．しかし，仮想的に画像を見るたびにその位置が変動する状況を考えて， $n$  回画像を見れば  $n$  個の位置が得られる．それらから真の位置を最適に推定するには正規分布モデルのもとではサンプル平均をとればよい．サンプル平均の分散はもとの分散の  $1/n$  になるから，このような仮想的な推定は式ノイズレベル  $\varepsilon$  を  $1/\sqrt{n}$  倍することと等価である．この結果， $\varepsilon \rightarrow 0$  の摂動解析は仮想的な  $n$  回の観測の  $n \rightarrow \infty$  の漸近解析に相当し， $n$  に関する漸近評価  $\dots + O(1/\sqrt{n^k})$  が  $\varepsilon$  に関する漸近評価  $\dots + O(\varepsilon^k)$  として現れる [13]．

このような解釈の双対性はモデル選択に対しても当てはまり，統計的推定に対する赤池の AIC [1] や Rissanen の MDL [19] の幾何学的当てはめに対する規準として幾何学的 AIC と幾何学的 MDL が得られる [11]．

## 8. 最適性の条件

式 (11) の KCR 下界は発表以来まったく認められず，その成立を疑う者さえあった<sup>8</sup>．しかし，最近 Chernov ら [3] は，これが筆者が示したより弱い条件のもとで成立することを証明した．筆者は  $\hat{u}$  が任意の不偏推定量，すなわち

$$E[\hat{u}] = u \quad (16)$$

のときに式 (11) が成立することを示したが [9, 10]，Chernov ら [3] はこれが次の一貫性のもとで成立することを示した．

$$\lim_{\varepsilon \rightarrow 0} \hat{u} = u \quad (17)$$

<sup>8</sup>その理由は，この結果が査読付きの (国際的な) 論文に発表されず (投稿したが，役に立たないという理由でリジェクトされた)，筆者の著書のみになら書かれているためとされた (日本語では発表している [9])．

これは、「データに誤差がないときその推定法によって真の値が得られる」ことを意味し、通常はどんな推定法でも自明に確認できる<sup>9</sup>。そもそも、これが成り立たないと推定としての意味がない。

式 (17) の一致性から次の事実が導かれる。各データ  $x_\alpha$  は  $m$  次元ベクトル、パラメータ  $u$  は  $p$  次元ベクトルであるとし、式 (6) の右辺に式 (8) を代入してテイラー展開して、式 (17) を用いると次のようになる。

$$\hat{u} = u + \sum_{\alpha=1}^N (\nabla_{x_\alpha} \hat{u}) \Delta x_\alpha + O(\varepsilon^2) \quad (18)$$

ただし  $\nabla_{x_\alpha} \hat{u}$  はその  $(ij)$  要素が  $\partial \hat{u}_i / \partial x_{j\alpha}$  の  $p \times m$  行列であり、微分は  $x_\alpha = \bar{x}_\alpha$ ,  $\alpha = 1, \dots, N$  において評価するものとする。上式から

$$(\hat{u} - u)(\hat{u} - u)^\top = \sum_{\alpha, \beta=1}^N (\nabla_{x_\alpha} \hat{u}) \Delta x_\alpha \Delta x_\beta^\top (\nabla_{x_\beta} \hat{u})^\top + \{\Delta x_\alpha\} \text{ の 3 次以上の項} \quad (19)$$

となる。両辺の期待値を取ると式 (7) の共分散行列が次のように表せる。

$$V[\hat{u}] = \varepsilon^2 \sum_{\alpha=1}^N (\nabla_{x_\alpha} \hat{u}) (\nabla_{x_\alpha} \hat{u})^\top + O(\varepsilon^4) \quad (20)$$

ここで誤差は各  $\alpha$  に対して独立であることから、式 (8) より

$$E[\Delta x_\alpha \Delta x_\beta^\top] = \delta_{\alpha\beta} \varepsilon^2 I \quad (21)$$

となることを用いた<sup>10</sup>。また式 (20) の右辺の最後の項が  $O(\varepsilon^4)$  となるのは、誤差分布の対称性から  $\Delta x_\alpha$  の奇数次の項の期待値が 0 になるためである。

## 9. KCR 下界の導出

Chernov ら [3] は筆者が文献 [9, 10] で用いたのと同じように、データの真の値  $\{\bar{x}_\alpha\}$  とパラメータ  $u$  に関する変分原理を用いて KCR 下界を導出している。 $\bar{x}_\alpha$  と  $u$  を式 (3) を満たすように  $\bar{x}_\alpha + \delta \bar{x}_\alpha$ ,  $u + \delta u$  と変分させる。式  $F(\bar{x}_\alpha + \delta \bar{x}_\alpha; u + \delta u) = 0$  より任意の変分  $\{\delta \bar{x}_\alpha\}$ ,  $\delta u$  に対して次式が成り立つ<sup>11</sup>。

$$(\nabla_x \bar{F}_\alpha, \delta \bar{x}_\alpha) + (\nabla_u \bar{F}_\alpha, \delta u) = 0 \quad (22)$$

<sup>9</sup> それに対して、統計的推定における一致性、すなわち推定量が観測数  $n$  を無限に増やせば何らかの意味で真の値に確率的に収束することを証明するのは単純ではない。

<sup>10</sup>  $\delta_{\alpha\beta}$  はクロネッカのデルタであり、 $\alpha = \beta$  のとき 1、そうでなければ 0 と約束する。

<sup>11</sup> これは通常のテイラー展開ではなく、「無限小」の変分  $\delta \bar{x}_\alpha$ ,  $\delta u$  に関する「恒等式」であるから、高次の項は現れない。力学では仮想仕事の原理と呼ばれる。 $\delta \bar{x}_\alpha$  は「真の値」 $\bar{x}_\alpha$  の「仮想的な変化」であり、観測誤差  $\Delta x_\alpha$  ではないことに注意。

ただし、ベクトル  $a, b$  の内積を  $(a, b)$  と書く。 $\nabla_x \bar{F}_\alpha$ ,  $\nabla_u \bar{F}_\alpha$  の定義は式 (10) に対するものと同じである。

推定量  $\hat{u}$  の定義式 (6) と式 (17) の一致性より  $u = \hat{u}(\bar{x}_1, \dots, \bar{x}_N)$  が恒等的に成立する。これに上記の変分を施すと、式 (22) を満たす任意の変分  $\{\delta \bar{x}_\alpha\}$ ,  $\delta u$  に対して次式が成り立つ。

$$\sum_{\alpha=1}^N (\nabla_{x_\alpha} \hat{u}) \delta \bar{x}_\alpha = \delta u \quad (23)$$

これからより次の結論が得られる。

$$\sum_{\alpha=1}^N (\nabla_{x_\alpha} \hat{u}) (\nabla_{x_\alpha} \hat{u})^\top \succ \left( \sum_{\alpha=1}^N \frac{(\nabla_u \bar{F}_\alpha)(\nabla_u \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} \right)^{-1} \quad (24)$$

これは次の補題を証明することによって得られる。

【補題 1】 0 でない  $N$  本の  $m$  次元ベクトル  $a_1, \dots, a_N$  と 0 でない  $N$  本の  $p$  次元ベクトル  $b_1, \dots, b_N$  が与えられたとする。ただし、ベクトル  $a_1, \dots, a_N$  はランクが  $m$  とする<sup>12</sup>。このとき

$$(a_\alpha, x_\alpha) + (b_\alpha, y) = 0, \quad \alpha = 1, \dots, N \quad (25)$$

が成り立つような任意の  $m$  次元ベクトル  $x_1, \dots, x_N$  と任意の  $p$  次元ベクトル  $y$  に対して、ある  $N$  個の  $p \times m$  行列  $A_1, \dots, A_N$  が存在して、

$$\sum_{\alpha=1}^N A_\alpha x_\alpha = y \quad (26)$$

が成り立つなら、次式が成り立つ。

$$\sum_{\alpha=1}^N A_\alpha A_\alpha^\top \succ \left( \sum_{\alpha=1}^N \frac{b_\alpha b_\alpha^\top}{\|a_\alpha\|^2} \right)^{-1} \quad (27)$$

この補題を Chernov ら [3] は筆者が文献 [9, 10] で用いたとほぼ同じ論法を用いて証明している。式 (24) の右辺は筆者が導いた KCR 下界 ( $\varepsilon^2$  は除いた形) にほかならない。そして式 (10) より、最尤推定量  $\hat{u}_{ML}$  の共分散行列  $V[\hat{u}_{ML}]$  が  $O(\varepsilon^4)$  を除いて下界を到達していることが結論される。

## 10. 考察と課題

Chernov ら [3] の証明では推定量の不偏性 (16) を仮定していないという意味で、筆者の理論 [9, 10] の拡張となっている。また付録 B に示す筆者の証明と比較しても筋道が明快でわかりやすい点が優れている。

<sup>12</sup> 複数のベクトルのランクとはそれらが張る部分空間の次元、すなわち線形独立なものの最大個数のことである。

る<sup>13</sup>。しかし、付録 B に示すように、筆者は共分散行列  $V[\hat{u}]$  自体に ( $\varepsilon$  に無関係に) 下界を与えたのに対して、Chernov ら [3] は  $V[\hat{u}]$  の式 (20) の  $\varepsilon$  に関する展開の第 1 項の下界を与えているのみである。一貫性があれば当然  $\varepsilon \rightarrow 0$  の極限で不偏性が成立するから、この議論は自然である。条件が弱められた分だけ結論も弱められているが、 $\varepsilon = 0$  の近傍での性質、特に最尤推定の最適性の証明にはこれで十分である。

一方、Chernov ら [3] は KCR 下界を用いて意外な事実を指摘している。それは、従来は最適でないと思われた手法も、KCR 下界の意味で最適であり得るということである。その一例は点列  $\{(x_\alpha, y_\alpha)\}$ ,  $\alpha = 1, \dots, N$  に円

$$(x - a)^2 + (y - b)^2 = R^2 \quad (28)$$

を当てはめる問題である。式 (2) の最小二乗法は次のように書ける。

$$J_{LS} = \sum_{\alpha=1}^N ((x_\alpha - a)^2 + (y_\alpha - b)^2 - R^2)^2 \rightarrow \min \quad (29)$$

これに対して、式 (4) の最尤推定は次のように書ける。

$$J_{ML} = \frac{1}{4} \sum_{\alpha=1}^N \frac{((x_\alpha - a)^2 + (y_\alpha - b)^2 - R^2)^2}{(x_\alpha - a)^2 + (y_\alpha - b)^2} \rightarrow \min \quad (30)$$

しかし、解析するところの場合には両方とも KCR の下界を満たし、その意味で両方とも最適である [3]。一般に、式 (4) を変形した

$$\tilde{J}_{ML} = \sum_{\alpha=1}^N \frac{c(\mathbf{u})F(\mathbf{x}_\alpha; \mathbf{u})^2}{\|\nabla_{\mathbf{x}}F_\alpha\|^2} \rightarrow \min \quad (31)$$

も解が KCR の下界を満たす。ただし、 $c(\mathbf{u})$  は  $\mathbf{u}$  の任意の正の関数である (付録 D 参照)。式 (28) は式 (30) の分子に  $c(a, b, R) = R$  を挿入した結果になっている。式 (30) の分母を  $R$  に置き換えても解の挙動は第 1 近似では影響を受けない。

Chernov ら [3] はシミュレーションを行い、誤差が極めて小さいときは式 (29) の最小二乗法と式 (30) の最尤推定の精度の挙動が同等であることを観察している。しかし、誤差を大きくすると一般に最尤推定が最小二乗法を上回る。ところが奇妙なことに、ある誤差以上では逆に最小二乗法が最尤推定を上回る。Chernov ら [3] はその原因が式 (28) が表現の特異性にあること指摘している<sup>14</sup>。

<sup>13</sup>ただし、補題 1 の証明は筆者の証明 [9, 10] と同じ程度の技巧が必要であり、単純ではない。

<sup>14</sup>したがって、この議論は楕円当てはめを始めとするコンピュータビジョンの代表的な問題には当てはまらず、一般には最尤推定が最小二乗より断然優れている。

## 11. まとめ

以上述べたように、KCR 下界は幾何学的当てはめ問題の最適性の最も基本的な指標である。ただし、KCR 下界の意味で最適でも高次の誤差項の影響は問題ごとに異なるので、誤差を大きくした場合の挙動は問題ごとに実験・解析する必要がある。この意味で、KCR 下界を補完する新たな指標の研究が望まれる。

最近、Mühlich ら [15] は拘束条件 (1) が線形化できる問題 (直線や楕円当てはめ、射影変換や基礎行列の計算を含む) に、平衡化 (whitening) と呼ばれる手法を拡張して、誤差が小さいときは KCR 下界の意味では最適でないが、誤差が大きいつきに、くりこみ法や HEIV 法や FNS 法 (いずれも KCR 下界の意味で最適) よりも精度が安定する (計算が破綻しない) 手法を提案している。

従来このような研究に目が向けられなかった原因は、コンピュータビジョン研究者が統計学の教科書や著名な統計学者の思想に引きずられてデータ数  $N \rightarrow \infty$  の漸近解析に目を奪われてしまったことにあると思われる。統計学の輸入ではなく、コンピュータビジョンに特化した理論解析を提唱する必要がある。Chernov ら [3] や Mühlich ら [15] の成果はそのよい例である。

謝辞: 本研究の一部は文部科学省科学研究費基盤研究 C (2) (No. 15500113) によった。

## 参考文献

- [1] 赤池弘次, 情報量基準 AIC とは何か—その意味と将来への展望, 数理科学, **153** (1976), 5–11.
- [2] 甘利俊一, 川鍋元明, 線形関係の推定—最小 2 乗法は最良であるのか?, 応用数理, **6-2** (1996-6), 96–109.
- [3] N. Chernov and C. Lesort, Statistical efficiency of curve fitting algorithms, *Comput. Stat. Data Anal.*, **47-4** (2004-11), 713–728.
- [4] W. Chojnacki, M. J. Brooks, A. van den Hengel and D. Gawley, On the fitting of surfaces to data with covariances, *IEEE Trans. Patt. Anal. Mach. Intell.*, **22-11** (2000), 1294–1303.
- [5] 遠藤利生, 鳥生隆, 田川憲男, フローからの 3 次元推定における最尤推定量が最適ではない証明, 情報処理学会研究報告 93-CVIM-86-2 (1993-11), 7–12.
- [6] T. Endoh, T. Toriu, and N. Tagawa, A superior estimator to the maximum likelihood estimator on 3-D motion estimation from noisy optical flow, *IEICE Trans. Inf. & Sys.*, **E77-D-11** (1994-11), 1240–1246.
- [7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, U.K., 2000.
- [8] 金谷健一, コンピュータビジョンのためのくりこみ法, 情報処理学会論文誌, **35-2** (1994-2), 201–209.
- [9] 金谷健一, 当てはめ問題の最適推定と精度の理論限界, 情報処理学会論文誌, **36-8** (1995-8), 1865–1873.
- [10] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam, The Netherlands, 1996.
- [11] 金谷健一, 幾何学的当てはめにおけるモデル選択, 電子情報通信学会論文誌 A, **J84-A-11** (2001-11), 1385–1393.
- [12] 金谷健一, くりこみ法その後: 波紋と発展情報処理学会研究報告 2003-CVIM-139-5 (2003-7), 33–40.

- [13] 金谷 健一, 画像からの幾何学的推論はどういう統計的モデルに基づくのか, 電子情報通信学会論文誌 D-II, **J86-D-II-7** (2003-7), 966–973.
- [14] Y. Leedan and P. Meer, Heteroscedastic regression in computer vision: Problems with bilinear constraint, *Int. J. Comput. Vision.*, **37-2** (2000), 127–150.
- [15] M. Mühlich and R. Mester, Unbiased errors-in-variables estimation using generalized eigensystem analysis, *Proc. 2nd Workshop on Statistical Methods in Video Processing*, May 2004, Prague, Czech, pp. 38–49.
- [16] 長尾淳平, 韓 太舜, かく乱母数を含む場合の MDL 基準の構築と空間図形モデル推定問題への応用, 電子情報通信学会論文誌 A, **J83-A-1** (2000), 83–95.
- [17] 太田 直哉, 栗原 祐介, かく乱母数を含まないオブティカルフローからの運動パラメータ推定, 電子情報通信学会論文誌 A, **J86-A-7** (2003-7), 772–780.
- [18] 岡谷貴之, 出口光一郎, 画像からのカメラの姿勢・3次元形状復元における推定精度の限界について, 第6回画像の認識・理解シンポジウム講演論文集, 2002年7–8月, 名古屋, pp. 335–340.
- [19] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore, 1989.
- [20] 上田修功, ベイズ学習 [I]–[IV], 電子情報通信学会誌, **85-4** (2002), 265–271; **85-6** (2002), 421–426; **85-7** (2002), 504–509. **85-8** (2002), 633–638.

#### 付録 A : 最尤推定の線形近似

式 (3) に  $\bar{x}_\alpha = x_\alpha - \Delta x_\alpha$  を代入して誤差  $\Delta x_\alpha$  が小さいと仮定して線形近似を行うと次のようになる .

$$F_\alpha - (\nabla_x F_\alpha, \Delta x_\alpha) = \mathbf{0} \quad (32)$$

この拘束条件に対するラグランジュ乗数  $\lambda_\alpha$  を導入して

$$L = \frac{1}{2} \sum_{\alpha=1}^N \|\Delta x_\alpha\|^2 + \sum_{\alpha=1}^N \lambda_\alpha (F_\alpha - (\nabla_x F_\alpha, \Delta x_\alpha)) \quad (33)$$

と置くと, 条件 (32) のもとで式 (4) を最小化する解  $\Delta x_\alpha$  は  $\nabla_{\Delta x_\alpha} L = \mathbf{0}$ ,  $\alpha = 1, \dots, N$  を満たす . これは次のように書ける .

$$\Delta x_\alpha - \lambda_\alpha \nabla_x F_\alpha = \mathbf{0} \quad (34)$$

したがって  $\Delta x_\alpha = \lambda_\alpha \nabla_x F_\alpha$  であり, 式 (32) に代入すると次のようになる .

$$F_\alpha - (\nabla_x F_\alpha, \lambda_\alpha \nabla_x F_\alpha) = \mathbf{0} \quad (35)$$

これから  $\lambda_\alpha$  が次のように得られる .

$$\lambda_\alpha = \frac{F_\alpha}{\|\nabla_x F_\alpha\|^2} \quad (36)$$

したがって式 (4) は次のように書ける .

$$\begin{aligned} J_{\text{ML}} &= \sum_{\alpha=1}^N \|\lambda_\alpha \nabla_x F_\alpha\|^2 = \sum_{\alpha=1}^N \frac{F_\alpha^2}{\|\nabla_x F_\alpha\|^4} \|\nabla_x F_\alpha\|^2 \\ &= \sum_{\alpha=1}^N \frac{F_\alpha^2}{\|\nabla_x F_\alpha\|^2} \end{aligned} \quad (37)$$

#### 付録 B : 最尤推定量の共分散行列の導出

式 (5) に式 (8), (9) を代入して展開すると,  $J_{\text{ML}}$  は次のようになる .

$$J_{\text{ML}} = \sum_{\alpha=1}^N \frac{((\nabla_x \bar{F}_\alpha, \Delta x_\alpha) + (\nabla_u \bar{F}_\alpha, \Delta u))^2}{\|\nabla_x \bar{F}_\alpha\|^2} + O(\varepsilon^3) \quad (38)$$

分母が  $\|\nabla_x \bar{F}_\alpha\|^2$  であるのは, 式 (5) の右辺の分子が  $O(\varepsilon^2)$  であるため, 分母の誤差を考慮してもその影響は剰余項  $O(\varepsilon^3)$  に吸収されるためである . 上式が最小になる  $\Delta u$  を求めれば  $u + \Delta u$  が最尤推定量  $\hat{u}_{\text{ML}}$  となる . 上式の第1項は  $\Delta u_\alpha$  の2次式であるから,  $\Delta u_\alpha$  で微分して0と置くと次式を得る .

$$\begin{aligned} &2 \sum_{\alpha=1}^N \frac{((\nabla_x \bar{F}_\alpha, \Delta x_\alpha) + (\nabla_u \bar{F}_\alpha, \Delta u)) \nabla_u \bar{F}_\alpha}{\|\nabla_x \bar{F}_\alpha\|^2} \\ &= O(\varepsilon^2) \end{aligned} \quad (39)$$

書き直すと次のようになる .

$$\begin{aligned} &\sum_{\alpha=1}^N \frac{(\nabla_u \bar{F}_\alpha)(\nabla_u \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} \Delta u \\ &= - \sum_{\alpha=1}^N \frac{(\nabla_u \bar{F}_\alpha)(\nabla_x \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} \Delta x_\alpha + O(\varepsilon^2) \end{aligned} \quad (40)$$

これから次の関係が得られる .

$$\begin{aligned} &\sum_{\alpha=1}^N \frac{(\nabla_u \bar{F}_\alpha)(\nabla_u \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} \Delta u \Delta u^\top - \sum_{\beta=1}^N \frac{(\nabla_u \bar{F}_\beta)(\nabla_u \bar{F}_\beta)^\top}{\|\nabla_x \bar{F}_\beta\|^2} \\ &= \sum_{\alpha, \beta=1}^N \frac{(\nabla_u \bar{F}_\alpha)(\nabla_x \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} \Delta x_\alpha \Delta x_\beta^\top \frac{(\nabla_x \bar{F}_\beta)(\nabla_u \bar{F}_\beta)^\top}{\|\nabla_x \bar{F}_\beta\|^2} \\ &\quad + O(\varepsilon^3) \end{aligned} \quad (41)$$

両辺の期待値を取ると,  $V[\hat{u}_{\text{ML}}] = E[\Delta u \Delta u^\top]$  であり, 式 (21) が成り立ち,  $E[O(\varepsilon^3)] = O(\varepsilon^4)$  であることから, 次式を得る .

$$\begin{aligned} &\sum_{\alpha=1}^N \frac{(\nabla_u \bar{F}_\alpha)(\nabla_u \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} V[\hat{u}_{\text{ML}}] - \sum_{\beta=1}^N \frac{(\nabla_u \bar{F}_\beta)(\nabla_u \bar{F}_\beta)^\top}{\|\nabla_x \bar{F}_\beta\|^2} \\ &= \sum_{\alpha=1}^N \frac{(\nabla_u \bar{F}_\alpha)(\nabla_x \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} \frac{(\nabla_x \bar{F}_\alpha)(\nabla_u \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} + O(\varepsilon^4) \\ &= \sum_{\alpha=1}^N \frac{(\nabla_u \bar{F}_\alpha)(\nabla_x \bar{F}_\alpha)^\top}{\|\nabla_x \bar{F}_\alpha\|^2} + O(\varepsilon^4) \end{aligned} \quad (42)$$

これから式 (10) が得られる .

#### 付録 C : KCR 下界のオリジナルな導出

筆者による KCR 下界の導出の道筋は以下の通りである [9, 10] . 式 (16) の不偏性は

$$E[\hat{u} - u] = \mathbf{0} \quad (43)$$

と書き直せる . これが式 (3) を満たす任意の  $\{\bar{x}_\alpha\}$ ,  $u$  に対して恒等的に成立しなければならない . 期待値の定義より左辺の変分は次のようになる<sup>15</sup>

$$\delta \int (\hat{u} - u) p_1 \cdots p_N dx = - \int (\delta u) p_1 \cdots p_N dx$$

<sup>15</sup>変分は  $\{\bar{x}_\alpha\}$  と  $u$  に関するものである . 推定量  $\hat{u}$  はデータ  $\{x_\alpha\}$  の関数であるから対応する変分は0である . 変分は  $\delta u$  データ  $\{x_\alpha\}$  に関係ないから積分  $\int dx$  の外に出せる . また,  $\int p_1 \cdots \delta p_\alpha \cdots p_N dx = 1$  であることに注意 .

$$\begin{aligned}
& + \sum_{\alpha=1}^N \int (\hat{u} - u) p_1 \cdots \delta p_\alpha \cdots p_N dx \\
& = -\delta u + \int (\hat{u} - u) \sum_{\alpha=1}^N (p_1 \cdots \delta p_\alpha \cdots p_N) dx \quad (44)
\end{aligned}$$

ただし,  $\int dx$  は  $\int \cdots \int dx_1 \cdots dx_N$  の略記である. 仮定よりデータ  $x_\alpha$  の確率密度は

$$p(x_\alpha) = \frac{1}{(\sqrt{2\pi})^{n_{\varepsilon^n}}} e^{-\|x_\alpha - \bar{x}_\alpha\|^2 / 2\varepsilon^2} \quad (45)$$

である. これを  $p_\alpha$  と略記した.  $\bar{x}_\alpha$  を変分させると, 上式の変分が次のように書ける.

$$\delta p_\alpha = (l_\alpha, \delta \bar{x}_\alpha) p_\alpha \quad (46)$$

ただし, スコア関数  $l_\alpha$  を次のように置いた.

$$l_\alpha \equiv \nabla_{\bar{x}_\alpha} \log p_\alpha = \frac{x_\alpha - \bar{x}_\alpha}{\varepsilon^2} \quad (47)$$

式 (43) が式 (3) を満たす任意の  $\{\bar{x}_\alpha\}$ ,  $u$  に対して恒等的に成立するから, 式 (44) は式 (22) を満たす任意の変分  $\{\delta \bar{x}_\alpha\}$ ,  $\delta u$  に対して恒等的に 0 である. 式 (46) を式 (44) に代入すると次式を得る.

$$E[(\hat{u} - u) \sum_{\alpha=1}^N l_\alpha^\top \delta \bar{x}_\alpha] = \delta u \quad (48)$$

ここで次のような変分  $\delta \bar{x}_\alpha$  を考える.

$$\delta \bar{x}_\alpha = -\frac{(\nabla_{\mathbf{x}} \bar{F}_\alpha)(\nabla_{\mathbf{x}} \bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}} \bar{F}_\alpha\|^2} \delta u \quad (49)$$

これは式 (22) を恒等的に満足することが容易に確認できる. 式 (49) を式 (48) を代入すると

$$E[(\hat{u} - u) \sum_{\alpha=1}^N m_\alpha^\top] \delta \bar{x}_\alpha = -\delta \bar{x}_\alpha \quad (50)$$

となる. ただし, 次のように置いた.

$$m_\alpha = \frac{(\nabla_{\mathbf{u}} \bar{F}_\alpha)(\nabla_{\mathbf{x}} \bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}} \bar{F}_\alpha\|^2} l_\alpha \quad (51)$$

式 (48) が式 (22) を満たす任意の変分  $\{\delta \bar{x}_\alpha\}$ ,  $\delta u$  に対して成り立つから, 特に式 (49) のように置いた変分に対しても成り立つ. したがって, 式 (50) が任意の (制約のない) 変分  $\delta u$  に対して成り立たなければならない. ゆえに, 次の関係を得る.

$$E[(\hat{u} - u) \sum_{\alpha=1}^N m_\alpha^\top] = -I \quad (52)$$

これを用いると, 式 (7) の共分散行列  $V[\hat{u}]$  の定義から, 次の関係を得る.

$$\begin{aligned}
& E\left[\begin{pmatrix} \hat{u} - u \\ \sum_{\alpha=1}^N m_\alpha \end{pmatrix} \begin{pmatrix} \hat{u} - u \\ \sum_{\alpha=1}^N m_\alpha \end{pmatrix}^\top\right] \\
& = \begin{pmatrix} V[\hat{u}] & -I \\ -I & M \end{pmatrix} \quad (53)
\end{aligned}$$

ただし, 行列  $M$  を次のように置いた.

$$\begin{aligned}
M & = E\left[\begin{pmatrix} \sum_{\alpha=1}^N m_\alpha \\ \sum_{\beta=1}^N m_\beta \end{pmatrix}^\top\right] \\
& = \sum_{\alpha, \beta=1}^N \frac{(\nabla_{\mathbf{u}} \bar{F}_\alpha)(\nabla_{\mathbf{x}} \bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}} \bar{F}_\alpha\|^2} E[l_\alpha l_\beta] \frac{(\nabla_{\mathbf{x}} \bar{F}_\beta)(\nabla_{\mathbf{u}} \bar{F}_\beta)^\top}{\|\nabla_{\mathbf{x}} \bar{F}_\beta\|^2} \\
& = \frac{1}{\varepsilon^2} \frac{(\nabla_{\mathbf{u}} \bar{F}_\alpha)(\nabla_{\mathbf{u}} \bar{F}_\alpha)^\top}{\|\nabla_{\mathbf{x}} \bar{F}_\alpha\|^2} \quad (54)
\end{aligned}$$

上式の変形において, 式 (21), (47) より  $E[l_\alpha l_\beta^\top] = \delta_{\alpha\beta} I / \varepsilon^4$  となることを用いた.  $J_\alpha \equiv E[l_\alpha l_\alpha^\top]$  は分布  $p_\alpha$  のフィッシャー情報行列であり, 各分布が独立なときは  $E[l_\alpha l_\beta^\top] = \delta_{\alpha\beta} J_\alpha$  となる.

式 (53) の左辺の期待値  $E[\cdot]$  の中身は明らかに半正値対称行列であるから, 右辺も半正値対称行列である. したがって, 次の行列も半正値対称行列である.

$$\begin{aligned}
& \begin{pmatrix} I & M^{-1} \\ & M^{-1} \end{pmatrix} \begin{pmatrix} V[\hat{u}] & -I \\ -I & M \end{pmatrix} \begin{pmatrix} I & \\ & M^{-1} \end{pmatrix} \\
& = \begin{pmatrix} V[\hat{u}] - M^{-1} & \\ & M^{-1} \end{pmatrix} \quad (55)
\end{aligned}$$

これから  $V[\hat{u}] - M^{-1}$  が半正値対称行列であること, すなわち式 (11) が成り立つことが結論される.

以上は最も単純な場合であるが, より一般の場合にも成り立つ<sup>16</sup>. 拘束条件が複数あったり, それらが独立ではなかったり, データやパラメータの自由度が拘束されている場合は一般逆行列や射影作用素を導入すればよい. 誤差は正規分布とは限らず, データごとに異なってもよい. スコア関数  $l_\alpha$  やフィッシャー情報行列  $J_\alpha$  がより複雑な形となるが, 論理は同じである.

#### 付録 D: 重みつき最小二乗法

式 (4) と比較すると式 (31) は次のように書ける.

$$\tilde{J}_{ML}(u) = c(u) J_{ML}(u) \quad (56)$$

しかし,  $c(u)$  を  $c(u + \Delta u) = c(u) + (\nabla_{\mathbf{u}} c, \Delta u) + \cdots$  としても  $J_{ML}$  が  $O(\varepsilon^2)$  であるから  $\tilde{J}_{ML}(u + \Delta u) = c(u) J_{ML}(u + \Delta u) + O(\varepsilon^3)$  である. したがって上式の微分は

$$\nabla \tilde{J}_{ML} = c(u) \nabla J_{ML} + O(\varepsilon^2) \quad (57)$$

となり,  $\nabla \tilde{J}_{ML} = 0$  の解と  $\nabla J_{ML} = 0$  の解とは  $O(\varepsilon^2)$  を除いて一致する. ゆえにそれらの共分散行列も  $O(\varepsilon^4)$  を除いて一致する. Chernov ら [3] は一般に重みつき最小二乗法

$$\tilde{J} = \sum_{\alpha=1}^N w_\alpha(x_\alpha; u) F(x_\alpha; u)^2 \rightarrow \min \quad (58)$$

が KCR 下界の意味で最適である必要十分条件が

$$w_\alpha(x_\alpha; u) = \frac{c(u)}{\|\nabla_{\mathbf{x}} F_\alpha\|^2} \quad (59)$$

であること, すなわち式 (31) の形以外には在り得ないことを証明している.

<sup>16</sup>ただし, 記号が増えて式が著しく煩雑になる. 筆者の理論 [10] が論文誌にリジェクトされたり, その成立さえ疑われた一因は, 筆者が最も一般的な状況で証明したことにあると思われる.