

しきい値を用いないオプティカルフローからの移動物体検出

太田 直哉[†] 金谷 健一[†] 木村 和弘[‡]

[†]群馬大学工学部情報工学科 [‡]CSK(株)

画像による自律走行車ではシーン中の人物や他の車両などの移動物体を検出することが重要である。本論文ではオプティカルフローから移動物体がフローの誤差の大きさに関する知識や経験的に設定するしきい値を用いることなく検出できることを示す。基本原理はフローが静止したシーンによるものであると移動物体によって生じたものであるなどのようなフローの解釈を「幾何学的モデル」とみなし、各モデルの“よさ”を「幾何学的AIC」によって相対的に比較することである。本論文ではまず一般理論を述べ、実験例を示す。本論文に示した原理は一般にしきい値処理に依存し過ぎる“知的処理”から経験的しきい値を減らすことに役立ち、実用的を高めることが期待される。

キーワード: 幾何学的AIC, 幾何学的モデル, モデル選択, オプティカルフロー, 移動物体検出, 統計的推定

Moving Object Detection from Optical Flow without Thresholding

Naoya Ohta[†], Kenichi Kanatani[†] and Kazuhiro Kimura[‡]

[†]Department of Computer Science
Gunma University, Kiryu, Gunma 376, Japan
[‡]CSK, Co. 2-6-1 Nishi-Shinjuku, Tokyo 163-02, Japan

In autonomous vehicle navigation by visual sensing, detecting objects that are moving independently of the background, such as people and other vehicles, is of utmost importance. In this paper, we show that moving objects can be detected from optical flow without using any knowledge about the magnitude of the noise in the flow or any thresholds to be adjusted empirically. The underlying principle is viewing a particular interpretation about the flow, such as being caused by a moving observer relative to a stationary scene or by an independently moving object, as a *geometric model* and comparing the relative “goodness” of candidate models measured by the *geometric AIC*. We first state a general theory and then show experiment examples. The principle proposed here is expected to play an essential role in reducing the number of empirical thresholds, thereby upgrading practicality, for “intelligent systems”, which are often overburdened with an excessive number of empirical thresholds.

Key words: geometric AIC, model selection, optical flow, moving object detection, statistical inference

[†] 376 桐生市天神町 1-5-1 群馬大学工学部情報工学科, Tel: (0277)30-1841, Fax: (0277)30-1801,
E-mail: ohta@cs.gunma-u.ac.jp, kanatani@cs.gunma-u.ac.jp

[‡] (株)CSK 〒163-02 東京都新宿区西新宿 2-6-1 新宿住友ビル TEL 03-3344-1811

1 はじめに

画像による移動ロボットの制御、特に自律走行車への応用ではシーン中の人物や他の車両のような移動物体の検出が重要な課題である。すなわち、画像中の各点の変化が、静止したシーンに相対的にカメラが移動したために生じた変化か、それともシーンに対して物体が移動したために生じた変化かを区別する必要がある。これをオプティカルフローから検出することを考える。

常識的な解決法は、ある小領域のフローがその周辺のフローと比較して著しい相違を示す場合にそれが移動物体であると判定することである。太田 [8] はフローが静止したシーンに対するカメラの移動から生じたと仮定してシーンの3次元復元を行ない、計算した奥行き値に著しい不自然さがある部分を移動物体と判定している。

より定量的には、フローは理想的には「エッジ極線拘束条件」と呼ぶ関係を満たすので、その条件からのずれがある程度以上大きい部分を移動物体と判定してもよい。例えばカメラが前方に移動している場合は静止したシーンによるフローは前方の「出現点」と呼ばれる一点から発散するように生じる。これがこの場合のエッジ極線拘束条件である。渡辺ら [11] はこのような場合に、フローの方向がその出現点の方向に比べてある程度以上隔たっている場合にその部分を移動物体と判定している。

これらの方法はすべて「観測したフローは静止したシーンに対してカメラが移動したために生じた」と仮定することに対する“不自然さ”を表す何らかの評価量を導入して、その“絶対的な大きさ”によって判定するものである。このとき判定のしきい値をどう設定すればよいのであろうか。太田 [8] や渡辺ら [11] はあらかじめ移動物体部分が既知の画像を用いて、その移動物体がよく検出できるようにしきい値を調節している。しかし、判定は精度の高いフローに対しては厳しく、精度の低いフローに対しては緩くする必要がある。

これを理論的に行なうとすれば、観測したフローの想定されるフロー（例えば仮定した出現点から発散する方向）からのずれの二乗和を計算し、それをフローの平均二乗誤差で割ったものが χ^2 分布に従うことから、ある有意水準を設定して χ^2 検定を行なえばよい [2]。しかし、フローの平均二乗誤差をどう推定すればよいのであろうか。また有意水準をどう設定すればよいのであろうか。これらはフローの大きさやそのパターン、移動物体の大きさやその3次元的位置、その移動の方向や移動速度、その他その状況を規定するもろも

ろの条件に依存するであろう。したがって、いくつかの例から試行錯誤的に経験値を設定しても、状況が少しでも変わればその値は意味をもたなくなる。

移動物体検出に限らず、今日「知的処理」と呼ばれるほとんどの処理はこのような経験的なしきい値に余りにも依存し、これが実用化を妨げる要因の一つである。ところが「研究」と称してよく行なわれるのは、このような判定をいっそう増やして複雑化する方向であり、経験値が当てはまる状況でデモを行なって高度な機能が実現されたとか人間の認識機能に近づいたなどと宣伝されている。しかし、そのような処理を真の意味で実用化するには、経験的なしきい値を少しでも減らすことが重要である。

もちろん現実の状況は複雑であるから、最終的には適当なしきい値によって調節する必要があるが、状況のある程度モデル化し、その状況では判定のしきい値が理論的に定まることが望ましい。そのようなしきい値を基準にし、モデル化し切れない要因に対処するために最終的な調節を行なうべきであり、始めから経験的なしきい値に頼るのはあるべき姿ではない。

本論文ではオプティカルフローから何らのしきい値も用いず、かつフローの精度に関する知識を必要とすることなく移動物体を検出する理論的な枠組みを提案する。その基本原理はフローが静止したシーンによるものであるとか移動物体によって生じたものであるなどのようなフローの解釈を幾何学的モデルとみなし、各モデルの“よさ”を評価する基準を導入し、そのよさを相対的に比較する。モデルのよさを評価する基準としてはAICやMDLが知られているが、単純な方程式で記述される幾何学的問題ではAICのほうが統計的な意味が明確であり、かつ定式化が単純である。ただし問題の性格上、通常の（赤池の）AIC [10] はそのままでは適用できないので、それを変形した幾何学的AIC [4, 5] を用いる。以下ではまず一般理論を述べた後に移動物体検出の実験例を示す。

2 フローのモデルと幾何学的AIC

オプティカルフローのモデルとは「誤差がなければ満たすであろうと想定される条件」をいう。オプティカルフローから3次元が復元できるのは何らかのモデルを仮定するからである。レンズの中心を原点とし、光軸方向をZ軸とするカメラ座標系をとり、焦点距離を長さの単位とすると、画像面は $Z = 1$ と同一視できる。そのような単位での画像座標 (x, y) の点に速度 (\dot{x}, \dot{y}) のフローが与えられているとき、これらをそ

それぞれ次のベクトルで表す。

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \quad \dot{\mathbf{x}} = \begin{pmatrix} \dot{x} \\ \dot{y} \\ 0 \end{pmatrix} \quad (1)$$

カメラの瞬間的な並進速度を \mathbf{v} 、回転速度を $\boldsymbol{\omega}$ とし、 $\{\mathbf{v}, \boldsymbol{\omega}\}$ を運動パラメータと呼ぶ。点 $\{\mathbf{x}_\alpha\}$, $\alpha = 1, \dots, N$ にフロー $\{\dot{\mathbf{x}}_\alpha\}$ が与えられているとする。このフローは誤差を含み、真の値 (未知) は $\{\bar{\mathbf{x}}_\alpha\}$ であるとする。

シーンが静止しているとき代表的なモデルとして次のものが考えられる。

一般モデル S : 運動パラメータ $\{\mathbf{v}, \boldsymbol{\omega}\}$ には特に制約がない。

このとき誤差がなければフローは次のエビ極線方程式を満たす¹⁾[4]。

$$|\mathbf{x}_\alpha, \bar{\mathbf{x}}_\alpha, \mathbf{v}| + (\mathbf{v} \times \mathbf{x}_\alpha, \boldsymbol{\omega} \times \mathbf{x}_\alpha) = 0 \quad (2)$$

一般並進モデル S_{gt} : カメラは回転せず ($\boldsymbol{\omega} = \mathbf{0}$) に並進している。

このとき誤差がなければフローは次の式を満たす。

$$|\mathbf{x}_\alpha, \bar{\mathbf{x}}_\alpha, \mathbf{v}| = 0 \quad (3)$$

特定並進モデル S_{st} : カメラは回転せず ($\boldsymbol{\omega} = \mathbf{0}$)、特定方向 \mathbf{v}^0 へ並進している。

このとき誤差がなければフローは次の式を満たす。

$$|\mathbf{x}_\alpha, \bar{\mathbf{x}}_\alpha, \mathbf{v}^0| = 0 \quad (4)$$

フロー $\dot{\mathbf{x}}_\alpha$ の共分散行列を $\epsilon^2 V_0[\dot{\mathbf{x}}_\alpha]$ とする。 $V_0[\dot{\mathbf{x}}_\alpha]$ はフロー検出時に定数倍を除いて求まるフローの共分散行列であり [7, 6, 8]、正規化共分散行列と呼ぶ。 ϵ は精度の絶対的な大きさを表す未知の定数であり、ノイズレベルと呼ぶ。上の各モデルのもとでの3次元復元とその幾何学的 AIC は次のようになる [4, 5]。

一般モデル: 運動パラメータ $\{\mathbf{v}, \boldsymbol{\omega}\}$ の最尤推定量は次の関数を最小にするものである²⁾。

$$J[\mathbf{v}, \boldsymbol{\omega}] = \sum_{\alpha=1}^N \frac{|\mathbf{x}_\alpha, \dot{\mathbf{x}}_\alpha, \mathbf{v}| + (\mathbf{v} \times \mathbf{x}_\alpha, \boldsymbol{\omega} \times \mathbf{x}_\alpha)^2}{(\mathbf{v}, (\mathbf{x}_\alpha \times V_0[\dot{\mathbf{x}}_\alpha] \times \mathbf{x}_\alpha) \mathbf{v})} \quad (5)$$

¹⁾ $|\cdot, \cdot, \cdot|$ はベクトルのスカラー三重積を表す。

²⁾ 式 (5) は \mathbf{v} の定数倍に不変であるから、任意に (例えば $\|\mathbf{v}\| = 1$) と正規化する。

その最小値を \hat{J} と置くと、幾何学的 AIC は次のようになる。

$$AIC(S) = \hat{J} + 2(N+5)\epsilon^2 \quad (6)$$

第1近似において \hat{J}/ϵ^2 が自由度 $N-5$ の χ^2 分布に従うから、二乗ノイズレベルの不偏推定量が次のように得られる。

$$\hat{\epsilon}^2 = \frac{\hat{J}}{N-5} \quad (7)$$

一般並進モデル: 並進速度 \mathbf{v} の最尤推定量は次の関数を最小にするものである。

$$J_{gt}[\mathbf{v}] = \sum_{\alpha=1}^N \frac{|\mathbf{x}_\alpha, \dot{\mathbf{x}}_\alpha, \mathbf{v}|^2}{(\mathbf{v}, (\mathbf{x}_\alpha \times V_0[\dot{\mathbf{x}}_\alpha] \times \mathbf{x}_\alpha) \mathbf{v})} \quad (8)$$

その最小値を \hat{J}_{gt} と置くと、幾何学的 AIC は次のようになる。

$$AIC(S_{gt}) = \hat{J}_{gt} + 2(N+2)\epsilon^2 \quad (9)$$

第1近似において \hat{J}_{gt}/ϵ^2 が自由度 $N-2$ の χ^2 分布に従うから、二乗ノイズレベルの不偏推定量が次のように得られる。

$$\hat{\epsilon}^2 = \frac{\hat{J}_{gt}}{N-2} \quad (10)$$

特定並進モデル: 未知数が含まれていないので

$$\hat{J}_{st} = \sum_{\alpha=1}^N \frac{|\mathbf{x}_\alpha, \dot{\mathbf{x}}_\alpha, \mathbf{v}^0|^2}{(\mathbf{v}^0, (\mathbf{x}_\alpha \times V_0[\dot{\mathbf{x}}_\alpha] \times \mathbf{x}_\alpha) \mathbf{v}^0)} \quad (11)$$

と置くと、幾何学的 AIC は次のようになる。

$$AIC(S_{st}) = \hat{J}_{st} + 2N\epsilon^2 \quad (12)$$

これによって次のようにモデルが比較できる。

一般モデル vs. 一般並進モデル:

一般モデル S からノイズレベル ϵ^2 が式 (7) によって推定できる。これを代入すると $K_{gt} = \sqrt{AIC(S_{gt})/AIC(S)}$ は次のようになる。

$$K_{gt} = \sqrt{\frac{N-5}{3N+5} \left(\frac{\hat{J}_{gt}}{\hat{J}} + \frac{2(N+2)}{N-5} \right)} \quad (13)$$

$K_{gt} < 1$ であれば一般並進モデルが選ばれる。

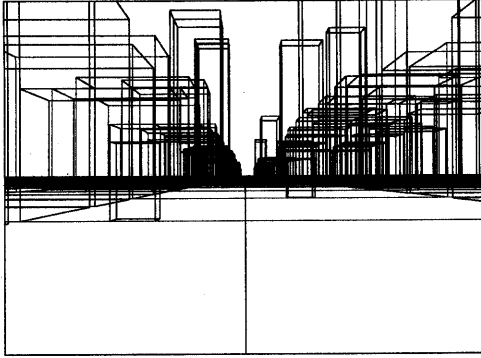


図 1: 3次元シーンのモデル。

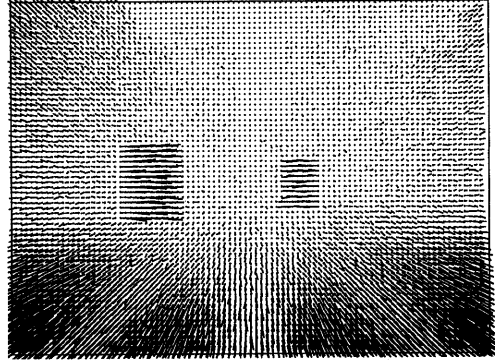


図 2: オプティカルフロー。

一般並進モデル vs. 特定並進モデル:

一般並進モデルからノイズレベル ϵ^2 が式 (10) によって推定できる。これを代入すると $K_{st} = \sqrt{AIC(S_{st})/AIC(S_{gt})}$ は次のようになる。

$$K_{st} = \sqrt{\frac{N-2}{3N+2} \left(\frac{\hat{J}_{st}}{\hat{J}_{gt}} + \frac{2N}{N-2} \right)} \quad (14)$$

$K_{st} < 1$ であれば特定並進モデルが選ばれる。

3 移動物体の検出

上記の理論は移動物体検出に次のように応用できる。カメラがある向きに車に固定され、車が前方に進行しているとす。シーンが静止しているとき観測されるのは誤差がなければ特定並進フローである。画像中のある領域に移動する物体があり、移動物体は(回転せずに)並進している。その領域で観測されるのは誤差がなければ一般並進フローである。そこでウィンドウを移動しながら画像上を走査し、ウィンドウ内のフローに対して一般並進モデルと特定並進モデルとを比較し、特定並進モデルが選ばれたら「移動物体は存在しない」と判定する(すなわち「移動物体の非存在の検出」である)。

これには式 (14) を計算し、 $K_{st} < 1$ であるかどうかを調べればよい。しかし、これは統計的な判定であり、種々の不確定要因があることを考えると、機械的にどちらかに判定するよりも K_{st} の値自身を「移動物体非存在の確からしさ」とみなすのが現実的である(K_{st} が 1 より小さいほど非存在の程度が高くなる)。

図 1 は走行中 (40km/h で直進) の車から前方に見える都市風景のワイヤーフレーム表示である (640 × 480 画素、焦点距離 600 画素、画角 56.1° × 43.6°)。そこには静止したシーン中を 6 個の直方体物体が車の進行

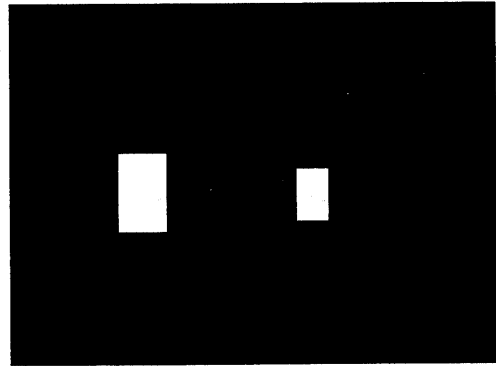


図 3: 移動物体領域 (白) と背景 (黒)。

方向に直交する方向に移動している (4km/h)。図 2 は理論的に計算したオプティカルフローの各方向に平均 0、標準偏差 0.2 画素の正規乱数で誤差を加えたものである。フローは 7 画素おきに定義する。ただし見やすくするために長さを 1.5 倍して表示している。図 3 は移動物体のある領域を白、ない領域を黒とした 2 値画像である。

図 4 はフローの定義された 5 × 5 画素を含む正方形ウィンドウを画像上で移動させながら計算した K_{st} の値をそのウィンドウの中心の画素の値とし、 $K_{st} = 0.95$ を黒、 $K_{st} = 2$ を白として濃淡表示したものである。 K_{st} を計算するには式 (8) の $J_{gt}[v]$ を最小にする v を計算しなければならないが、これには「くりこみ法」と呼ぶ手法 [9] に計算を安定化させる手法を付加したものをを用いた (付録参照)。条件 $K_{st} \geq 1$ によって機械的に移動物体を判定した結果が図 5 である。

上記の計算には経験的に設定するしきい値を何も含んでいないし、フローの誤差の大きさに関する何の知識も用いていない。それでもこのように移動物体の存在 / 非存在に関する情報が得られる。もちろん実際の

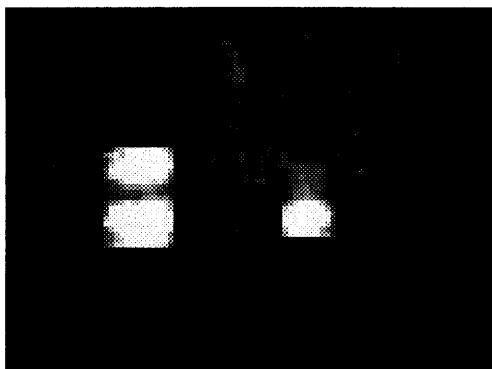


図 4: K_{st} の濃淡表示。



図 5: 条件 $K_{st} \geq 1$ による移動物体の判定。

応用に際してはフローの精度や移動物体の性格などのより具体的な知識や情報を付加すればさらに精度が向上するであろう。また、経験的なしきい値による処理を導入すればさらに能力が向上するであろう。例えば上記の例では図 4 と図 5 を比較してわかるように、実用上は理論的しきい値 1 よりも多少大きい値を用いたほうが好ましい結果が得られる。しかし、それらは 2 次的な調節であり、上記の実験は少なくとも基本的には精度の知識やしきい値処理なしに判定できることを示している。

4 まとめ

本論文ではオプティカルフローから移動物体が、少なくとも原理的にはフローの誤差の大きさに関する知識や経験的に設定するしきい値を用いることなく検出できることを示した。基本原理はフローが静止したシーンによるものであるとか移動物体によって生じたものであるなどのようなフローの解釈を「幾何学的モデル」とみなし、各モデルの“よさ”を「幾何学的 A

IC」によって相対的に比較することである。本論文ではその一般理論を述べ、実験例を示した。このような考え方は、一般にしきい値処理に依存し過ぎる「知的処理」から経験的しきい値を減らすことに役立ち、実用的を高めることが期待される。

参考文献

- [1] K. Daniilidis and H.-H. Nagel, Analytical results on error sensitivity of motion estimation from two views, *Image Vis. Comput.*, **8-4** (1990), 287-303.
- [2] 金谷健一, 画像の 3 次元解釈の統計的信頼性, 情報処理学会論文誌, **34-10** (1993), 2062-2070.
- [3] 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), 1865-1873.
- [4] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, 1996.
- [5] 金谷健一, 情報量基準による幾何学的モデル選択, 情報処理学会論文誌, **37-6** (1996), 1073-1080.
- [6] 木村和広, 太田直哉, 金谷健一, 精密なノイズモデルによるオプティカルフローの検出, 情報処理学会コンピュータビジョン研究会, 96-CV-99-6 (1996-3).
- [7] N. Ohta, Image movement detection with reliability indices, *IEICE Trans. Inf. & Syst.*, **E74-D-10** (1991), 3379-2288.
- [8] 太田直哉, 信頼性情報を持ったオプティカルフローからの形状復元とその移動物体検出への応用, 電子情報通信学会論文誌, **76-D-II-8** (1993), 1562-1571.
- [9] N. Ohta and K. Kanatani, Optimal structure from motion algorithm for optical flow, *IEICE Trans. Inf. & Syst.*, **E78-D-12** (1995), 1559-1566.
- [10] 坂本慶行, 石黒真木夫, 北川源四郎, 「情報量統計学」, 共立出版, 1983.
- [11] 渡辺陸, 武田信之, 小野口一則, 主成分分析を用いた移動物体認識法の検討, 情報処理学会研究報告, 96-CV-99 (1996-3).

付録: 特異点をもつくりこみ法の安定化

拘束条件 $F^{(k)}(a, u) = 0, k = 1, \dots, L$ が与えられたとき、誤差のあるデータ $\{a_\alpha\}, \alpha = 1, \dots, N$ から未知パラメータ u を推定する問題を考える。オプティ

カルフローからの3次元復元もこのクラスに属する。その最尤推定は一般に次の形の評価関数の最小化となる[4]。

$$J[\mathbf{u}] = \sum_{\alpha=1}^N \sum_{k,l=1}^L W_{\alpha}^{(kl)} F_{\alpha}^{(k)} F_{\alpha}^{(l)} \quad (15)$$

ここに $F_{\alpha}^{(k)}$ は $F^{(k)}(\mathbf{a}_{\alpha}, \mathbf{u})$ の略記であり、 $W_{\alpha}^{(k,l)}$ は \mathbf{a}_{α} と \mathbf{u} から定まる重み関数である。式(5), (8), (11)はいずれもこの形である。

多くの問題では $F^{(k)}(\mathbf{a}, \mathbf{u})$ が \mathbf{u} に関して線形である。このときもし重み関数 $W_{\alpha}^{(k,l)}$ が \mathbf{u} に依存しない定数であれば評価関数 $J[\mathbf{u}]$ は \mathbf{u} の2次形式となるから、固有値問題を解いて解が定まる。しかし一般に $W_{\alpha}^{(k,l)}$ が \mathbf{u} に依存するので、まず \mathbf{u} を何らかの方法で推定し、 $W_{\alpha}^{(k,l)}$ にそれを代入して定数とみなして $J[\mathbf{u}]$ を最小化し、その解を用いてこれを反復することが考えられる。しかし、そのようにしたのでは統計的な偏差が生じて精度のよい解が得られない。

「くりこみ法」はその反復の過程でその偏差を推定してはそれを差し引いて補正するという発想に基いている。そして例外的な場合を除いて数回の反復で精度の高い解が効率よく計算されることが確かめられている。その例外的な場合というのはあるデータ \mathbf{a}_{α} が存在して、 \mathbf{u} の真の値 $\bar{\mathbf{u}}$ と \mathbf{a}_{α} の真の値 $\bar{\mathbf{a}}_{\alpha}$ とに対して $W_{\alpha}^{(kl)} = \infty$ となる場合である。そのようなデータ \mathbf{a}_{α} を特異データと呼ぶ[3]。

データには誤差があるため特異データであっても ∞ の発散は実際には生じないが、くりこみ法の反復の過程で不安定な振動が生じることがある。直線やコニックなどの曲線当てはめでは無限遠点が特異データとなるので有限範囲には特異データは存在しないが、オプティカルフローでは「出現点」がそれに当たる。出現点が最尤推定の特異点となることは以前から多くの研究者によって指摘されてきた(例えば Danilidis ら[1])。したがって画像中に出現点が現われる場合(カメラがほぼ前進または後退する場合)にこの現象が生じる。

これを防ぐ有効な方法は、まず \mathbf{u} を適当な方法で推定し(例えば式(15)で $W_{\alpha}^{(kl)} = 1$ として固有値問題を解き)、それを代入して式(15)の右辺のデータに関する総和の各項 $\sum_{k,l=1}^L W_{\alpha}^{(kl)} F_{\alpha}^{(k)} F_{\alpha}^{(l)}$ を計算し、これがある値より大きくなる α に対するデータ \mathbf{a}_{α} を用いないことである。その大きさは次のように評価できる。データに含まれる誤差の大きさを表すノイズレベル ϵ は $J[\mathbf{u}]/\epsilon^2$ が自由度 $rN - n'$ の χ^2 分布に従うことから推定できる(r は拘束条件のランク、 n' はパラメータ空間の自由度[4])。これによって \mathbf{u} の初期推定値から ϵ を推定すると、 $\sum_{k,l=1}^L W_{\alpha}^{(kl)} F_{\alpha}^{(k)} F_{\alpha}^{(l)}/\epsilon^2$ が

ほぼ自由度 r の χ^2 分布に従うから、

$$\sum_{k,l=1}^L W_{\alpha}^{(kl)} F_{\alpha}^{(k)} F_{\alpha}^{(l)} > \epsilon^2 \chi_{r,\alpha}^2 \quad (16)$$

なるものを除去すればよい。ただし $\chi_{r,\alpha}^2$ を自由度 r の χ^2 分布の上側 $\alpha\%$ 点であり、大きいほうの約何%を除去するかを指定するものである。

この方法は式(15)を次のように書き換えることに相当している。

$$J[\mathbf{u}] = \sum_{\alpha=1}^N \rho_{\epsilon^2 \chi_{r,\alpha}^2} \left(\sum_{k,l=1}^L W_{\alpha}^{(kl)} F_{\alpha}^{(k)} F_{\alpha}^{(l)} \right) \quad (17)$$

ここに $\rho_{\epsilon^2 \chi_{r,\alpha}^2}(x)$ は $x \leq c$ なら $\rho(x) = x$, $x > c$ なら $\rho(x) = 0$ とする関数である。一般に最適化計算の安定化手法はロバスト推定と呼ばれ、特にこのような何らかの関数 $\rho(x)$ を用いる方法はM推定と呼ばれている。上述の方法による特異データの除去もM推定の変種とみなすことができる。