

# 最小2乗法とロバスト推定法

———直線推定の例を通じてその本質を学ぶ———

## 1. はじめに

画像処理には図1に示したように様々な応用があります。そういった応用において基本的な機能として円や直線を精度良く検出しなければならない場面に遭遇することは多いでしょう。

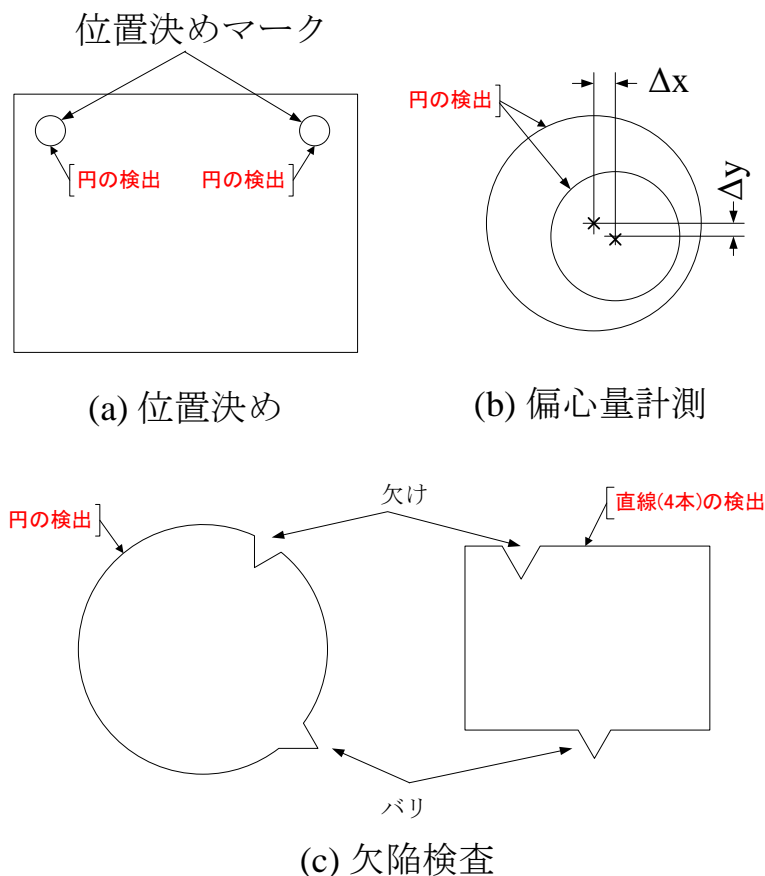


図1 画像処理の応用例

ご存じのように円は3点の座標が判れば決定出来ますし、直線は2点で決定出来ます。しかし、もしそのうちの1点においてたまたま精度が悪かった場合、円や直線の精度が悪くなってしまいうでしょう(図2)。その際、なるべく多くの点を使って円や直線を算出できれば、何らかの事情で多少ずれてもその他の点が頑張ってくれるので精度の悪化は軽減出来ます。円は4点以上の点があった場合は全体の点をなるべくほどよく加味しながらそれらしい箇所に決定する必要があります。その決定法として、得られたデータに**偏りが無く**、その**ばらつきが正規分布**をしているような場合に用いられる理論が**最小2乗法**であり、データに**偏りがあっても**比較的高い精度を出せる理論が**ロバスト推定法**なのです。

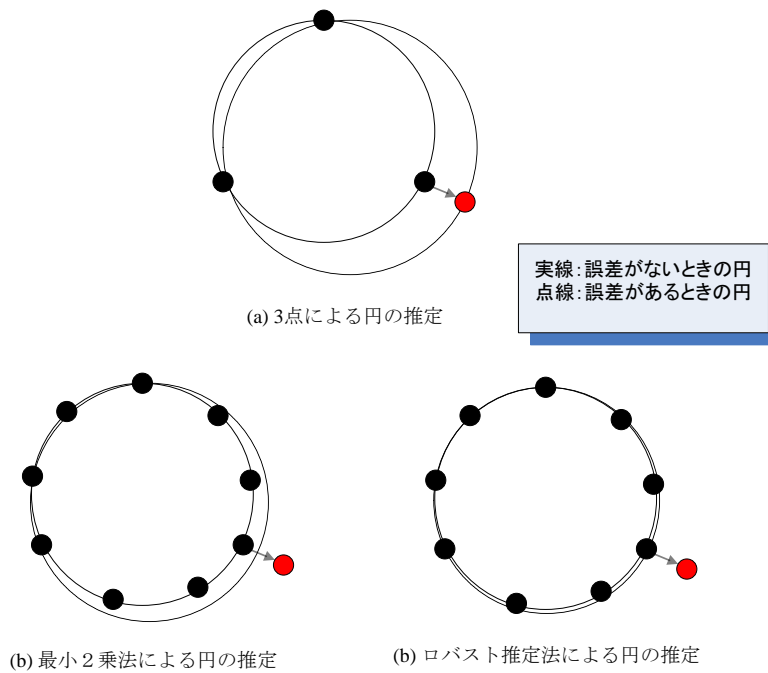


図2 円を推定する際の各種手法

実はいきなり円を対象に説明しますと式の複雑さ(円は陰関数のためです)に気を取られて本質を見失うので、ここでは最も簡単な直線を基軸に話を進めていきましょう。基本原理さえ理解できれば複雑なものへの対応はそれほど時間を要しません。よって、本内容は次のような要望に応えるものになっています。

カメラ座標系  $\Sigma_v$  上にある3点以上の座標が  $(x_i, y_i), i=1 \sim n$  で与えられたときに、それらの座標を最良に近似する直線  $y = ax + b$  を推定したい。

前述したように、この問題は最小2乗法によって解くことが出来ます。最小2乗法による直線推定はこの誤差を最小化するかにより、**回帰分析による解法**と**主成分分析による解法**の2つがあります。図3にそれらを示しますが、図から判るように誤差  $\epsilon_i$  の取り方が異なります。ただ、何れにしても(1)式に示した誤差の2乗和  $R$  を定義してこれが最小になるような直線を求める事が目的になります。言うまでもないことですが、この2つの解法によって推定された直線は異なります。

$$R = \sum_{i=1}^n \epsilon_i^2 \tag{1}$$

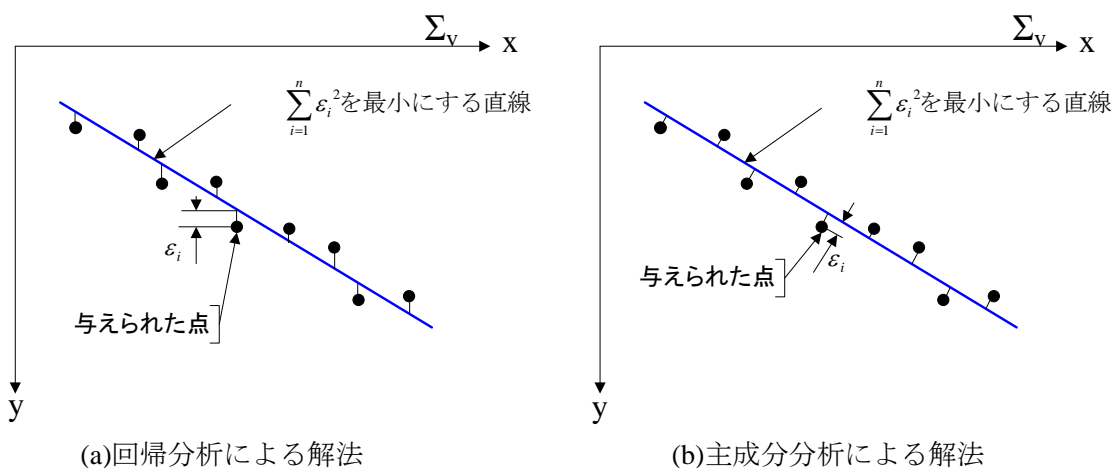


図3 最小2乗法による直線推定

$y$  軸方向に比べて  $x$  軸方向の精度が十分に良い場合(例えば音声処理では  $x$  軸は精度の良い時間軸なのでこれに当てはまります)は回帰分析による解法、 $x$  軸方向と  $y$  軸方向が同程度の精度の場合(例えば画像処理は主にこちらです)は主成分分析による解法が一般的です。

本来、画像処理用途なら主成分分析による解法の説明が良いのですが、行列の固有値や固有ベクトルといった概念が出てくるため、やはり本質を見失いやすいので、最小2乗法の原理を素直に表現している回帰分析による解法で説明します(主成分分析による解法は添付資料1に載せておきます)。以下、判りやすさを優先して数学的厳密性は多少犠牲にして説明させていただきます。

## 2. 最小2乗法による直線推定(回帰分析による解法)

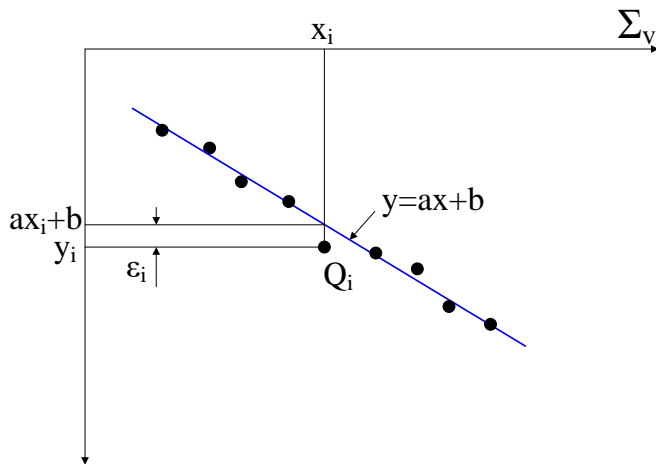


図4 回帰分析による解法(記号の定義)

以下に最小2乗法の理論を説明します。ちなみに、最小2乗法の専門書等では重みを加味したもので説明をしています。それは各データの**ばらつき(分散)がデータ毎に異なる**ことを前提にそのばらつきの値を重みにしたものです。ここでは説明を簡単にするために各データのばらつきは同じと仮定して、重みを考慮していないもので説明します(実用上はこれで何ら問題はありません)。なお、この重みは後述するロバスト推定法で出てくる重みとは**本質的に異なる**ものです。

点 $Q_i$ における誤差を $\epsilon_i$ としたとき、 $\epsilon_i$ は(2)式で表されます。

$$\epsilon_i = y_i - (ax_i + b) \quad \text{-----(2)}$$

誤差の2乗和 $R$ は

$$R = \sum_{i=1}^n \epsilon_i^2 \quad \text{-----(1)再記}$$

となりますので、これに(2)式を代入すると(3)式になります。

$$R = \sum_{i=1}^n (y_i - (ax_i + b))^2 \quad \text{-----(3)}$$

$$= \left( \sum_{i=1}^n x_i^2 \right) a^2 + \left( \sum_{i=1}^n 1 \right) b^2 - 2 \left( \sum_{i=1}^n x_i y_i \right) a - 2 \left( \sum_{i=1}^n y_i \right) b + 2 \left( \sum_{i=1}^n x_i \right) ab + \sum_{i=1}^n y_i^2 \quad \text{---(4)}$$

(4)式は $a, b$ のみの関数になっていますので、これを $R = f(a, b)$ と考え、更に $a^2, b^2$ の係数が必ず正になることから図5のようなグラフになります。

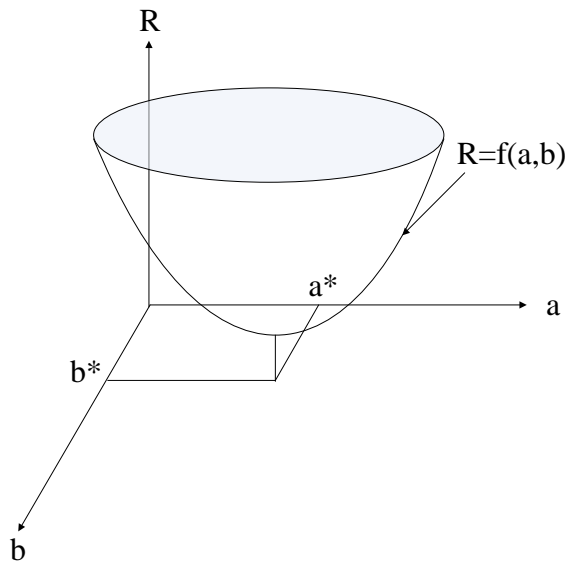


図5 直線のパラメータと誤差の関係

$R$  を最小にする  $a, b$ 、つまり  $a^*, b^*$  を見つけることにより直線を推定します。具体的には  $a, b$  の偏微分係数を求めて  $\frac{\partial R}{\partial a} = 0, \frac{\partial R}{\partial b} = 0$  とし、それらを連立させて求めれば良いことになります。なお、 $\frac{\partial R}{\partial a}, \frac{\partial R}{\partial b}$  の概念は添付資料2で説明しています。

$$\frac{\partial R}{\partial a} = 2\left(\sum_{i=1}^n x_i^2\right)a + 2\left(\sum_{i=1}^n x_i\right)b - 2\left(\sum_{i=1}^n x_i y_i\right) = 0 \quad \text{----- (5)}$$

$$\frac{\partial R}{\partial b} = 2\left(\sum_{i=1}^n x_i\right)a + 2\left(\sum_{i=1}^n 1\right)b - 2\left(\sum_{i=1}^n y_i\right) = 0 \quad \text{----- (6)}$$

(5)式～(6)式は

$$\begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix} \quad \text{----- (7)}$$

となるので、求めたい  $a^*, b^*$  は(8)式となります。

$$\begin{pmatrix} a^* \\ b^* \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n 1 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix} \quad \text{----- (8)}$$

なお、以上の偏微分係数を利用した算出法は最小2乗法概念を理解するのに役に立ちますが、実際の応用には汎用性の高い**一般逆行列**を利用したものを使う場合が多いので、そちらも簡単に説明しておきます。

直線は次のように内積の形に変形することが出来ます。

$$y = ax + b \quad \text{-----(9)}$$

$$= (x \ 1) \begin{pmatrix} a \\ b \end{pmatrix} \quad \text{-----(10)}$$

(10)式に  $(x_i, y_i), i=1 \sim n$  を代入して現れた式を行列およびベクトルを使ってまとめると、

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \quad \text{-----(11)}$$

となります。これを

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}, \vec{p} = \begin{pmatrix} a \\ b \end{pmatrix} \quad \text{-----(12)}$$

とおくと、

$$\vec{y} = X\vec{p} \quad \text{-----(13)}$$

となります。ここで、

$$\vec{p} = X^{-1}\vec{y} \quad \text{-----(14)}$$

としたいところですが、計算に使う座標が3点以上あると  $X$  は正方行列ではなくなり、逆行列が定義できません。そこで、正方行列にするため両辺に  $X^T$  を掛けます。

$$X^T\vec{y} = X^T X\vec{p} \quad \text{-----(15)}$$

これで  $X^T X$  は正方行列(ここでは2行2列)になり、逆行列が定義できます。よって、

$$\vec{p} = (X^T X)^{-1} X^T \vec{y} \quad \text{-----(16)}$$

となり、直線のパラメータ  $\vec{p}$  が計算できます。なお、 $T$  は転置の意味です(以下同)。言うまでもないことですが、(16)式を具体的に展開すると(8)式と同じ式になります。

以上、最小2乗法を用いた直線推定に関して説明しましたが、最小2乗法の応用は広く、**円・楕円・平面・2次曲線・アフィン変換**等いろいろなものがあります。その他、 $m$  個のパラメータ  $p_i$  と変数  $q_i$  ( $q_i$

は  $x^2$  や  $\log x$  でも構いません)、それと定数  $c$  に関して  $\sum_{i=1}^m p_i q_i = c$  のように表現できものは全て最小2乗

法で解くことができます。中には一見複雑過ぎて無理と思われる式でも工夫することにより解くことが出来るものもあります。

例えば、 $z = x^a y^b$  はそのままでは解けそうもないのですが、両辺に対数を取ることで、

$$\log z = a \log x + b \log y \quad \text{-----(17)}$$

と変形できます。ここで、

$$Z = \log z, X = \log x, Y = \log y \quad \text{-----(18)}$$

と置き直せば、

$$Z = aX + bY \quad \text{-----(19)}$$

となり、最小2乗法による解法が可能になるわけです。実際、こういったことを見出したり考え出したりすることが腕の見せ所かもしれません。

### 3. 最小2乗法の問題点

強力で応用範囲の広い最小2乗法ですが、一つ問題があります。それはデータに偏りが無いということを前提としているため、全ての点を同じ**寄与度**(直線推定をする際の計算に点が**影響する度合い**)で計算していることです。そのため、何らかの事情により誤差の大きな点が混入してしまうと、推定する直線がその点に引っ張られてしまうということになります。図6にその一例を示します。本来なら青線を推定結果としたいところですが、点Pの存在により赤線の位置にずれてしまっています。

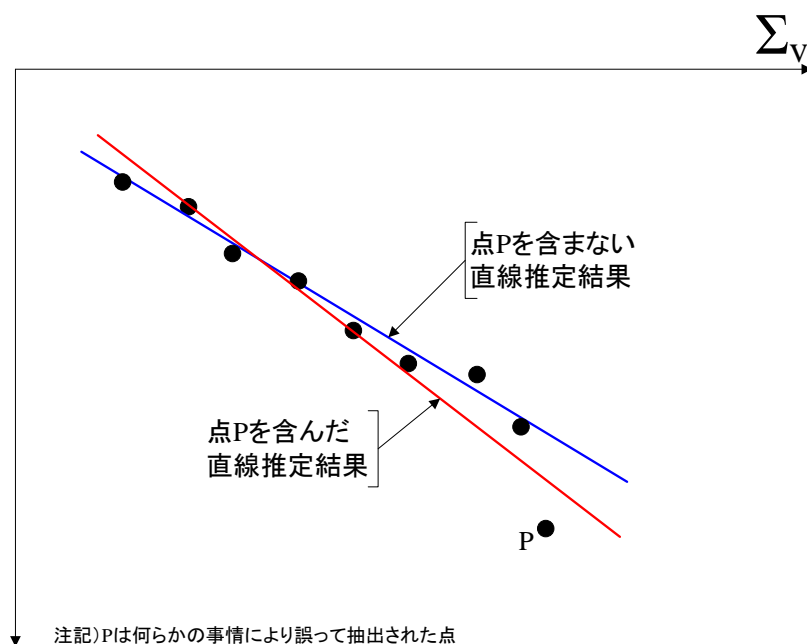


図6 最小2乗法の問題点

これは点Pを含んだ全点の寄与度を個別に制御できれば解決出来ます。つまり、全ての点を同じ寄与度で直線推定に反映させるのではなく、各点に対して寄与度を定量化した**重み**を考慮して計算すればよいわけです(最小2乗法の章でも述べましたが、この重みは最小2乗法そのものの理論で現れるデータのばらつきから定義される重みとは異なります)。これがロバスト推定法の基本原理の1つになります。



#### 4. ロバスト推定法による直線推定

ロバスト推定法は最小2乗法に重みの概念を組み込んだものですが、それに伴いもう1つの機能が必要になってきます。最小2乗法のように1回の計算で算出したのでは重みを計算の中に取り入れることが出来ません。詳細は後述しますが、重みを調整しながら**反復処理**させ、より良い結果に収束させていくことで重みを上手に生かしています。**重み付き反復最小2乗法**がロバスト推定法の別名といっても良いでしょう。

前述したように、最小2乗法は(20)式に示した連立方程式を解いています。

$$\begin{cases} y_1 = ax_1 + b \\ y_2 = ax_2 + b \\ \vdots \\ y_n = ax_n + b \end{cases} \quad \text{-----(20)}$$

それに対して、ロバスト推定法は各々の式に重みを掛けた(21)式を連立させて解くことになります。

$$\begin{cases} w_1 y_1 = w_1 (ax_1 + b) \\ w_2 y_2 = w_2 (ax_2 + b) \\ \vdots \\ w_n y_n = w_n (ax_n + b) \end{cases} \quad \text{-----(21)}$$

つまり、重み  $w_1, w_2, \dots, w_n$  を考慮しながら最小2乗法を反復して計算します。重みは次のように算出します。

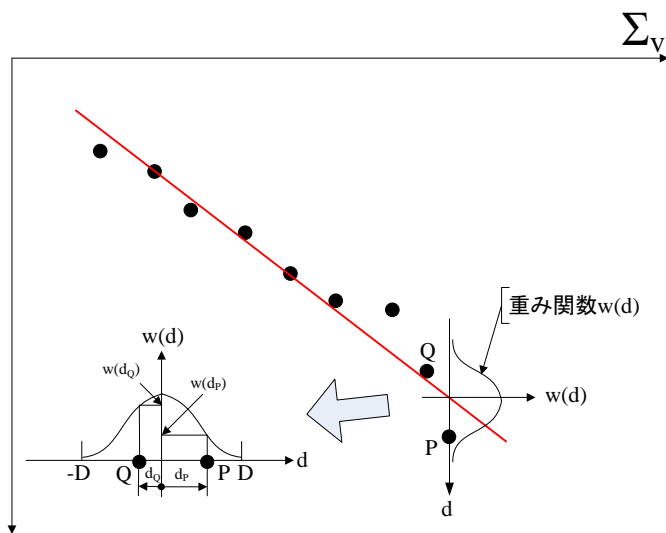


図7 重み関数

各点の重みを算出するのに(22)式～(23)式に示した**重み関数  $w(d)$** を導入します。この関数は、その時点までに推定した直線(図7の例では赤線)の位置を中心とした関数になります(重み関数はここで述べた以外にも種々のものが定義されています)。 $D$ は外部より入力されるパラメータです。画像のS/Nにより  $D$  を制御します。もちろん、「S/N→大」なら「 $D$ →小」です。ちなみに理論上  $D$  を無限大にすると最小2乗法と**等価**になります。

$$w(d) = \begin{cases} 1 - \left(\frac{d}{D}\right)^2 & |d| \leq D \\ 0 & |d| > D \end{cases} \quad \text{-----(22)}$$

$$\text{-----(23)}$$

本例では点Pの重みは $w(d_p)$ 、点Qの重みは $w(d_q)$ となります。 $w(d_p) < w(d_q)$ ですから、点Pより点Qの方が高い寄与度を持つこととなります。全ての点においてこの重みを算出し、この重みを使ってロバスト推定法が実現されます。

まず、 $\bar{y}, X, \bar{p}, {}^{(k)}W$ を次のように定義します。

$$\bar{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}, \bar{p} = \begin{pmatrix} a \\ b \end{pmatrix}, {}^{(k)}W = \begin{pmatrix} {}^{(k)}w_1 & 0 & 0 & 0 \\ 0 & {}^{(k)}w_2 & \ddots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & {}^{(k)}w_n \end{pmatrix} \quad \text{-----(24)}$$

これにより、(21)式は(25)式のように表記されます。

$${}^{(k)}W\bar{y} = {}^{(k)}WX\bar{p} \quad \text{-----(25)}$$

(25)式を変形すると、

$$X^T {}^{(k)}W\bar{y} = X^T {}^{(k)}WX\bar{p} \quad \text{-----(26)}$$

$$\bar{p} = (X^T {}^{(k)}WX)^{-1} X^T {}^{(k)}W\bar{y} \quad \text{-----(27)}$$

となるので、 $k=1 \sim K$ の範囲で反復計算をすれば、 $\bar{p}$ を求めることが出来ます。図8に(27)式を使った処理フローを示します。

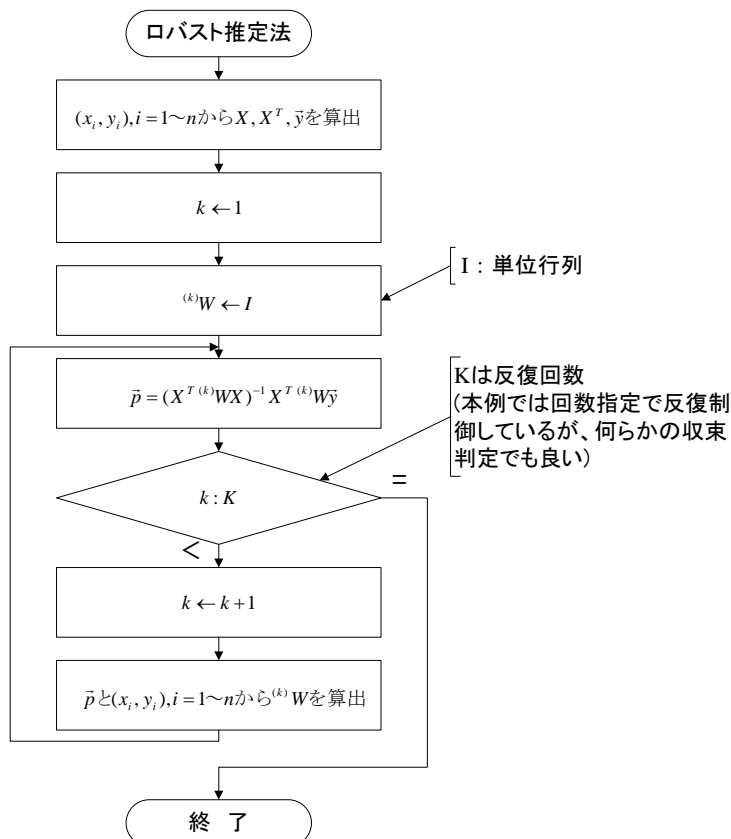


図8 ロバスト推定法の処理フロー

この処理により、目的とする直線に徐々に収束していきます(図 9)。

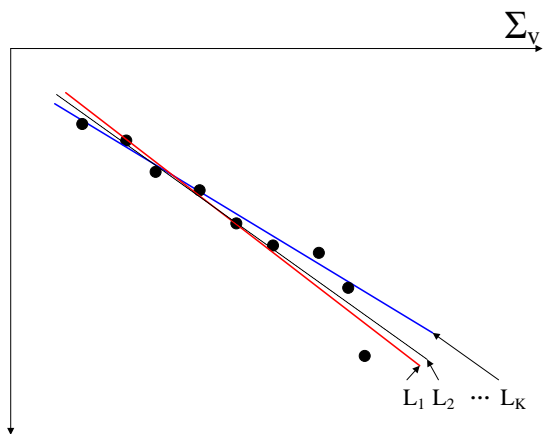


図 9 ロバスト推定法における収束過程

ここで、記号の意味は以下の通りです。

$L_1$ : 通常の最小 2 乗法により推定された直線(全点の重みを 1.0 として計算)

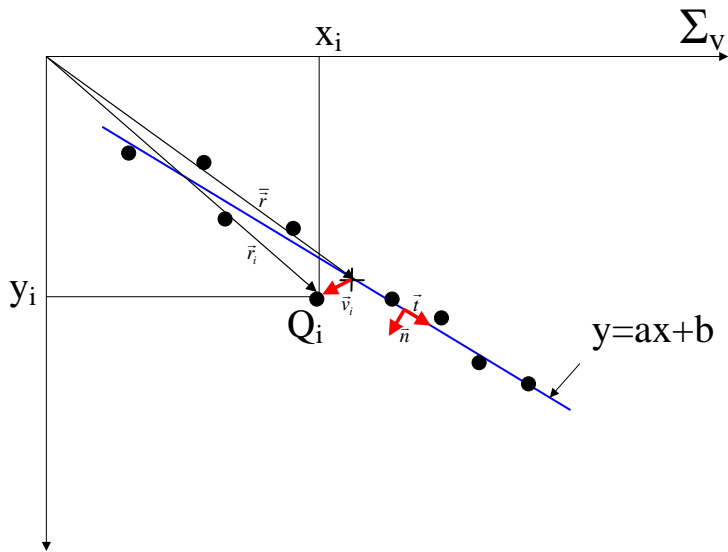
$L_2$ : 各点毎の重みを考慮した重み付き最小 2 乗法により推定された直線

.....

$L_K$ : 重み付き最小 2 乗法を繰り返して最終的に推定された直線

以 上

添付資料1 最小2乗法による直線推定(主成分分析による解法)



付図1 主成分分析による解法(記号の定義)

推定したい直線の方角ベクトルを  $\vec{t}$  とします。

$$\vec{t} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad \text{ただし } \|\vec{t}\| = 1 \quad \text{----- (付1)}$$

よって、法線ベクトル  $\vec{n}$  は  $\vec{t}$  を  $\frac{\pi}{2}$  回転させて、

$$\vec{n} = \begin{pmatrix} \cos \frac{\pi}{2} & -\sin \frac{\pi}{2} \\ \sin \frac{\pi}{2} & \cos \frac{\pi}{2} \end{pmatrix} \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad \text{----- (付2)}$$

$$= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad \text{----- (付3)}$$

$$= \begin{pmatrix} -t_y \\ t_x \end{pmatrix} \quad \text{----- (付4)}$$

となります。本方法では、直線は与えられた点群の重心を通ることが理論的に保証されていますので、あとは傾きである  $\vec{n}$  を求めることで直線を推定します。

点  $Q_i$  の位置ベクトル  $\vec{r}_i$  を

$$\vec{r}_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \quad \text{----- (付5)}$$

とした時、その平均ベクトル  $\vec{r}$  は

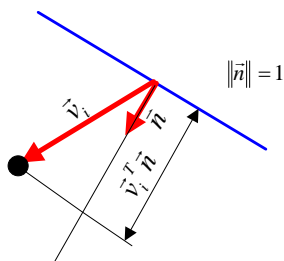
$$\vec{r} = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} x_i \\ y_i \end{pmatrix} \quad \text{----- (付6)}$$

となります。これらから、

$$\vec{v}_i = \vec{r}_i - \bar{\vec{r}} \quad \text{-----} \quad \text{(付7)}$$

で $\vec{v}_i$ を定義して、さらに誤差 $\varepsilon_i$ を(付8)式で定義します。

$$\varepsilon_i = \vec{v}_i^T \vec{n} \quad \text{-----} \quad \text{(付8)}$$



付図2  $\vec{v}_i^T \vec{n}$ の意味

誤差の2乗和 $R$ は

$$R = \sum_{i=1}^n \varepsilon_i^2 \quad \text{-----} \quad \text{(付9)}$$

$$= \sum_{i=1}^n (\vec{v}_i^T \vec{n})^2 \quad \text{-----} \quad \text{(付10)}$$

$$= \sum_{i=1}^n (\vec{v}_i^T \vec{n})^T (\vec{v}_i^T \vec{n}) \quad \text{-----} \quad \text{(付11)}$$

$$= \sum_{i=1}^n (\vec{n}^T \vec{v}_i) (\vec{v}_i^T \vec{n}) \quad \text{-----} \quad \text{(付12)}$$

$$= \sum_{i=1}^n \vec{n}^T (\vec{v}_i \vec{v}_i^T) \vec{n} \quad \text{-----} \quad \text{(付13)}$$

$$= \vec{n}^T \left( \sum_{i=1}^n (\vec{v}_i \vec{v}_i^T) \right) \vec{n} \quad \text{-----} \quad \text{(付14)}$$

となります。ここで、

$$V = \sum_{i=1}^n \vec{v}_i \vec{v}_i^T \quad \text{-----} \quad \text{(付15)}$$

とおくと、

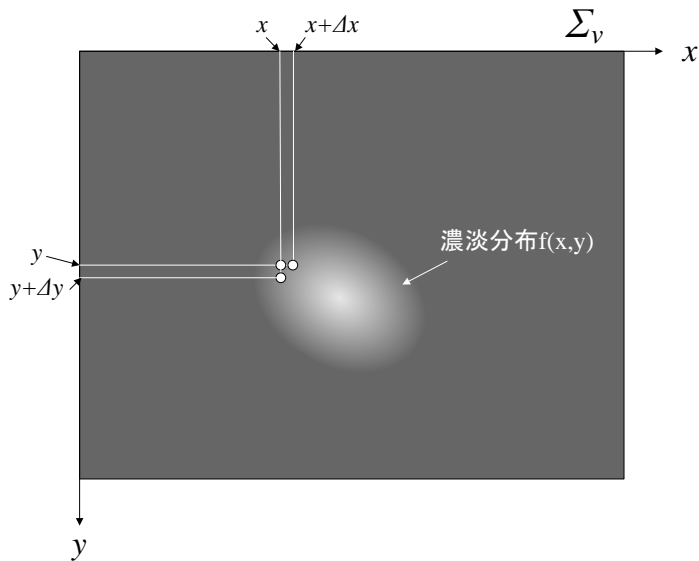
$$R = \vec{n}^T V \vec{n} \quad \text{-----} \quad \text{(付16)}$$

となります。線形代数の定理より、 $R$ が最小となる $\vec{n}$ は $V$ の最小固有値 $\lambda_{\min}$ に対応する固有ベクトル $\vec{u}_{\min}$ となりますので、これから $\vec{n}$ が求められます。ちなみに、最大固有値 $\lambda_{\max}$ に対応する固有ベクトル $\vec{u}_{\max}$ は $\vec{t}$ となります。

以上

## 添付資料2 偏微分係数 $\frac{\partial f}{\partial x}$ , $\frac{\partial f}{\partial y}$ の概念

カメラ座標系  $\Sigma_v$  上に濃淡分布  $f(x, y)$  があるとき、任意の位置  $(x, y)$  と  $x$  の近傍  $\Delta x$ 、 $y$  の近傍  $\Delta y$  を下図の通り定義します。



$\frac{\partial f}{\partial x}$  の定義は次式となります。  $\frac{df}{dx}$  との違いは  $x$  以外の変数を全て定数と見なすことです。

$$\frac{\partial f}{\partial x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

しかし、画像処理の世界では最小単位が1画素なので  $\Delta x \rightarrow 0$  ではなく  $\Delta x \rightarrow 1$  となります。つまり、

$$\begin{aligned} \frac{\partial f}{\partial x} &= \lim_{\Delta x \rightarrow 1} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x} \\ &= f(x + 1, y) - f(x, y) \end{aligned}$$

となります。

同様に  $\frac{\partial f}{\partial y}$  は、

$$\frac{\partial f}{\partial y} = f(x, y + 1) - f(x, y)$$

となります。

以上