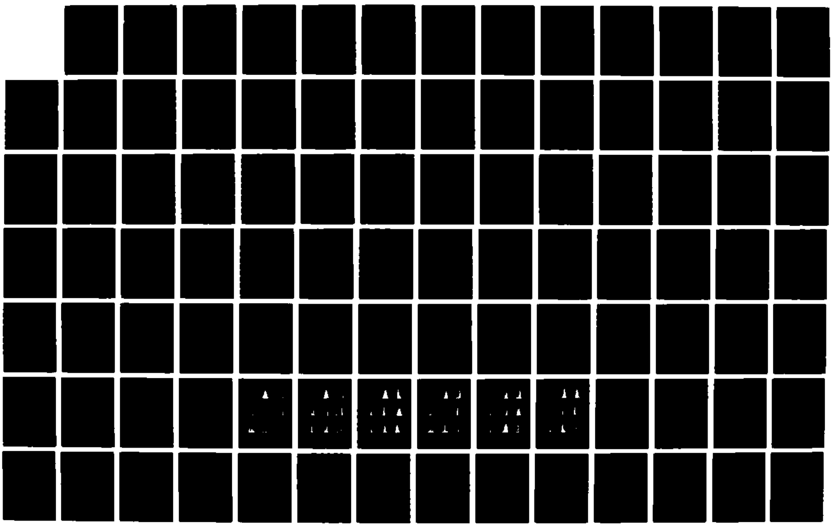
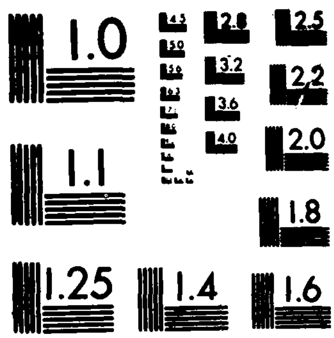
AD-A183 034 THE ESTIMATION OF A RIGID BODY MOTION IN THE PRESENCE OF NOISE(U) HARVARD UNIV CAMBRIDGE MASS DEPT OF STATISTICS C H PARK 31 JUL 87 TR-ONR-C-1 1/2

UNCLASSIFIED N00014-86-K-0246 F/G 14/4 NL




1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

AD-A183 834

1b. RESTRICTIVE MARKINGS

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

3. DISTRIBUTION / AVAILABILITY OF REPORT
Unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
TR-NO. ONR-C-1

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
Department of Statistics
Harvard University

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION

6c. ADDRESS (City, State, and ZIP Code)
Dept. of Statistics, SC713
Harvard University
Cambridge, MA 02138

7b. ADDRESS (City, State, and ZIP Code)

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
ONR

8b. OFFICE SYMBOL *(If applicable)*
CODE 1111

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
N00014-86-K-0246

8c. ADDRESS (City, State, and ZIP Code)
Office of Naval Research
Arlington, VA 22217-5000

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*
The Estimation of a Rigid Body Motion in the Presence of Noise

12. PERSONAL AUTHOR(S)
Chang Hoon Park

13a. TYPE OF REPORT
Technical Report

13b. TIME COVERED
FROM ______ TO ______

14. DATE OF REPORT (Year, Month, Day)
July 31, 1987

15. PAGE COUNT
15 105

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
IMAGE PROCESSING MOTION ESTIMATION

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

SEE REVERSE SIDE

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Herman Chernoff

22b. TELEPHONE *(Include Area Code)*
617-495-5462

22c. OFFICE SYMBOL

19. ABSTRACT

The problem of estimating a rigid body motion from two noisy images of an object taken at two different times is studied.

The available data consist of the unordered locations of some of the prominent points of the object. Because these points are not individually recognized and some of them may be missed in one or both of the images, it is not obvious which points of the two images correspond to one another. Moreover the observed locations are subject to error.

A computational procedure, capitalizing on the rigidity of the object, is proposed for estimating the motion parameters of the object in the presence of Gaussian noise $N(o,\sigma^2 I_2)$ independently added to the positions in the images of the points observed.

In principle, one might apply maximum likelihood to the estimation problem, but the difficulty in formulating and calculating the likelihood function under the above mentioned assumptions if formidable. However one ought to expect to do better when the observed points are recognized and the common points among them are matched without error in identification. Hence this latter situation, which is readily treated by maximum likelihood, should provide us with lower bounds for the error in our estimates for the original problem.

Asymptotic normality and consistency of the maximum likelihood estimate as $\sigma \to 0$ are derived for this favorable situation. Similar results hold even when σ is incorrectly assumed to be $\sigma_a = k\sigma$ for positive $k \neq 1$.

A simulation study has been carried out to compare the efficiency of the proposed estimator for the more complex problem with that of the maximum likelihood estimate in the favorable situation and to test the robustness of the proposed estimator under the misspecification of the value of σ — Sigma

HARVARD UNIVERSITY

DEPARTMENT OF STATISTICS

THE ESTIMATION OF A RIGID BODY MOTION IN THE PRESENCE OF NOISE

BY

CHANG HOON PARK

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Technical Report No. ONR-C-1

September, 1987

THE ESTIMATION OF A RIGID BODY MOTION IN THE PRESENCE OF NOISE

by

CHANG HOON PARK

Submitted to the Department of Mathematics on July 31, 1987
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Mathematics

ABSTRACT

The problem of estimating a rigid body motion from two noisy images of an object taken at two different times is studied.

The available data consist of the unordered locations of some of the prominent points of the object. Because these points are not individually recognized and some of them may be missed in one or both of the images, it is not obvious which points of the two images correspond to one another. Moreover the observed locations are subject to error.

A computational procedure, capitalizing on the rigidity of the object, is proposed for estimating the motion parameters of the object in the presence of Gaussian noise $\mathcal{N}(0, \sigma^2 \mathbf{I}_2)$ independently added to the positions in the images of the points observed.

In principle, one might apply maximum likelihood to the estimation problem, but the difficulty in formulating and calculating the likelihood function under the above mentioned assumptions is formidable. However one ought to expect to do better when the observed points are recognized and the common points among them are matched without error in identification. Hence this latter situation, which is readily treated by maximum likelihood, should provide us with lower bounds for the error in our estimates for the original problem.

Asymptotic normality and consistency of the maximum likelihood estimate as $\sigma \to 0$ are derived for this favorable situation. Similar results hold even when σ is incorrectly assumed to be $\sigma_a = k\sigma$ for positive $k \neq 1$.

A simulation study has been carried out to compare the efficiency of the proposed estimator for the more complex problem with that of the maximum likelihood estimate in the favorable situation and to test the robustness of the proposed estimator under the misspecification of the value of σ.

Key Words: Image Processing, Motion Estimation
AMS:]980 Subject Classification: Primary-8699; Secondary-62N99

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest thanks to my advisor, Herman Chernoff, who, from the very beginning of my study at M.I.T., has been a consistent source of encouragement and support. He has been literally a father-like figure to me with his thoughtfulness and direction.

I am also very grateful to Richard Dudley, who has given many important suggestions and careful corrections toward the final version of this thesis. Greta Ljung has served as the third member of my thesis committee and deserves my word of thanks. I should also mention Bill DuMouchel for his help and concern while he was here at M.I.T.

Among my colleagues, past and present, at the Statistics Center, I would especially like to thank Tom Lane, Guy Manuel, Irvin Schick, Dong Kim, Lisa Newton and Alan Zaslavsky, most of whom I have bothered so many times with non-trivial questions. Nick Lange also has suffered many times for the same reason and I am grateful to him.

Outside the Statistics Center, I must thank all my friends, especially Soonchul Lee, for the good times and laughs we shared. I wish him the best luck in his future.

I also would like to acknowledge the financial support from M.I.T. as a research assistant to Herman Chernoff and Richard Dudley during my stay at M.I.T.

All these years, my parents, parents-in-laws, brothers, sister, and their families have shown me their never-ending support both in the United States and back in Korea. In particular, without the sacrificing support from my parents and parents-in-laws during the critical period last year, I could never have finished my thesis.

Finally, but not least, I would like to thank my wife Hye Kyung and my son Sukhyun for their love, encouragement, endurance, and everything during the last five difficult years. I hope I can become a better husband and father from now on.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction and Summary

The problem of estimating a rigid body motion from two consecutive image frames of an object taken at two different times is an important issue in image sequence analysis. For example, motion estimation cannot be separated from the problem of image matching and image registration, which in turn, have a wide range of direct applications. Consequently, there is a sizable amount of literature dealing with the subject, especially for three dimensional rigid body motion.

A large number of the existing methods for the three-dimensional problem basically depend on solving simultaneous non-linear equations that one gets from the known matched points in the two image frames. In the event of noisy images, least squares methods are applied by simply requiring a greater number of matched points than that of the parameters to be estimated. However, they all produce rather unsatisfactory results when the angle of rotation of the motion is not small, or when they are applied to real images (*i.e.*, noisy images).

In this thesis, an estimation method for two dimensional motion, capitalizing on the rigidity of the object involved with the assumption that the images are noisy, is proposed. It is assumed that the data consist of the unordered locations of some of the prominent points of the object. Because it is assumed here that these points are not individually recognized and some of them may be missed in one or both of the image frames, it is not obvious which observed points in the two image frames correspond to one another. Moreover it is also assumed that $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_2)$ Gaussian noise is independently added to the locations of the points observed in the two image frames. Methods for this two dimensional problem should prove to be helpful in attacking the more practical three dimensional problem where some points may become occluded.

In Chapter 2, a survey of the existing literature on motion estimation and the related matching problems is presented. In Chapter 3, the details of the proposed estimation procedure and the logic behind it are given. Then, in Chapter 4, the maximum likelihood estimate in a more favorable situation, where the observed points are recognized and the common points among them are matched without error in identification, is studied. Asymptotic normality and consistency of the maximum likelihood estimate as $\sigma \to 0$ for this latter situation are derived. Also similar results are derived even when σ is incorrectly assumed to be $\sigma_a = k\sigma$ for positive $k \neq 1$. These provide us with lower bounds for the error in our estimates for the original problem. Chapter 5 shows some simulation results. Finally Chapter 6 has conclusions and suggestions for further work.

In the next section, we list some definitions and notation used throughout.

1.2 Definitions and Notation

Let $\mathbf{X}$ be an m-dimensional real vector valued random variable. We denote the *law* (*distribution*) of $\mathbf{X}$ by $\mathcal{L}[\mathbf{X}]$. Sometimes we use the notation $\mathcal{L}[\mathbf{X} \mid \boldsymbol{\beta}]$ instead of $\mathcal{L}[\mathbf{X}]$ to emphasize the fact that the distribution of $\mathbf{X}$ is dependent on the parameter $\boldsymbol{\beta}$. We write $\mathrm{E}[\mathbf{X}]$, and $\mathrm{Cov}[\mathbf{X}]$ to denote the mean vector, and covariance matrix of $\mathbf{X}$ respectively. In case $\mathbf{X}$ is a 1-dimensional real random variable, $\mathrm{Cov}[\mathbf{X}]$ is reduced to the variance $\mathrm{Var}[\mathbf{X}]$.

A Gaussian (normal) distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

We interpret any vector to be a column vector when it is referred to only by its name in equations or formulae unless otherwise specified.

We denote by $\mathbf{R}^m$, the m-dimensional Euclidean real vector space, and use $(-\pi, \pi]$ to denote the set $\{u \in \mathbf{R}^1 \mid -\pi < u \leq \pi\}$.

The notation ":=" is to be read "by definition is equal to".

The m by m identity matrix is denoted by $\mathbf{I}_m$:

$$\mathbf{I}_m := \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{1.1}$$

We denote by $\mathbf{1}_m$, the m-dimesional vector with m 1's as its components:

$$\mathbf{1}_m := \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbf{R}^m. \tag{1.2}$$

For any matrix $\mathbf{A}$, we use $\mathbf{A}^T$ to denote the transpose of $\mathbf{A}$.

For any p_2 by q_2 matrix $\mathbf{A}$ and p_1 by q_1 matrix $\mathbf{B}$, we define the *direct product* or *Kronecker product* of $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{A} \otimes \mathbf{B}$, to be the $p_1 p_2$ by $q_1 q_2$ matrix $\mathbf{C}$ where

$$\mathbf{C} := \begin{bmatrix} \mathbf{A}b_{11} & \mathbf{A}b_{12} & \cdots & \mathbf{A}b_{1q_1} \\ \mathbf{A}b_{21} & \mathbf{A}b_{22} & \cdots & \mathbf{A}b_{2q_1} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}b_{p_1 1} & \mathbf{A}b_{12} & \cdots & \mathbf{A}b_{p_1 q_1} \end{bmatrix}, \tag{1.3}$$

where b_{ij} is the (i,j)-th element of $\mathbf{B}$, for $i = 1, \ldots, p_1$, $j = 1, \ldots, q_1$.

For any vector $\mathbf{x} := (x_1, \ldots, x_m)^T \in \mathbf{R}^m$, we use $\|\mathbf{x}\|$ to denote the Euclidean norm (length) of $\mathbf{x}$:

$$\|\mathbf{x}\| := \sqrt{\sum_{i=1}^{m} x_i^2} = \sqrt{\mathbf{x}^T\mathbf{x}}. \tag{1.4}$$

We denote the *unit circle* by $\mathbf{S}^1$:

$$\mathbf{S}^1 := \{(x,y) \in \mathbf{R}^2 \mid \|(x,y)\| = 1\}, \tag{1.5}$$

and define the function $u : \mathbf{R}^2 \setminus \{0\} \longrightarrow \mathbf{S}^1$ by

$$u(x,y) := \frac{(x,y)}{\|(x,y)\|} \tag{1.6}$$

for each $(x,y) \in \mathbf{R}^2 \setminus \{0\}$.

The *covering map on* $\mathbf{S}^1$, $c^* : \mathbf{S}^1 \longrightarrow (-\pi, \pi]$, assigns to each $\mathbf{x} \in \mathbf{S}^1$, a $c^*(x) \in (-\pi, \pi]$ where

$$(\cos c^*(\mathbf{x}), \sin c^*(\mathbf{x})) = \mathbf{x}. \tag{1.7}$$

We will also define the *covering map on* $\mathbf{R}^1$, $c : \mathbf{R}^1 \longrightarrow (-\pi, \pi]$, by

$$c(x) := c^*(\cos x, \sin x), \tag{1.8}$$

for each $x \in \mathbf{R}^1$. Note that for any $x \in \mathbf{R}^1$, we have

$$c(x) \equiv x \pmod{2\pi}. \tag{1.9}$$

The function arctan : $\mathbf{R}^2 \setminus \{0\} \longrightarrow (-\pi, \pi]$ is defined by

$$\arctan(x, y) := c^* u(x, y) \tag{1.10}$$

for each $(x, y) \in \mathbf{R}^2 \setminus \{0\}$. We call $\arctan(x, y)$, the *angle* of the vector (x, y). Note that for any $(x, y) \in \mathbf{R}^2 \setminus \{0\}$, we have that

$$\tan(\arctan(x, y)) = \frac{y}{x} \tag{1.11}$$

and

$$\arctan(x, y) \neq \arctan(-x, -y). \tag{1.12}$$

For any two vectors v_1, $v_2 \in \mathbf{R}^2$, the *difference vector* of v_1 and v_2 denoted by $D(v_1, v_2)$ is defined by

$$D(v_1, v_2) := v_2 - v_1. \tag{1.13}$$

For a function $t : \mathbf{R}^m \longrightarrow \mathbf{R}^d$, where

$$t(v) := \begin{bmatrix} t_1(v) \\ \vdots \\ t_d(v) \end{bmatrix}, \tag{1.14}$$

we define the *Jacobian matrix* or *derivative* of t with respect to $\boldsymbol{v}$, when it exists, to be the d by m matrix $\partial t(\boldsymbol{v})/\partial \boldsymbol{v}$ whose (i,j)-th element is given by

$$\left[\frac{\partial t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right]_{ij} := \frac{\partial t_i(v_1,\ldots,v_m)}{\partial v_j} \quad \text{for } i=1,\ldots,d,\ j=1,\ldots,m, \tag{1.15}$$

where $\boldsymbol{v} = (v_1,\ldots,v_m)$.

The 2 by 2 orthogonal matrix representing the rotation by θ radians about the origin in a plane is denoted by $\mathbf{U}(\theta)$:

$$\mathbf{U}(\theta) := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{1.16}$$

We use $\boldsymbol{T}$ to denote the vector of translation by T_x, T_y in the x- and y-coordinates respectively, *i.e.*,

$$\boldsymbol{T} := \begin{bmatrix} T_x \\ T_y \end{bmatrix}. \tag{1.17}$$

A *rigid body motion* of an object in a plane is a transformation of the object that can be obtained through a rotation and a translation only. Thus a general rigid body motion can be uniquely characterized in terms of a rotation about the origin followed by a translation in the plane. In other words, a rigid body motion is a map

$$(x,y) \longmapsto (x',y')$$

for any point (x,y) on the object, where

$$\begin{bmatrix} x' \\ y' \end{bmatrix} := \mathbf{U}(\theta) \begin{bmatrix} x \\ y \end{bmatrix} + \boldsymbol{T} \tag{1.18}$$

for some uniquely defined $\mathbf{U}(\theta)$ and $\boldsymbol{T}$.

An *image frame* is a two-dimensional plane that contains an image of the object in which we are interested, where some globally fixed (for different image frames) Euclidean coordinates, which we call the *image frame coordinates*, are superimposed.

A *location* (*position*) of a point in an image frame is the image frame coordinate vector of the point.

We will treat an object as represented by a finite set of points on it, called *prominent* points. In the simulations in Chapter 5, we will assume that there are 10 prominent points. A prominent point is said to be *observed* if we see the point in an image frame but cannot necessarily distinguish which of the several prominent points it is. Also a prominent point is said to be *recognized* if we see the point and can identify which of the several prominent points it is. The noisy locations of the observed points on the object form the data in our problem.

Chapter 2

Related Literature

There are two separate approaches in motion estimation problems. One of them, the so called *"low level"* or *"signal processing"* approach, is the one where you have grey level images as your data and you try to estimate the motion by operating on the raw data. One of the methods in that approach is the *Fourier method* for the two-dimensional motion problem, where you take advantage of the fact that sharp straight edges give rise to line spectra in the frequency domain, and hence allow you to estimate the angle of rotation. Another low level method is the *matching* or *correlation method* that sets up a *cost* function depending on the grey level functions of pixels in each image frame and tries to find the value of the motion parameters which minimizes the cost function. There is another method called the *method* of *differentials* in the low level approach which uses the idea of relating time differences to spatial differences. All of the three methods mentioned above are described in [Huang 81].

The other approach is the *"high level"* or *"feature based"* approach, where we assume that some features, e.g., points, lines, contours, etc., in the images have already been extracted by some other means and the data consist of the locations of the features. Normally it is assumed that the correspondence between the features has already been established in the feature extraction process. In this sense, the estimation method to be proposed in this thesis is somewhere between the two approaches but closer to the feature based approach because we have as data the unordered observed locations of some of the prominent points in two image frames but they are not recognized and may be missing in one of the two image frames.

The high level methods may then be further classified into two classes, namely the *"equation solvers"* and the *"merit score* or *weight maximizers"*. The equation solvers basically assume that the correspondence between features in each image frame are known and hence derive sets of simultaneous equations, usually non-linear, with motion parameters as unknowns. Researchers in this line of approach include Ullman [Ullman 79], Roach and Aggarwal [Roach 79,Roach 80], Nagel [Nagel 81a], Nagel and Neumann [Nagel 81b], Tsai and Huang [Tsai 81,Tsai 84], Tsai, Huang and Zhu [Tsai 82], Fang and Huang [Fang 83a,Fang 83b]. Methods for solving such equations are generally iterative and require good initial guesses of unknowns. However, sensitivity to noise is shown in the experiments reported by Fang and Huang [Fang 83a,Fang 83b].

The merit score maximizers often bear the names of "*point pattern matching*" or "*image matching*" instead of motion estimation. They generally deal with two-dimensional motion only. Even though they usually assume that the first image frame contains the prototype or model pattern with which the observed point pattern in the second is to be matched, they essentially attack the same estimation problem as mentioned in Section 1.1. In these settings, as the name "matching" says, the individual correspondence between features in different image frames are not assumed. They implicitly or explicitly assign merit scores or weights to possible matches of features according to their compatibility with the candidate value of motion parameters, and then try to maximize the total merit scores to get the maximizing value of motion parameters. Our method to be proposed in this thesis belongs to this class but is not concerned with explicit matching. We list some of the existing methods that fall into this class below.

Simon, Checroun and Roche [Simon 72] compute all interpoint distances in each point pattern and then use a comparison of the sorted list of these with some relaxation rule to match the point patterns. Their method is applicable only to patterns that contain equal numbers of points, i.e., there are no missing prominent points.

Seidl [Seidl 74] measures similarity of point patterns, without seeking an explicit matching, using nearest neighbor relations. Again the method is restricted to patterns that contain equal numbers of points in application.

Zahn [Zahn 74] compares the minimal spanning trees of point patterns, but this method is also sensitive to missing and extra points.

Kahl, Rosenfeld and Danker [Kahl 80] only consider small ($\leq 10°$) angles of rotation. Even though their method is sensitive to noise exceeding a preset level, it is insensitive to missing or extra points.

Ranade and Rosenfeld [Ranade 80] consider the situation where only translation is allowed. Each individual point matching is rated by its effect on other points, and through relaxation, an overall matching is converged to. Experimental results are given showing a tolerance to some noise.

Lavine, Lambird and Kanal [Lavine 81] try to recognize point patterns without finding an explicit matching using sequences of interpoint distances and show the "correctness" of their method under a plausible noise model.

Bolles [Bolles 79] and Ogawa [Ogawa 86] apply maximal clique techniques to the point pattern matching problems, and the latter uses the Delaunay triangulation to partition a point pattern into a set of triangles reducing the computational cost of matching.

Baird [Baird 84] proposes a method for the situation where there are no missing or extra points, allowing noise whose bounds are specified as arbitrary convex polygons about each model point location. He also gives the comparisons of various methods in terms of the asymptotic order of computational runtime as the number of feature points increases.

In summary, only a few researchers have provided methods to attack the motion estimation problem which allows an arbitrary rigid body motion, when there are missing and extra points, and at the same time the locations of points are observed

with moderate noise. It is exactly this situation that we want to deal with in this thesis.

Chapter 3

Method of Estimation in the Presence of Noise

3.1 Problem Statement

Suppose that we have two consecutive noisy image frames, taken at times t_1 and t_2, ($t_1 < t_2$), of an object in two dimensional space and that the object was subjected to a rigid body motion in the time between the two image frames. Assume that all points of the object lie on a plane called the object plane parallel to the image frames.

The case we consider is where we observe some of the prominent points but do not recognize them individually and some of the observed points may be missed in one or the other of the two image frames. Thus it is not obvious which subsets of the observed points in the two image frames correspond to one another. Moreover the observed locations are assumed to be noisy with Gaussian noise $\mathcal{N}(0, \sigma^2 \mathbf{I}_2)$ independently added to each. The assumption of equal variance σ^2 for both image frames corresponds, in one interpretation, to an assumption that both image frames are at the same distance from the observer.

Now the problem is to estimate the motion parameters, namely the angle θ of rotation about the origin and the translation vector $\boldsymbol{T}$, given the data of the locations of the observed points on the object in the two image frames.

In Figure 3.1, we have an example of simulated data, namely, two image frames showing the noisy locations of the observed points at two different times t_1 and t_2, when there has been a rotation by 90° ($= \pi/2$ radians) followed by a slight translation.

3.2 Basic Idea and Motivation

3.2.1 The Number of Common Observed Points m_0

Let us assume that there are n *distinct* prominent points on the object in which

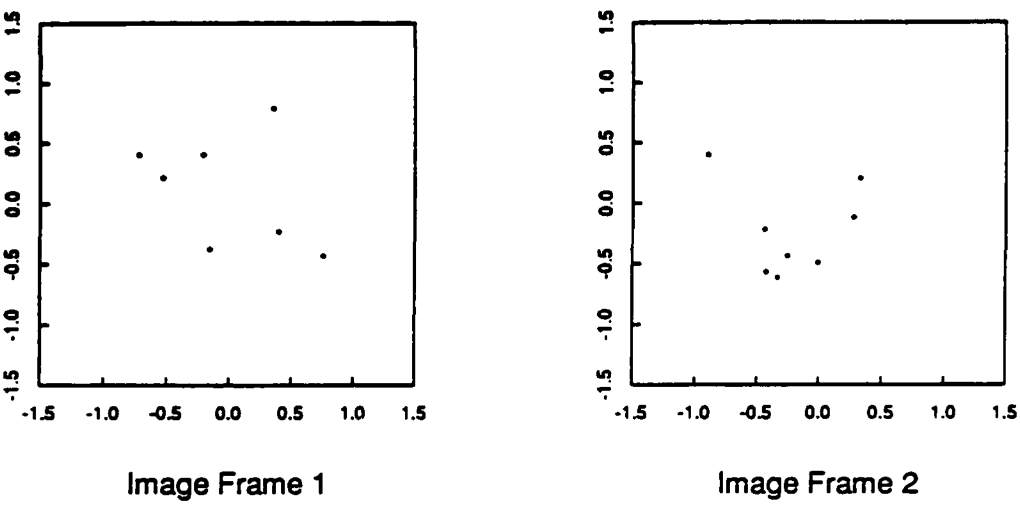


Figure 3.1: An example of simulated data.

we are interested. Let m_1, m_2 be the numbers of the observed points in image frames 1, 2 respectively. Also let m_0 be the number of common observed points in the two image frames. For example, in the simulated data illustrated in Figure 3.1, we have

$$n = 10,\ m_1 = 7,\ m_2 = 8,\ m_0 = 5.$$

If we assume that each of the n prominent points on the object has independent probability p of being observed in any image frame, then the above mentioned m_1, m_2, and m_0 become random variables which we call M_1, M_2, and M_0 respectively. The probability $P(m_1, m_2)$ of getting m_1 observed points in image frame 1, and m_2 in image frame 2 is given by

$$\begin{aligned} P(m_1, m_2) &:= P\{M_1 = m_1, M_2 = m_2\} \\ &= P\{M_1 = m_1\}P\{M_2 = m_2\} \\ &= \binom{n}{m_1} p^{m_1}(1-p)^{n-m_1} \binom{n}{m_2} p^{m_2}(1-p)^{n-m_2} \\ &= \binom{n}{m_1}\binom{n}{m_2} p^{m_1+m_2}(1-p)^{2n-m_1-m_2}, \end{aligned} \tag{3.1}$$

which is just the product of two binomial densities, each with parameters (n, p).

The conditional probability of getting exactly m_0 common observed points given that we have m_1 observed points in image frame 1, and m_2 in image frame 2 is given by

$$P\{M_0 = m_0 \mid M_1 = m_1, M_2 = m_2\} = \frac{\binom{n}{m_0}\binom{n-m_0}{m_1-m_0}\binom{n-m_1}{m_2-m_0}}{\binom{n}{m_1}\binom{n}{m_2}}$$

$$= \frac{m_1!m_2!(n-m_1)!(n-m_2)!}{n!m_0!(m_1-m_0)!(m_2-m_0)!(n-m_1-m_2+m_0)!} \tag{3.2}$$

which is the *hypergeometric* density with parameters (n, m_1, m_2). A hypergeometric random variable X with parameters (N, r, k) has the following interpretation:

> Consider a population of N items, of which k are of type I and $N-k$ are of type II. Assume that a random sample of size r is drawn without replacement from the population. Then X can be defined as the number of items of type I in the sample drawn.

From this interpretation, it is easy to see that

$$\mathrm{E}[X] = \frac{rk}{N} \tag{3.3}$$

and

$$\mathrm{Var}[X] = \frac{rk}{N}(1-\frac{k}{N})(1-\frac{r-1}{N-1}). \tag{3.4}$$

Therefore in our case, we get the conditional mean and standard deviation of M_0 given $\{M_1 = m_1, M_2 = m_2\}$ as

$$\mu(m_0 \mid m_1, m_2) := \mathrm{E}[M_0 \mid M_1 = m_1, M_2 = m_2] = \frac{m_1 m_2}{n} \tag{3.5}$$

and

$$\begin{aligned} \sigma(m_0 \mid m_1, m_2) &:= \sqrt{\mathrm{Var}[M_0 \mid M_1 = m_1, M_2 = m_2]} \\ &= \frac{1}{n}\sqrt{\frac{m_1 m_2 (n-m_1)(n-m_2)}{n-1}}. \end{aligned} \tag{3.6}$$

For example, if $n = 10$, $p = .8$, $m_1 = 7$, $m_2 = 8$, then we have

$$\begin{aligned} P(7,8) &\approx .0608 \\ \mu(m_0 \mid 7,8) &= 5.6 \\ \sigma(m_0 \mid 7,8) &\approx .6110. \end{aligned}$$

For the tabulation of $P(m_1, m_2)$, $\mu(m_0 \mid m_1, m_2)$, $\sigma(m_0 \mid m_1, m_2)$ for m_1, $m_2 \geq 3$ when $n = 10$, $p = .8$, see Table 3.1. For the cases where m_1, $m_2 < 3$, we have $P(m_1, m_2) < 10^{-4}$.

In the simulation study in Chapter 5, we will evaluate the characteristics of the estimation procedure to be proposed for two examples, in each of which m_0, m_1 and m_2 attain certain values. Thus our evaluations may be considered to be, in part, conditional on those values of m_0, m_1 and m_2.

3.2.2 Heuristics of the Method of Estimation

In order to understand the main idea which is employed in the estimation method to be proposed in the next section, let us first consider the situation where no noise is present in the locations of the points observed in the two image frames.

Because of the rigidity of the object involved, the difference vector of any two locations of the m_0 common observed points should not change its length from one image frame to the other. In addition, the change in angle should be $\theta (\mathrm{mod}\ 2\pi)$ where $\theta \in (-\pi, \pi]$ is the amount of the angle of rotation involved. Therefore, if we were to look at all the possible pairs of difference vectors (of the locations), one from each image frame with the lengths of the two being the same, then a clear majority of them would show θ as the exact difference in angles of the two in each pair.

However, this does not give a complete answer to this rather easy-looking problem. For example, if, in considering the prominent points, there are two difference

	m_2							
m_1	10	9	8	7	6	5	4	3
10	.0115 10.0000 .0000	.0288 9.0000 .0000	.0324 8.0000 .0000	.0216 7.0000 .0000	.0095 6.0000 .0000	.0028 5.0000 .0000	.0006 4.0000 .0000	.0001 3.0000 .0000
9	.0288 9.0000 .0000	.0721 8.1000 .3000	.0811 7.2000 .4000	.0540 6.3000 .4583	.0236 5.4000 .4899	.0071 4.5000 .5000	.0015 3.6000 .4899	.0002 2.7000 .4583
8	.0324 8.0000 .0000	.0811 7.2000 .4000	.0912 6.4000 .5333	.0608 5.6000 .6110	.0266 4.8000 .6532	.0080 4.0000 .6666	.0017 3.2000 .6532	.0002 2.4000 .6110
7	.0216 7.0000 .0000	.0540 6.3000 .4583	.0608 5.6000 .6110	.0405 4.9000 .7000	.0177 4.2000 .7483	.0053 3.5000 .7638	.0011 2.8000 .7483	.0002 2.1000 .7000
6	.0095 6.0000 .0000	.0236 5.4000 .4899	.0266 4.8000 .6532	.0177 4.2000 .7483	.0078 3.6000 .8000	.0023 3.0000 .8165	.0005 2.4000 .8000	.0001 1.8000 .7483
5	.0028 5.0000 .0000	.0071 4.5000 .5000	.0080 4.0000 .6667	.0053 3.5000 .7638	.0023 3.0000 .8165	.0007 2.5000 .8333	.0001 2.0000 .8165	.0000 1.5000 .7638
4	.0006 4.0000 .0000	.0015 3.6000 .4899	.0017 3.2000 .6532	.0011 2.8000 .7483	.0005 2.4000 .8000	.0001 2.0000 .8165	.0000 1.6000 .8000	.0000 1.2000 .7483
3	.0001 3.0000 .0000	.0002 2.7000 .4583	.0002 2.4000 .6110	.0002 2.1000 .7000	.0001 1.8000 .7483	.0000 1.5000 .7638	.0000 1.2000 .7483	.0000 .9000 .7000

Table 3.1: The probability distribution of (M_1, M_2) and conditional moments of M_0. Each cell shows $[P(m_1, m_2), \mu(m_0 \mid m_1, m_2), \sigma(m_0 \mid m_1, m_2)]^T$ when $n = 10$.

vectors of the same length, in either one of the two image frames, then they may contribute to incorrectly estimating the value of θ when there are missing points in the image frames. Even if these difference vectors were parallel and contributed to estimating the value of θ correctly, they could still give confusing and misleading information about the translation vector T. In order to alleviate this problem to a certain extent, we need to find additional evidence that a pair of difference vectors, one in each image frame, is a truly corresponding pair with respect to the value of θ which it seems to support.

We propose that being one of the three corresponding pairs of *edges* of two *triangles*, each triangle formed by three observed points, say A_i, B_i, C_i in each image frame i, $i = 1, 2$, so that all three pairs of difference vectors $(D(A_1, B_1), D(A_2, B_2))$, $(D(B_1, C_1), D(B_2, C_2))$, $(D(C_1, A_1), D(C_2, A_2))$ have contributed in supporting the same value of θ, constitutes that evidence. We call such a pair of triangles an *admissible match* of triangles for that particular value of θ. So by maximizing the number of admissible matches of triangles over the possible values of θ, we are likely to get a quite reliable estimate of θ.

Once we have an estimate of θ, it is easy to get an estimate of T. After rotating the points in the first image frame by an angle of θ about the origin, the differences in location of matching points will provide an estimate of T. What constitutes matching points can be determined from the *vertices* of triangles that have been admissibly matched.

It should be noted that it is still possible that a triangle may form admissible matches with several triangles even for a common value of θ.

Now let us return to the problem where we have Gaussian noise $\mathcal{N}(0, \sigma^2 \mathbf{I}_2)$ independently added to each location of the points observed in the two image frames. Here we may expect neither the lengths of the two truly corresponding difference vectors, one from each image frame to be exactly the same, nor the angles of them to differ by exactly $\theta (\mathrm{mod}\ 2\pi)$ as in the noise-free case discussed above. Nevertheless, we would expect them to hold to a reasonable extent if we assume a noise of moderate size. The method we propose, which is presented in detail in the next section, is based primarily on this consideration.

In order to estimate θ and T, we first consider a fixed candidate angle of rotation, θ_0, and try to find approximate matches of the difference vectors, one from each image frame, by two criteria, one for the lengths and the other for the angles, that are similar to the ones used in the noise-free case above except that these are more flexible depending on the size of the noise involved. Any candidate pair of difference vectors, one from each image frame, is removed from further consideration unless it satifies the criteria to a prespecified extent that is set in advance according to the knowledge of the size of the noise.

To each of the *surviving matches* of difference vectors, we then assign a non-zero *weight* according to the angle compatibility with respect to θ_0. We now go further by considering all matches of two triangles, each one formed by three observed points in each image frame, that satisfy the following *admissibility* condition with respect to θ_0:

> A match of two triangles, each one formed by three observed points, say A_i, B_i, C_i in each image frame i, $i = 1, 2$, is *admissible* if all the three pairs of difference vectors $(D(A_1, B_1), D(A_2, B_2))$, $(D(B_1, C_1), D(B_2, C_2))$, $(D(C_1, A_1), D(C_2, A_2))$ are surviving matches of difference vectors.

Then we assign to each admissible match of triangles, a weight which is the geometric mean of the three weights associated with the constituent surviving matches of difference vectors. Now by summing all such weights over all the admissible matches of triangles, we get a *total weight* for the fixed candidate angle θ_0.

Again it should be noted that a triangle may form admissible matches with several triangles even for a common value of θ_0.

Now we first maximize the number of admissible matches of triangles over the possible values of θ_0, and then over the set of all such maximizing values of θ_0, we select the value $\tilde{\theta}_0$ of θ_0 that maximizes the total weight. Our final estimate $\hat{\theta}$ of θ will be a weighted average of the differences in two angles of surviving matches of the difference vectors forming the admissible matches of triangles for $\tilde{\theta}_0$, weights being the ones with respect to $\tilde{\theta}_0$. The final step is to get an estimate $\hat{\boldsymbol{T}}$ of $\boldsymbol{T}$ in exactly the same way as we did in the noise-free case before. Note that if there is no admissible match of triangles, we simply return the answer "Two image frames do not seem to show the same object".

It should be noted that if the prominent points on the object form the vertices of a regular polygon, we may have lack of identifiability, i.e., there may be several possible estimates consistent with the data.

3.3 Method of Estimation

In the last section, we claimed heuristically that the differences in lengths and angles of each pair of difference vectors may play a major role in establishing the *admissible* matches of observed points in two *noisy* image frames taken at times t_1, and t_2 respectively. Now to support that claim, we will discuss in detail the approximate behavior of the differences in lengths and angles.

3.3.1 Assumptions

Suppose that the physical points $P_1, \ldots, P_n$ are the n *distinct* prominent points on the object in which we are interested. For simplicity, but without any loss of generality, let us assume that the first m_0 of these points are observed in both image frames, the next $m_1 - m_0$ points only in image frame 1, and the next $m_2 - m_0$ points only in image frame 2. Since our data consist of the noisy locations of the points observed in the two image frames, again without loss of generality, we will assume that there are only $\tilde{n}$ ($\leq n$) prominent points to begin with, where

$$\tilde{n} := m_0 + (m_1 - m_0) + (m_2 - m_0) = m_1 + m_2 - m_0. \tag{3.7}$$

After making these assumptions, let

$$\mathbf{X}_0 := ((x_1, y_1), \ldots, (x_{\tilde{n}}, y_{\tilde{n}})) \tag{3.8}$$

be the true location of the $\tilde{n}$ observed points $P_1, \ldots, P_{\tilde{n}}$ in the image frame coordinates at time t_1 in that order. Also let

$$\mathbf{X}_1 := ((X_{11}, Y_{11}), \ldots, (X_{1m_1}, Y_{1m_1})) \tag{3.9}$$

be the observed locations of the m_1 points $P_1, \ldots, P_{m_1}$ in image frame 1, and

$$\mathbf{X}_2 := ((X_{21}, Y_{21}), \ldots, (X_{2m_2}, Y_{2m_2})) \tag{3.10}$$

be the observed locations of the m_2 points $P_1, \ldots, P_{m_0}, P_{m_1+1}, \ldots, P_{\tilde{n}}$ in image frame 2 in their respective orders. However, remember that what we have as actual data are two *unordered* sets of observed locations, one from each image frame.

Our independent Gaussian noise assumption can be expressed in the following matrix notation:

All components of $\mathbf{X}_1$ and $\mathbf{X}_2$ are independently distributed with

$$\mathcal{L}\left(\begin{bmatrix} X_{1i} \\ Y_{1i} \end{bmatrix}\right) = \mathcal{N}(\mu_{1i}, \sigma^2 \mathbf{I}_2), \quad i = 1, \ldots, m_1, \tag{3.11}$$

and

$$\mathcal{L}\left(\begin{bmatrix} X_{2j} \\ Y_{2j} \end{bmatrix}\right) = \mathcal{N}(\mu_{2j}, \sigma^2 \mathbf{I}_2), \quad j = 1, \ldots, m_2, \tag{3.12}$$

where

$$\mu_{1i} := \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \quad i = 1, \ldots, m_1, \tag{3.13}$$

and

$$\mu_{2j} := \begin{cases} \mathbf{U}(\theta) \begin{bmatrix} x_j \\ y_j \end{bmatrix} + T & \text{if } j = 1, \ldots, m_0 \\ \\ \mathbf{U}(\theta) \begin{bmatrix} x_{j+m_1-m_0} \\ y_{j+m_1-m_0} \end{bmatrix} + T & \text{if } j = m_0 + 1, \ldots, m_2 \end{cases} . \tag{3.14}$$

The notational correspondence is described again in Table 3.2.

3.3.2 The Length Difference and Angle Difference

Let us consider any two prominent points P_k, P_l with $1 \leq k < l \leq m_0$, so that they are among the common observed points. Let us fix k and l for the time being.

We first look at their true locations μ_{ik}, μ_{il} in each image frame i, $i = 1, 2$. We have

$$\begin{aligned} D(\mu_{1k}, \mu_{1l}) &:= \mu_{1l} - \mu_{1k} \\ &= \begin{bmatrix} x_l - x_k \\ y_l - y_k \end{bmatrix}, \end{aligned} \tag{3.15}$$

and

$$\begin{aligned} D(\mu_{2k}, \mu_{2l}) &:= \mu_{2l} - \mu_{2k} \\ &= \mathbf{U}(\theta) \begin{bmatrix} x_l - x_k \\ y_l - y_k \end{bmatrix}. \end{aligned} \tag{3.16}$$

Clearly we know

$$\|D(\mu_{2k}, \mu_{2l})\| = \|D(\mu_{1k}, \mu_{1l})\|, \tag{3.17}$$

namely, the true difference vectors corresponding to the same prominent points P_k, P_l in the two image frames have the same lengths. Moreover if we consider their *angles*,

$$\varphi_i := \arctan(D(\mu_{ik}, \mu_{il})), \ i = 1, 2, \tag{3.18}$$

Physical Points	$\mathbf{X}_0$	True Locations in Image Frame 1	$\mathbf{X}_1$	True Locations in Image Frame 2	$\mathbf{X}_2$
P_1	$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$	$\mu_{11} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$	$\begin{bmatrix} X_{11} \\ Y_{11} \end{bmatrix}$	$\mu_{21} = \mathbf{U}(\theta)\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \boldsymbol{T}$	$\begin{bmatrix} X_{21} \\ Y_{21} \end{bmatrix}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
P_{m_0}	$\begin{bmatrix} x_{m_0} \\ y_{m_0} \end{bmatrix}$	$\mu_{1m_0} = \begin{bmatrix} x_{m_0} \\ y_{m_0} \end{bmatrix}$	$\begin{bmatrix} X_{1m_0} \\ Y_{1m_0} \end{bmatrix}$	$\mu_{2m_0} = \mathbf{U}(\theta)\begin{bmatrix} x_{m_0} \\ y_{m_0} \end{bmatrix} + \boldsymbol{T}$	$\begin{bmatrix} X_{2m_0} \\ Y_{2m_0} \end{bmatrix}$
P_{m_0+1}	$\begin{bmatrix} x_{m_0+1} \\ y_{m_0+1} \end{bmatrix}$	$\mu_{1m_0+1} = \begin{bmatrix} x_{m_0+1} \\ y_{m_0+1} \end{bmatrix}$	$\begin{bmatrix} X_{1m_0+1} \\ Y_{1m_0+1} \end{bmatrix}$	NA	NA
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
P_{m_1}	$\begin{bmatrix} x_{m_1} \\ y_{m_1} \end{bmatrix}$	$\mu_{1m_1} = \begin{bmatrix} x_{m_1} \\ y_{m_1} \end{bmatrix}$	$\begin{bmatrix} X_{1m_1} \\ Y_{1m_1} \end{bmatrix}$	NA	NA
P_{m_1+1}	$\begin{bmatrix} x_{m_1+1} \\ y_{m_1+1} \end{bmatrix}$	NA	NA	$\mu_{2m_0+1} = \mathbf{U}(\theta)\begin{bmatrix} x_{m_1+1} \\ y_{m_1+1} \end{bmatrix} + \boldsymbol{T}$	$\begin{bmatrix} X_{2m_0+1} \\ Y_{2m_0+1} \end{bmatrix}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$P_{\tilde{n}}$	$\begin{bmatrix} x_{\tilde{n}} \\ y_{\tilde{n}} \end{bmatrix}$	NA	NA	$\mu_{2m_2} = \mathbf{U}(\theta)\begin{bmatrix} x_{\tilde{n}} \\ y_{\tilde{n}} \end{bmatrix} + \boldsymbol{T}$	$\begin{bmatrix} X_{2m_2} \\ Y_{2m_2} \end{bmatrix}$

Table 3.2: The notational correspondence. Image frames 1 and 2 taken at times t_1 and t_2, respectively. "NA" represents the nonavailability of the information. $\tilde{n} = m_1 + m_2 - m_0$.

then we have

$$\varphi_2 = c(\varphi_1 + \theta),$$

or equivalently,

$$\theta = c(\varphi_2 - \varphi_1) \tag{3.19}$$

if we add the restriction that $\theta \in (-\pi, \pi]$. In other words, the angle θ of rotation can be recovered from the difference $\varphi_2 - \varphi_1$ of the two angles, one from each image frame, of the true difference vectors corresponding to the same prominent points P_k, P_l.

Now we look at the observed locations (X_{ik}, Y_{ik}), (X_{il}, Y_{il}) of P_k, P_l in each image frame i, and denote them by P_{ik}, P_{il} respectively, for $i = 1, 2$. Then it is easy to see that

$$\mathcal{L}[D(P_{ik}, P_{il})] = \mathcal{N}(D(\mu_{ik}, \mu_{il}), 2\sigma^2 \mathrm{I}_2), \quad \text{for } i = 1, 2, \tag{3.20}$$

and hence we get

$$\mathcal{L}\left[\frac{\|D(P_{ik}, P_{il})\|^2}{2\sigma^2}\right] = \chi_2^2\left(\frac{\|D(\mu_{ik}, \mu_{il})\|^2}{2\sigma^2}\right), \quad \text{for } i = 1, 2, \tag{3.21}$$

where $\chi_q^2(\delta^2)$ is the *non-central Chi-square* distribution with q *degrees of freedom* and *non-centrality parameter* δ^2. Since a $\chi_q^2(\delta^2)$ random variable has mean $q + \delta^2$ and variance $2q + 4\delta^2$ (see [Johnson 70, pp. 130–135] for derivation), we get for $i = 1, 2$,

$$\mathrm{E}\left[\frac{\|D(P_{ik}, P_{il})\|^2}{2\sigma^2}\right] = 2 + \frac{\|D(\mu_{ik}, \mu_{il})\|^2}{2\sigma^2}, \tag{3.22}$$

and

$$\mathrm{Var}\left[\frac{\|D(P_{ik}, P_{il})\|^2}{2\sigma^2}\right] = 4 + \frac{2\|D(\mu_{ik}, \mu_{il})\|^2}{\sigma^2}. \tag{3.23}$$

It follows that for $i = 1, 2$,

$$\mathrm{E}\left[\|D(P_{ik}, P_{il})\|^2\right] = 4\sigma^2 + \|D(\mu_{ik}, \mu_{il})\|^2, \tag{3.24}$$

and

$$\mathrm{Var}\left[\|D(P_{ik}, P_{il})\|^2\right] = 16\sigma^4 + 8\sigma^2\|D(\mu_{ik}, \mu_{il})\|^2. \tag{3.25}$$

In the next subsection, we will derive approximation formulae for $\mathrm{E}\left[\|D(P_{ik}, P_{il})\|\right]$, $\mathrm{Var}\left[\|D(P_{ik}, P_{il})\|\right]$, $\mathrm{E}\left[\arctan\left(D(P_{ik}, P_{il})\right)\right]$, and $\mathrm{Var}\left[\arctan\left(D(P_{ik}, P_{il})\right)\right]$, for sufficiently small $\sigma > 0$. To do that, let us first simplify the notation by taking

$$\begin{aligned} D_i &:= D(P_{ik}, P_{il}) && (3.26)\\ D_{0i} &:= D(\mu_{ik}, \mu_{il}) && (3.27)\\ d_i &:= \|D_i\| && (3.28)\\ d_{0i} &:= \|D_{0i}\| && (3.29) \end{aligned}$$

for $i = 1, 2$. Then (3.20)–(3.25) become

$$\begin{aligned} \mathcal{L}\left[D_i\right] &= \mathcal{N}(D_{0i}, 2\sigma^2 \mathrm{I}_2), && (3.30)\\ \mathcal{L}\left[\frac{d_i^2}{2\sigma^2}\right] &= \chi_2^2\left(\frac{d_{0i}^2}{2\sigma^2}\right), && (3.31)\\ \mathrm{E}\left[\frac{d_i^2}{2\sigma^2}\right] &= 2 + \frac{d_{0i}^2}{2\sigma^2}, && (3.32)\\ \mathrm{Var}\left[\frac{d_i^2}{2\sigma^2}\right] &= 4 + \frac{2d_{0i}^2}{\sigma^2}, && (3.33)\\ \mathrm{E}\left[d_i^2\right] &= 4\sigma^2 + d_{0i}^2, && (3.34)\\ \mathrm{Var}\left[d_i^2\right] &= 16\sigma^4 + 8\sigma^2 d_{0i}^2, && (3.35) \end{aligned}$$

for $i = 1, 2$.

3.3.3 Approximations

We begin this subsection with a statement without proof of the following theorem which justifies a method called the "*δ-method*" for finding the approximate mean and covariance of a function of a random variable. For a slight variation and a discussion, see [Bishop 75, pp. 492–494].

Theorem 3.1 **(δ-method)** *Let $\mathbf{V}_\tau$ be an m-dimensional real vector valued random variable, $\boldsymbol{v}_0 \in \mathbf{R}^m$, and let*

$$t(\boldsymbol{v}) := \begin{bmatrix} t_1(\boldsymbol{v}) \\ \vdots \\ t_d(\boldsymbol{v}) \end{bmatrix}$$

be a d-dimensional real vector valued function defined on some neighborhood N of $\boldsymbol{v}_0$ in $\mathbf{R}^m$ such that the Jacobian matrix $\partial t(\boldsymbol{v})/\partial \boldsymbol{v}$ exists and is continuous on N. Suppose also that

$$\mathcal{L}\left[\frac{\mathbf{V}_\tau - \boldsymbol{v}_0}{\tau}\right] \to \mathcal{L}\left[\mathbf{W}\right] \quad \text{as } \tau \to 0, \tag{3.36}$$

for some random variable $\mathbf{W}$.

Then $t(\mathbf{V}_\tau)$ admits the first order Taylor Expansion

$$t(\mathbf{V}_\tau) = t(\boldsymbol{v}_0) + \left.\frac{\partial t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\boldsymbol{v}_0} (\mathbf{V}_\tau - \boldsymbol{v}_0) + o_p(\tau) \quad \text{as } \tau \to 0. \tag{3.37}$$

Also

$$\mathbf{V}_\tau \to \boldsymbol{v}_0 \quad \text{in probability} \quad \text{as } \tau \to 0, \tag{3.38}$$

$$t(\mathbf{V}_\tau) \to t(\boldsymbol{v}_0) \quad \text{in probability} \quad \text{as } \tau \to 0, \tag{3.39}$$

and

$$\mathcal{L}\left[\frac{t(\mathbf{V}_\tau) - t(\boldsymbol{v}_0)}{\tau}\right] \to \mathcal{L}\left[\left.\frac{\partial t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\boldsymbol{v}_0} \mathbf{W}\right] \quad \text{as } \tau \to 0. \tag{3.40}$$

The conclusion of Theorem 3.1 is especially useful, when we have

$$\mathcal{L}\left[\mathbf{W}\right] = \mathcal{N}(0, \Sigma), \tag{3.41}$$

where it follows that

$$\mathcal{L}\left[\left.\frac{\partial t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\boldsymbol{v}_0} \mathbf{W}\right] = \mathcal{N}\left(0, \left.\frac{\partial t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\boldsymbol{v}_0} \Sigma \left.\frac{\partial t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\boldsymbol{v}_0}^T\right). \tag{3.42}$$

We can interpret this as saying that for small $\tau > 0$,

$$\mathcal{L}[t(\mathbf{V}_\tau)] \approx \mathcal{N}\left(t(\boldsymbol{v}_0), \tau^2 \left.\frac{\partial t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\boldsymbol{v}_0} \Sigma \left.\frac{\partial t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\boldsymbol{v}_0}^T\right), \tag{3.43}$$

and hence $t(\mathbf{V}_\tau)$ behaves like a Gaussian random variable t^* for which

$$\mathrm{E}[t^*] \approx t(\boldsymbol{v}_0), \tag{3.44}$$

and

$$\mathrm{Cov}[t^*] \approx \tau^2 \left.\frac{\partial t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\boldsymbol{v}_0} \Sigma \left.\frac{\partial t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right|_{\boldsymbol{v}=\boldsymbol{v}_0}^T . \tag{3.45}$$

First we apply Theorem 3.1 with $\mathbf{V}_\tau = d_i^2$, $\boldsymbol{v}_0 = d_{0i}^2$, $t(\boldsymbol{v}) = \sqrt{\boldsymbol{v}}$ and $\tau = \sigma$. To do so we must find the limiting distribution of $(d_i^2 - d_{0i}^2)/\sigma$. We have

$$\begin{aligned}
\frac{d_i^2 - d_{0i}^2}{\sigma} &= \frac{2\sigma^2\chi_2^2(\delta^2) - d_{0i}^2}{\sigma} \quad \text{where } \delta^2 = \frac{d_{0i}^2}{2\sigma^2} \\
&= \frac{2\sigma^2}{\sigma}\left[(Z_1+\delta)^2 + Z_2^2 - \delta^2\right] \quad \text{where } \mathcal{L}\left(\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}\right) = \mathcal{N}(0, \mathbf{I}_2) \\
&= 2\sigma\left[2\delta Z_1 + (Z_1^2 + Z_2^2)\right] = 2\sqrt{2}d_{0i}Z_1 + O_p(\sigma).
\end{aligned}$$

Hence

$$\mathcal{L}\left[\frac{d_i^2 - d_{0i}^2}{\sigma}\right] \to \mathcal{L}\left[\mathbf{W}\right] = \mathcal{N}(0, 8d_{0i}^2).$$

Therefore from (3.43), we get for small $\sigma > 0$,

$$\mathcal{L}[d_i] \approx \mathcal{N}\left(d_{0i}, \sigma^2\left(\frac{1}{2d_{0i}}\right)^2(8d_{0i}^2)\right) = \mathcal{N}\left(d_{0i}, 2\sigma^2\right). \tag{3.46}$$

Again we apply Theorem 3.1 with $\mathbf{V}_\tau = \boldsymbol{D}_i$, $\boldsymbol{v}_0 = \boldsymbol{D}_{0i}$, $t(\boldsymbol{v}) = \arctan(\boldsymbol{v})$ and $\tau = \sigma$. In this case, it is obvious that

$$\mathcal{L}\left[\frac{\boldsymbol{D}_i - \boldsymbol{D}_{0i}}{\sigma}\right] = \mathcal{N}(0, 2\mathbf{I}_2) \to \mathcal{N}(0, 2\mathbf{I}_2).$$

Now we have

$$\begin{aligned}
\frac{\partial t(v)}{\partial \boldsymbol{v}} &= \left[\frac{\partial \arctan(v_1, v_2)}{\partial v_1}, \frac{\partial \arctan(v_1, v_2)}{\partial v_2}\right] \\
&= \left[-\frac{v_2}{v_1^2}\cos^2(\arctan(v_1, v_2)), \frac{1}{v_1}\cos^2(\arctan(v_1, v_2))\right] \\
&= \left[-\frac{v_2}{v_1^2 + v_2^2}, \frac{v_1}{v_1^2 + v_2^2}\right],
\end{aligned}$$

and hence from (3.43), we get for small $\sigma > 0$,

$$\mathcal{L}[\arctan(\boldsymbol{D}_i)] \approx \mathcal{N}\left(\arctan(\boldsymbol{D}_{0i}), \frac{2\sigma^2}{d_{0i}^2}\right). \tag{3.47}$$

It should be noted that except for the case where $\arctan(\boldsymbol{D}_{0i}) = \pi$, the approximating normal distribution in (3.47) assigns positive probability outside the interval $(-\pi, \pi]$, but that this probability approaches zero as $\sigma \to 0$, while the actual distribution is confined to $(-\pi, \pi]$, *i.e.*, a distribution on the circle. The exceptional case may be treated by translating the probability in the approximating distribution over $(\pi, 2\pi)$ to $(-\pi, 0)$. However the need for this peculiar distribution on the line will disappear below, when we consider the distribution of a difference of angles which centers about 0.

Since $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are independent, and so are d_1 and d_2, we see from (3.46) and (3.47) that for small $\sigma > 0$,

$$\mathcal{L}[d_2 - d_1] \approx \mathcal{N}\left(d_{02} - d_{01}, 4\sigma^2\right) = \mathcal{N}\left(0, 4\sigma^2\right) \tag{3.48}$$

since

$$d_{02} = d_{01}$$

from (3.17). Also

$$\begin{aligned} \mathcal{L}[\arctan(\boldsymbol{D}_2) - \arctan(\boldsymbol{D}_1)] &\approx \mathcal{N}\left(\arctan(\boldsymbol{D}_{02}) - \arctan(\boldsymbol{D}_{01}), \frac{2\sigma^2}{d_{01}^2} + \frac{2\sigma^2}{d_{02}^2}\right) \\ &= \mathcal{N}\left(\arctan(\boldsymbol{D}_{02}) - \arctan(\boldsymbol{D}_{01}), \frac{4\sigma^2}{d_{01}^2}\right), \end{aligned} \tag{3.49}$$

or

$$\mathcal{L}[c(\arctan(\boldsymbol{D}_2) - \arctan(\boldsymbol{D}_1) - \theta)] \approx \mathcal{N}\left(0, \frac{4\sigma^2}{d_{01}^2}\right) \tag{3.50}$$

since

$$c(\arctan(\boldsymbol{D}_{02}) - \arctan(\boldsymbol{D}_{01}) - \theta) = 0 \tag{3.51}$$

from (3.18) and (3.19).

Therefore, we may conclude from (3.48) and (3.50) that for small $\sigma > 0$,

$$\mathcal{L}\left[\frac{d_2 - d_1}{2\sigma}\right] \approx \mathcal{N}(0,1) \tag{3.52}$$

and

$$\mathcal{L}\left[\frac{c(\arctan(D_2) - \arctan(D_1) - \theta)d_{01}}{2\sigma}\right] \approx \mathcal{N}(0,1). \tag{3.53}$$

We may go one step further by replacing d_{01} in (3.53) by the estimate $(d_1 + d_2)/2$ in view of (3.46) and get

$$\mathcal{L}\left[\frac{c(\arctan(D_2) - \arctan(D_1) - \theta)}{2\sigma}\frac{d_1 + d_2}{2}\right] \approx \mathcal{N}(0,1). \tag{3.54}$$

The approximation results (3.52) and (3.54) will play the key roles in the estimation procedure presented in the next subsection.

3.3.4 Estimation Procedure

The two approximate distributional results (3.52) and (3.54) provide us with valid statistics with which we can measure the significance of the correspondence between any pair of difference vectors, say $D(P_{1i}, P_{1j})$, $D(P_{2k}, P_{2l})$, of the observed locations P_{1i}, P_{1j} in image frame 1, and P_{2k}, P_{2l} in image frame 2, and thus enable us to assign a weight to each correspondence. To do that let us first define functions $h(u)$, $\rho_1(r)$ and $\rho_2(r)$ that are needed in our weighting scheme, by

$$h(u) := \max(2.5 - u^2, 0) \quad \text{for } u \in \mathbf{R}^1, \tag{3.55}$$

$$\rho_1(r) := \sqrt{.2 + r^2} \quad \text{for } r \geq 0, \tag{3.56}$$

$$\rho_2(r) := -4.119(r+1)\exp\{-(r+1)\} + 1.915 \quad \text{for } r \geq 0. \tag{3.57}$$

The roles of these functions are to be explained later after we indicate how they are used in our weighting scheme (see Figure 3.2 for their graphs).

In the following steps, let us describe the estimation procedure, where we take the value of σ to be $\sigma_a = k\sigma$ for $k > 0$. The effect of this misspecification of σ when $k \neq 1$ will be discussed in the next two chapters.

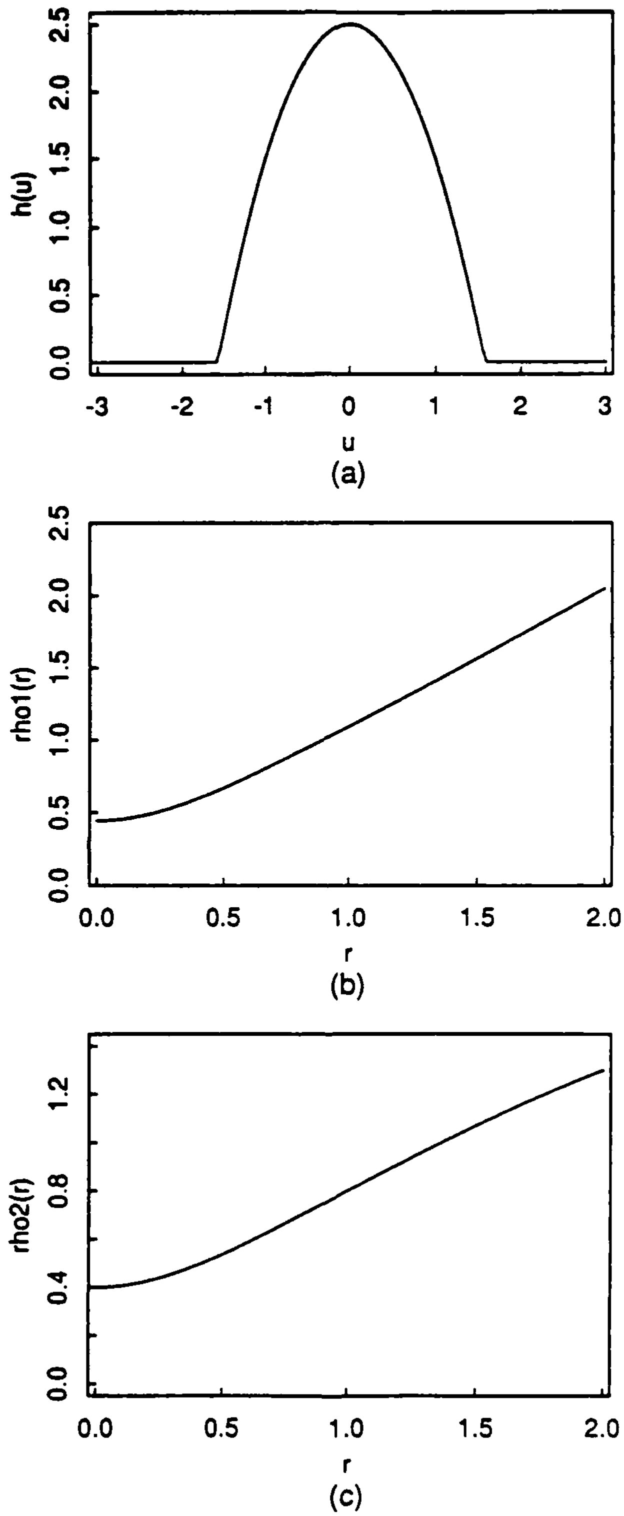


Figure 3.2: Graphs of the functions (a) $h(u)$, (b) $\rho_1(r)$ and (a) $\rho_2(r)$.

To make the procedure feasible on a computer, we fix a finite grid $\Gamma \subset (-\pi, \pi]$ of candidate angles θ_0 of rotation. Specifically, in the simulations in Chapter 5,

$$\Gamma = \{(.01)n \mid n = 0, \pm 1, \ldots, \pm 314\} \tag{3.58}$$

STEP 1 *Assign an arbitrary order to the observed locations in image frame 1, and label them by that order,* $P_{11}, \ldots, P_{1m_1}$. *Similarly label the observed locations in image frame 2 by* $P_{21}, \ldots, P_{2m_2}$.

STEP 2 *For each pair of observed locations* (P_{1i}, P_{1j}), $1 \leq i < j \leq m_1$, *in image frame 1 and each pair of observed locations* (P_{2k}, P_{2l}), $1 \leq k \neq l \leq m_2$ *in image frame 2, form the difference vectors* $D(P_{1i}, P_{1j})$ *and* $D(P_{2k}, P_{2l})$, *and then compute*

$$R_{1ij} := \|D(P_{1i}, P_{1j})\| \tag{3.59}$$

$$R_{2kl} := \|D(P_{2k}, P_{2l})\| \tag{3.60}$$

$$\varphi_{1ij} := arctan(D(P_{1i}, P_{1j})) \tag{3.61}$$

$$\varphi_{2kl} := arctan(D(P_{2k}, P_{2l})). \tag{3.62}$$

STEP 3 *For each* i, j, $1 \leq i < j \leq m_1$, *and* k, l, $1 \leq k \neq l \leq m_2$, *consider the pair of difference vectors* $[D(P_{1i}, P_{1j}), D(P_{2k}, P_{2l})]$. *Call* $[ij, kl] := [D(P_{1i}, P_{1j}), D(P_{2k}, P_{2l})]$ *to be a* candidate match of difference vectors *if*

$$h\left(\frac{R_{2kl} - R_{1ij}}{2\sigma_a}\right) > 0. \tag{3.63}$$

Then denote by $\mathcal{C}$, *the set of all candidate matches of difference vectors,* $[ij, kl]$.

STEP 4 *Denote the ordered triple* $([i_1i_2, j_1j_2], [i_1i_3, j_1j_3], [i_2i_3, j_2j_3]) \in \mathcal{C}^3$ *by* $[\vec{\imath}, \vec{\jmath}] := [i_1i_2i_3, j_1j_2j_3]$, *and call it a* candidate match of triangles. *Let*

$$\mathcal{C}^* := \{[\vec{\imath}, \vec{\jmath}] \in \mathcal{C}^3\} \tag{3.64}$$

be the set of all candidate matches of triangles.

STEP 5 *Fix a candidate angle of rotation,* $\theta_0 \in \Gamma$.

STEP 6 *Denote the set* $\{(1,2),(1,3),(2,3)\}$ *by* $\langle 123 \rangle$. *For each* $[\vec{i}, \vec{j}] \in C^*$ *and each* $(s,t) \in \langle 123 \rangle$, *compute the* weight *of* $[i_s i_t, j_s j_t]$ *for* θ_0 *by*

$$w_{[i_s i_t, j_s j_t]}(\theta_0) := \rho_1 \left(\frac{R_{1 i_s i_t} + R_{2 j_s j_t}}{2} \right) h \left(\frac{c\,(\varphi_{2 j_s j_t} - \varphi_{1 i_s i_t} - \theta_0)}{2\sigma_a} \rho_2 \left(\frac{R_{1 i_s i_t} + R_{2 j_s j_t}}{2} \right) \right). \tag{3.65}$$

If $w_{[i_s i_t, j_s j_t]}(\theta_0) > 0$, *then call* $[i_s i_t, j_s j_t]$ *a* surviving match of difference vectors for θ_0, *and let*

$$S(\theta_0) := \{[i_s i_t, j_s j_t] \in C \mid [\vec{i}, \vec{j}] \in C^*, (s,t) \in \langle 123 \rangle, w_{[i_s i_t, j_s j_t]}(\theta_0) > 0\} \tag{3.66}$$

be the set of all surviving matches of difference vectors for θ_0.

We digress here to elaborate on the roles of the functions h, ρ_1, and ρ_2. As can be seen from Figure 3.2 (a), $h(u)$ is a non-negative even function which serves the following two purposes.

First, h serves as an "*aperture*" through which the argument u has to pass in order to attain a non-zero h value at all. Specifically, it sets $\sqrt{2.5} \approx 1.5811$ as the upper limit by which $|u|$ is bounded to attain a non-zero h value. In particular, if we have a real random variable Z, with $\mathrm{E}[Z] = 0$ and $\mathrm{Var}[Z] = 1$, in place of u, then we can interpret this as saying that only the observations Z within ± 1.5811 standard deviations from the expected value will get the non-zero h values. Note that the probability of such an event,

$$P\{h(Z) > 0\} \approx P\{|Z| < 1.5811\}$$

is approximately .886 if $\mathcal{L}[Z] \approx \mathcal{N}(0,1)$. Remember that, from (3.52) and (3.54), if $[ij, kl]$ is a true match and $\sigma_a = \sigma$, then we have

$$\mathcal{L}\left[\frac{R_{2kl} - R_{1ij}}{2\sigma_a} \right] \approx \mathcal{N}(0,1),$$

and

$$\mathcal{L}\left[\frac{c\,(\varphi_{2kl}-\varphi_{1ij}-\theta_0)}{2\sigma_a}\frac{R_{1ij}+R_{2kl}}{2}\right]\approx \mathcal{N}(0,1).$$

Secondly, h, by itself can be used as a weight of a special kind assigned to the argument value u. The resulting weights will have the property that as $|u|$ decreases from 1.5811, the weight increases sharply and reaches a smooth maximum, namely 2.5, at $u = 0$. This property insures that, once a value of u passes through the non-zero aperture, it will get a weight which is not too close to zero, and at the same time, the values of u which are close to 0 will get almost the same weights as $u = 0$ does.

The functions ρ_1 and ρ_2 are used to adjust the basic weighting scheme provided by h in order to incorporate the following idea. Suppose that we have defined, in STEP 6, the weight of $[i_s i_t, j_s j_t]$ for θ_0 by

$$w^*_{[i_s i_t, j_s j_t]}(\theta_0) := h\left(\frac{c\,(\varphi_{2j_s j_t}-\varphi_{1i_s i_t}-\theta_0)}{2\sigma_a}\frac{R_{1i_s i_t}+R_{2j_s j_t}}{2}\right). \qquad (3.67)$$

This weight fails to take into account that matching long difference vectors provide more accurate information about θ than do short ones. So by defining the weight by (3.65) instead in STEP 6, we enlarge the non-zero aperture for the pairs of large difference vectors using ρ_2, and at the same time, we get a weight roughly of the order of the average length of the two difference vectors using ρ_1. The reason that we use $\sqrt{.2+r^2}$ for $\rho_1(r)$ instead of r is to bound the function away from 0 in the neighborhood of $r = 0$. For $\rho_2(r)$, functions of the form

$$-a(r+1)\exp\{-(r+1)\}+b$$

were first considered in order to get a point of inflection at $r = 1$, and then the coefficients a, b were determined to satisfy the conditions

$$\rho_2(0) = .4 \text{ and } \rho_2(1) = .8.$$

The choice of these functions is somewhat arbitrary.

We now describe the next estimation step.

STEP 7 *For each* $[\vec{\imath},\vec{\jmath}] \in C^*$, *if* $[\vec{\imath},\vec{\jmath}] \in S(\theta_0)^3$, *then call it an* admissible match of triangles for θ_0, *and let*

$$A(\theta_0) := \{[\vec{\imath},\vec{\jmath}] \in C^* \mid [\vec{\imath},\vec{\jmath}] \in S(\theta_0)^3\} = C^* \cap S(\theta_0)^3 \tag{3.68}$$

be the set of all admissible matches of triangles for θ_0.

STEP 8 *For each* $[\vec{\imath},\vec{\jmath}] \in A(\theta_0)$, *define its weight for* θ_0 *by*

$$w_{[\vec{\imath},\vec{\jmath}]}(\theta_0) := \Big(\prod_{(s,t)\in\langle 123\rangle} w_{[i_s i_t, j_s j_t]}(\theta_0)\Big)^{1/3} \tag{3.69}$$

and let

$$w(\theta_0) := \sum_{[\vec{\imath},\vec{\jmath}]\in A(\theta_0)} w_{[\vec{\imath},\vec{\jmath}]}(\theta_0). \tag{3.70}$$

STEP 9 *Let*

$$N(\theta_0) := |A(\theta_0)| \tag{3.71}$$

be the number of the admissible matches of triangles for θ_0. *Define*

$$\Theta_{\max} := \{\check{\theta}_0 \in \Gamma \mid N(\check{\theta}_0) = \max_{\theta_0\in\Gamma} N(\theta_0) > 0\}. \tag{3.72}$$

If $\Theta_{\max} = \emptyset$, *then return the answer "Two image frames do not seem to show the same object."*

STEP 10 *Let* $\tilde{\theta}_0 \in \Theta_{\max}$ *be the smallest value in* Γ *of* θ_0 *such that*

$$w(\tilde{\theta}_0) = \max_{\check{\theta}_0\in\Theta_{\max}} w(\check{\theta}_0). \tag{3.73}$$

Remark: Usually, there will be only one $\tilde{\theta}_0$ satisfying (3.73).

STEP 11 *For each* $[\bar{\imath},\bar{\jmath}] \in \mathcal{A}(\tilde{\theta}_0)$ *and each* $(s,t) \in \langle 123 \rangle$, *define*

$$\nabla\angle_{[\bar{\imath},\bar{\jmath}]}(s,t) := c(\varphi_{2j_s j_t} - \varphi_{1i_s i_t} - \tilde{\theta}_0). \tag{3.74}$$

Then define the final estimate $\hat{\theta}$ *of the angle of rotation* θ *by*

$$\hat{\theta} := \tilde{\theta}_0 + \frac{\sum_{[\bar{\imath},\bar{\jmath}]\in\mathcal{A}(\tilde{\theta}_0)} w_{[\bar{\imath},\bar{\jmath}]}(\tilde{\theta}_0) \sum_{(s,t)\in\langle 123\rangle} \nabla\angle_{[\bar{\imath},\bar{\jmath}]}(s,t)}{3w(\tilde{\theta}_0)}. \tag{3.75}$$

Note that

$$w(\tilde{\theta}_0) = \sum_{[\bar{\imath},\bar{\jmath}]\in\mathcal{A}(\tilde{\theta}_0)} w_{[\bar{\imath},\bar{\jmath}]}(\tilde{\theta}_0).$$

STEP 12 *Define the estimate* $\hat{\boldsymbol{T}}$ *of the translation vector* $\boldsymbol{T}$ *by*

$$\hat{\boldsymbol{T}} := \frac{1}{3N(\hat{\theta}_0)} \sum_{[\bar{\imath},\bar{\jmath}]\in\mathcal{A}(\tilde{\theta}_0)} \left(\sum_{k=1}^{3} P_{2j_k} - \mathbf{U}(\hat{\theta}) \sum_{k=1}^{3} P_{1i_k}\right). \tag{3.76}$$

Chapter 4

Maximum Likelihood Estimate in the Recognized Points Problem

In the estimation method we proposed in the previous chapter, we have assumed that the observed points in the two image frames are not individually recognized. In principle one might apply maximum likelihood to the estimation problem, but the difficulty in formulating and calculating the likelihood function under the above assumption is formidable. However, one ought to expect to do better when all the observed points are recognized so that the m_0 common points among them in the two image frames are matched without error in identification. Hence, this latter situation, which is readily treated by maximum likelihood as will be seen in this chapter, should provide us with lower bounds for the error in our estimates for the original problem. We refer to the two problems by the labels, "*Unrecognized Points*" and "*Recognized Points*" problems, respectively.

In this chapter, we will prove the existence of the maximum likelihood estimate, and derive its asymptotic normality and consistency as $\sigma \to 0$, for the Recognized Points problem. Similar results are derived even when σ is incorrectly assumed to be $\sigma_a = k\sigma$ for positive $k \neq 1$ in constructing the likelihood function.

We continue to use the same notation as in Section 3.3, and assume that $m_0 \geq 2$ throughout this chapter.

4.1 Simplified Assumptions and the Maximum Likelihood Estimate

For the Recognized Points problem, where we recognize all the observed points, we can assume without loss of generality that

$$m_0 = m_1 = m_2 = \tilde{n}(\geq 2), \tag{4.1}$$

since the $m_i - m_0$ unmatched observed points in each image frame i, $i = 1, 2$, do not provide us with any additional information about the motion parameters θ and

$\boldsymbol{T}$. Accordingly, let us assume that the three vectors

$$\mathbf{X}_0 := ((x_1, y_1), \ldots, (x_{m_0}, y_{m_0})) \in \mathbf{R}^{2m_0} \tag{4.2}$$

$$\mathbf{X}_1 := ((X_{11}, Y_{11}), \ldots, (X_{1m_0}, Y_{1m_0})) \in \mathbf{R}^{2m_0} \tag{4.3}$$

$$\mathbf{X}_2 := ((X_{21}, Y_{21}), \ldots, (X_{2m_0}, Y_{2m_0})) \in \mathbf{R}^{2m_0} \tag{4.4}$$

are the true locations at time t_1, the observed locations in image frame 1 (at time t_1), the observed locations in image frame 2 (at time t_2), respectively, of the m_0 *distinct* prominent points $P_1, \ldots, P_{m_0}$ in that order. For notational simplicity, let us define the vector of observations $\mathbf{X}$ by

$$\mathbf{X} := \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \in \mathbf{R}^{4m_0}. \tag{4.5}$$

Similarly, let us define the vectors $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\mu}$ by

$$\boldsymbol{\mu}_1 := \begin{bmatrix} \mu_{11} \\ \vdots \\ \mu_{1m_0} \end{bmatrix} \in \mathbf{R}^{2m_0} \tag{4.6}$$

$$\boldsymbol{\mu}_2 := \begin{bmatrix} \mu_{21} \\ \vdots \\ \mu_{2m_0} \end{bmatrix} \in \mathbf{R}^{2m_0} \tag{4.7}$$

$$\boldsymbol{\mu} := \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \in \mathbf{R}^{4m_0}. \tag{4.8}$$

Also define the true parameter $\boldsymbol{\beta}$ representing the true locations of the *distinct* prominent points $P_1, \ldots, P_{m_0}$ at time t_1, the angle of rotation θ, and the translation vector $\boldsymbol{T}$ by

$$\boldsymbol{\beta} := \begin{bmatrix} \boldsymbol{\mu}_1 \\ \theta \\ \boldsymbol{T} \end{bmatrix} \in \mathbf{R}^{2m_0+3}. \tag{4.9}$$

Then we can write

$$\boldsymbol{\mu}_2 = g(\boldsymbol{\beta}) \tag{4.10}$$

where $g(\boldsymbol{\beta})$ is defined by

$$g(\boldsymbol{\beta}) := (\mathbf{U}(\theta) \otimes \mathbf{I}_{m_0})\boldsymbol{\mu}_1 + \boldsymbol{T} \otimes \mathbf{1}_{m_0}$$

$$= \begin{bmatrix} \mathbf{U}(\theta) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{U}(\theta) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{U}(\theta) \end{bmatrix} \boldsymbol{\mu}_1 + \begin{bmatrix} \boldsymbol{T} \\ \vdots \\ \boldsymbol{T} \end{bmatrix} \in \mathbf{R}^{2m_0}. \tag{4.11}$$

Hence we have

$$\boldsymbol{\mu} = G(\boldsymbol{\beta}), \tag{4.12}$$

where $G(\boldsymbol{\beta})$ is defined by

$$G(\boldsymbol{\beta}) := \begin{bmatrix} \boldsymbol{\mu}_1 \\ g(\boldsymbol{\beta}) \end{bmatrix} \in \mathbf{R}^{4m_0}. \tag{4.13}$$

The independent Gaussian noise assumption that we made in Subsection 3.3.1 can now be written as

$$\mathcal{L}[\mathbf{X}] = \mathcal{L}[\mathbf{X}\,|\,\boldsymbol{\beta}] = \mathcal{N}(\mathbf{G}(\boldsymbol{\beta}), \sigma^2 \mathbf{I}_{4\mathbf{m}_0}), \tag{4.14}$$

and hence the probability density function $f(\mathbf{x};\boldsymbol{\beta})$ of $\mathbf{X}$ is given by

$$f(\mathbf{x};\boldsymbol{\beta}) := \frac{1}{(2\pi\sigma^2)^{2m_0}} \exp\left[-\frac{1}{2}\frac{(\mathbf{x}-G(\boldsymbol{\beta}))^T(\mathbf{x}-G(\boldsymbol{\beta}))}{\sigma^2}\right]. \tag{4.15}$$

Therefore the log-likelihood function $L(\tilde{\boldsymbol{\beta}})$ of $\mathbf{X}$ is given by

$$\begin{aligned} L(\tilde{\boldsymbol{\beta}}) &:= L(\mathbf{X};\tilde{\boldsymbol{\beta}}) := \log f(\mathbf{X};\tilde{\boldsymbol{\beta}}) \\ &= -2m_0 \log(2\pi\sigma^2) - \frac{1}{2}\frac{(\mathbf{X}-G(\tilde{\boldsymbol{\beta}}))^T(\mathbf{X}-G(\tilde{\boldsymbol{\beta}}))}{\sigma^2} \end{aligned} \tag{4.16}$$

for any $\tilde{\boldsymbol{\beta}} \in \mathbf{R}^{2m_0+3}$.

The maximum likelihood estimate of $\boldsymbol{\beta}$ is defined to be the value $\hat{\boldsymbol{\beta}}_{\mathrm{ML}} := \hat{\boldsymbol{\beta}}_{\mathrm{ML}}(\mathbf{X})$ of $\tilde{\boldsymbol{\beta}}$ which maximizes $f(\mathbf{X};\tilde{\boldsymbol{\beta}})$, or equivalently, $L(\tilde{\boldsymbol{\beta}})$. Note that $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ is a solution of the *likelihood equation*

$$\frac{\partial L(\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}} = 0 \tag{4.17}$$

where $\partial L(\tilde{\boldsymbol{\beta}})/\partial\tilde{\boldsymbol{\beta}}$ is the row gradient vector of $L(\tilde{\boldsymbol{\beta}})$. Also $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ satisfies

$$\hat{\boldsymbol{\beta}}_{\mathrm{ML}}(G(\boldsymbol{\beta})) = \boldsymbol{\beta} \tag{4.18}$$

since we have from (4.16), for any $\tilde{\beta} \neq \beta$,

$$L(G(\beta);\beta) - L(G(\beta);\tilde{\beta}) = \frac{1}{2}\frac{(G(\beta) - G(\tilde{\beta}))^T(G(\beta) - G(\tilde{\beta}))}{\sigma^2} > 0,$$

where the strict inequality follows from Lemma 4.2 to be proved in the next section.

In the next section, we will derive the existence of $\hat{\beta}_{\mathrm{ML}}$, and its asymptotic properties as $\sigma \to 0$.

The proof of a strengthened result to the effect that $\hat{\beta}_{\mathrm{ML}}$ is unique with probability 1 will be deferred to Appendix A.1.

4.2 Asymptotic Normality and Consistency of the Maximum Likelihood Estimate

In this section, we assume that the true σ is known unless otherwise stated.

First note that, from (4.16), we have by differentiation,

$$\frac{\partial L(\tilde{\beta})}{\partial \tilde{\beta}}^T = \frac{1}{\sigma^2}\frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}}^T(\mathbf{X} - G(\tilde{\beta})). \tag{4.19}$$

We begin with the following lemma.

Lemma 4.1 *Let* $\mathbf{X}$ *be an m-dimensional Gaussian random vector whose true distribution is given by*

$$\mathcal{L}[\mathbf{X}] = \mathcal{N}(K(\beta), \Sigma) \tag{4.20}$$

for some $\beta \in \mathbf{R}^d$, *some* m *by* m *positive definite matrix* Σ, *and some function* $K : \mathbf{R}^d \longrightarrow \mathbf{R}^m$. *Let*

$$K(\tilde{\beta}) := \begin{bmatrix} K_1(\tilde{\beta}) \\ \vdots \\ K_m(\tilde{\beta}) \end{bmatrix}$$

for any $\tilde{\beta} \in \mathbf{R}^d$. *Suppose each* $K_i(\tilde{\beta})$ *is twice differentiable with respect to* $\tilde{\beta}$.

Then the log-likelihood function $L(\mathbf{X};\tilde{\boldsymbol\beta})$ of $\mathbf{X}$ satisfies the following.

$$\mathrm{E}\left[\left.\frac{\partial L(\mathbf{X};\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}}\right|_{\tilde{\boldsymbol\beta}=\boldsymbol\beta}^{T}\left.\frac{\partial L(\mathbf{X};\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}}\right|_{\tilde{\boldsymbol\beta}=\boldsymbol\beta}\right] = -\mathrm{E}\left[\left.\frac{\partial^2 L(\mathbf{X};\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}^2}\right|_{\tilde{\boldsymbol\beta}=\boldsymbol\beta}\right] \tag{4.21}$$

$$= -\left.\frac{\partial^2 L(\mathbf{X};\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}^2}\right|_{(\mathbf{X},\tilde{\boldsymbol\beta})=(K(\boldsymbol\beta),\boldsymbol\beta)} . \tag{4.22}$$

$$\tag{4.23}$$

The matrix

$$J := \mathrm{E}\left[\left.\frac{\partial L(\mathbf{X};\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}}\right|_{\tilde{\boldsymbol\beta}=\boldsymbol\beta}^{T}\left.\frac{\partial L(\mathbf{X};\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}}\right|_{\tilde{\boldsymbol\beta}=\boldsymbol\beta}\right] \tag{4.24}$$

is Fisher's information matrix.

Remark: The first relationship (4.21) in the conclusion above, holds under more general regularity conditions on $L(\mathbf{X};\tilde{\boldsymbol\beta})$ which allow the change of the order of differentiation and integration. Also, it is known that the second one, (4.22), holds under somewhat more general circumstances. For details, see [Kendall 77b, pp. 56–58] and [Huzurbazar 49].

Nevertheless, we will write out the proof of Lemma 4.1 under the assumptions given, for the later use.

Proof : Since $\mathcal{L}[\mathbf{X}] = \mathcal{N}(K(\boldsymbol\beta),\Sigma)$, we have

$$L(\mathbf{X};\tilde{\boldsymbol\beta}) := -\tfrac{1}{2}\log((2\pi)^m|\Sigma|) - \tfrac{1}{2}(\mathbf{X}-K(\tilde{\boldsymbol\beta}))^T\Sigma^{-1}(\mathbf{X}-K(\tilde{\boldsymbol\beta})) \tag{4.25}$$

and hence

$$\frac{\partial L(\mathbf{X};\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}}^T = \frac{\partial K(\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}}^T\Sigma^{-1}(\mathbf{X}-K(\tilde{\boldsymbol\beta})). \tag{4.26}$$

Differentiating one more time with respect to $\tilde{\boldsymbol\beta}$ gives us

$$\begin{aligned}\frac{\partial^2 L(\mathbf{X};\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}^2} &= \frac{\partial}{\partial\tilde{\boldsymbol\beta}}\left[\frac{\partial K(\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}}^T\Sigma^{-1}(\mathbf{X}-K(\tilde{\boldsymbol\beta}))\right]\\ &= -\frac{\partial K(\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}}^T\Sigma^{-1}\frac{\partial K(\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}} + \frac{\partial^2 K(\tilde{\boldsymbol\beta})}{\partial\tilde{\boldsymbol\beta}^2}^T\Sigma^{-1}(\mathbf{X}-K(\tilde{\boldsymbol\beta})), \end{aligned} \tag{4.27}$$

where we used the notation $(\partial^2 K(\tilde{\boldsymbol{\beta}})/\partial\tilde{\boldsymbol{\beta}}^2)^T$ to denote the row "vector" with matrix components,

$$\frac{\partial^2 K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}^2}^T := \left[\frac{\partial^2 K_1(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}^2}, \ldots, \frac{\partial^2 K_m(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}^2}\right], \tag{4.28}$$

with the understanding that it be treated just like an ordinary m-dimensional row vector. By substitution, we get

$$\begin{aligned}
J &= \mathrm{E}\left[\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}(\mathbf{X}-K(\boldsymbol{\beta}))(\mathbf{X}-K(\boldsymbol{\beta}))^T\Sigma^{-1}\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}}\right] \\
&= \left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}\mathrm{E}\left[(\mathbf{X}-K(\boldsymbol{\beta}))(\mathbf{X}-K(\boldsymbol{\beta}))^T\right]\Sigma^{-1}\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \\
&= \left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}\mathrm{Cov}[\mathbf{X}]\Sigma^{-1}\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \\
&= \left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}},
\end{aligned} \tag{4.29}$$

and

$$\begin{aligned}
&\mathrm{E}\left[\left.\frac{\partial^2 L(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}^2}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}}\right] \\
&= \mathrm{E}\left[-\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} + \left.\frac{\partial^2 K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}^2}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}(\mathbf{X}-K(\boldsymbol{\beta}))\right] \\
&= -\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} + \left.\frac{\partial^2 K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}^2}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}\mathrm{E}\left[\mathbf{X}-K(\boldsymbol{\beta})\right] \\
&= -\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}},
\end{aligned} \tag{4.30}$$

and

$$\begin{aligned}
&\left.\frac{\partial^2 L(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}^2}\right|_{(\mathbf{X},\tilde{\boldsymbol{\beta}})=(K(\boldsymbol{\beta}),\boldsymbol{\beta})} \\
&= -\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}\left.\frac{\partial K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} + \left.\frac{\partial^2 K(\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}^2}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \Sigma^{-1}(K(\boldsymbol{\beta})-K(\boldsymbol{\beta}))
\end{aligned}$$

$$= -\left.\frac{\partial K(\tilde{\beta})}{\partial \tilde{\beta}}\right|_{\tilde{\beta}=\beta}^{T} \Sigma^{-1} \left.\frac{\partial K(\tilde{\beta})}{\partial \tilde{\beta}}\right|_{\tilde{\beta}=\beta}. \tag{4.31}$$

This completes the proof.

Applying Equation (4.29) in the proof of Lemma 4.1 to our case, namely, when the log-likelihood function is given by (4.16), we get

$$J = \frac{1}{\sigma^2} \left.\frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}}\right|_{\tilde{\beta}=\beta}^{T} \left.\frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}}\right|_{\tilde{\beta}=\beta}. \tag{4.32}$$

We will now derive the existence of the maximum likelihood estimate $\hat{\beta}_{\mathrm{ML}}$ and its consistency as $\sigma \rightarrow 0$ for the Recognized Points problem. To do that, first let

$$\Omega := \mathbf{R}^{2m_0} \times (-\pi, \pi] \times \mathbf{R}^2 \subset \mathbf{R}^{2m_0+3}. \tag{4.33}$$

Then the space Ω, with the topology on $(-\pi, \pi]$ being the *identification topology* with respect to the covering map $c^* : \mathbf{S}^1 \longrightarrow (-\pi, \pi]$, (i.e., a set $V \subset (-\pi, \pi]$ is open if and only if $c^{*-1}(V)$ is open in $\mathbf{S}^1$), is the space of parameters $\tilde{\beta}$ over which we want to maximize the likelihood function $f(\mathbf{X}; \tilde{\beta})$, or the log-likelihood function $L(\tilde{\beta})$.

We need the following lemma to proceed. This lemma states that $G(\tilde{\beta})$ is one-to-one on that subset of Ω corresponding to at least two distinct prominent points. It essentially follows from the fact that the identity is the only rotation about 0 in $\mathbf{R}^2$, which has a non-zero fixed point. (See, for example, [Choquet 69, pp. 59–66].)

Lemma 4.2 *If*

$$\tilde{\beta} := \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\theta} \\ \tilde{T} \end{bmatrix}, \quad \text{and} \quad \tilde{\beta}' := \begin{bmatrix} \tilde{\mu}_1' \\ \tilde{\theta}' \\ \tilde{T}' \end{bmatrix} \in \Omega,$$

and $\tilde{\mu}_1$ represents at least two distinct *prominent points, then $G(\tilde{\beta}) = G(\tilde{\beta}')$ implies that $\tilde{\beta} = \tilde{\beta}'$.*

Proof : Suppose $G(\tilde{\boldsymbol{\beta}}) = G(\tilde{\boldsymbol{\beta}}')$. Then from (4.13) we have

$$\tilde{\boldsymbol{\mu}}_1 = \tilde{\boldsymbol{\mu}}_1', \tag{4.34}$$

and

$$g(\tilde{\boldsymbol{\beta}}) = g(\tilde{\boldsymbol{\beta}}'). \tag{4.35}$$

Then from (4.11), (4.34) and (4.35), we get

$$(\mathbf{U}(\tilde{\theta}) \otimes \mathbf{I}_{m_0})\tilde{\boldsymbol{\mu}}_1 + \tilde{\boldsymbol{T}} \otimes \mathbf{1}_{m_0} = (\mathbf{U}(\tilde{\theta}') \otimes \mathbf{I}_{m_0})\tilde{\boldsymbol{\mu}}_1 + \tilde{\boldsymbol{T}}' \otimes \mathbf{1}_{m_0}. \tag{4.36}$$

Without loss of generality, we can and will assume that $\tilde{\boldsymbol{\mu}}_{11}$ and $\tilde{\boldsymbol{\mu}}_{12}$ are distinct points. Then looking at the first four components of the vectors in (4.36), we get

$$\mathbf{U}(\tilde{\theta})\tilde{\boldsymbol{\mu}}_{11} + \tilde{\boldsymbol{T}} = \mathbf{U}(\tilde{\theta}')\tilde{\boldsymbol{\mu}}_{11} + \tilde{\boldsymbol{T}}', \tag{4.37}$$

and

$$\mathbf{U}(\tilde{\theta})\tilde{\boldsymbol{\mu}}_{12} + \tilde{\boldsymbol{T}} = \mathbf{U}(\tilde{\theta}')\tilde{\boldsymbol{\mu}}_{12} + \tilde{\boldsymbol{T}}'. \tag{4.38}$$

Subtracting (4.38) from (4.37) gives

$$\mathbf{U}(\tilde{\theta})(\tilde{\boldsymbol{\mu}}_{11} - \tilde{\boldsymbol{\mu}}_{12}) = \mathbf{U}(\tilde{\theta}')(\tilde{\boldsymbol{\mu}}_{11} - \tilde{\boldsymbol{\mu}}_{12}),$$

or

$$\mathbf{U}(\tilde{\theta}')^{-1}\mathbf{U}(\tilde{\theta})(\tilde{\boldsymbol{\mu}}_{11} - \tilde{\boldsymbol{\mu}}_{12}) = \tilde{\boldsymbol{\mu}}_{11} - \tilde{\boldsymbol{\mu}}_{12},$$

and hence

$$\mathbf{U}(\tilde{\theta} - \tilde{\theta}')(\tilde{\boldsymbol{\mu}}_{11} - \tilde{\boldsymbol{\mu}}_{12}) = \tilde{\boldsymbol{\mu}}_{11} - \tilde{\boldsymbol{\mu}}_{12}.$$

Since $\tilde{\boldsymbol{\mu}}_{11} \neq \tilde{\boldsymbol{\mu}}_{12}$, we must have

$$\mathbf{U}(\tilde{\theta} - \tilde{\theta}') = \mathbf{I}_2$$

and thus

$$\tilde{\theta} = \tilde{\theta}'. \tag{4.39}$$

Then from (4.37) and (4.39), we get

$$\tilde{T} = \tilde{T}'. \tag{4.40}$$

Therefore from (4.34), (4.39) and (4.40), we can conclude that

$$\tilde{\beta} = \tilde{\beta}'.$$

This completes the proof.

We will state without proof, the following theorem which proves the existence of the maximum likelihood estimate based on n *i.i.d.* observations, and its consistency as the sample size $n \to \infty$, under certain conditions. Then after that, we will show how the theorem applies to the maximum likelihood estimate $\hat{\beta}_{\mathrm{ML}}$ in the Recognized Points problem. For a proof see [Chernoff 79, pp. 52–54].

Theorem 4.1 *Suppose that it is possible to extend the space of parameters, Ω, to a compact set Ω^* such that*

(i) *for every $\tilde{\beta} \in \Omega^*$, $P_{\tilde{\beta}}$ is a* subdistribution *with density $f(\mathbf{x};\tilde{\beta})$,* (i.e., $f(\mathbf{x};\tilde{\beta}) \geq 0$ *and* $\int f(\mathbf{x};\tilde{\beta})\,d\mathbf{x} \leq 1$), *which is continuous with respect to $\tilde{\beta}$,*

(ii) *for every $\tilde{\beta} \in \Omega^*$, there is a neighborhood $N_{\tilde{\beta}}$ of $\tilde{\beta}$ such that*

$$\mathrm{E}_{\beta}\left[\inf_{\check{\beta} \in N_{\tilde{\beta}}} \log\left[\frac{f(\mathbf{x};\beta)}{f(\mathbf{x};\check{\beta})}\right]\right] > -\infty,$$

and

(iii) *for every $\tilde{\beta} \in \Omega^* \setminus \{\beta\}$, $P_{\tilde{\beta}} \neq P_{\beta}$.*

Then, for every neighborhood N_{β} of β, $\hat{\beta}_n$, the value of $\tilde{\beta} \in \Omega^*$ which maximizes the likelihood based on the first n observations, exists and satisfies

$$P_{\beta}\{\hat{\beta}_n \notin N_{\beta} \text{ infinitely often}\} = 0$$

and

$$P_{\boldsymbol{\beta}}\{\hat{\boldsymbol{\beta}}_n \notin N_{\boldsymbol{\beta}}\} \rightarrow 0 \quad \textit{as} \quad n \rightarrow \infty. \tag{4.41}$$

First note that, in the Gaussian case, Theorem 4.1, which deals with the maximum likelihood estimate when the sample size $n \rightarrow \infty$, can also be applied to the situation where we have a fixed sample size but $\sigma \rightarrow 0$, for the following reason, as mentioned in [Johansen 84, p. 77]. If $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ are independently identically distributed with

$$\mathcal{L}[\mathbf{X}^{(i)}] = \mathcal{N}(\boldsymbol{\mu}, \Sigma), \qquad i = 1, \ldots, n,$$

then

$$\overline{\mathbf{X}} := \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}^{(i)}$$

is a sufficient statistic for $\boldsymbol{\mu}$ and

$$\mathcal{L}[\overline{\mathbf{X}}] = \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \Sigma)$$

where

$$\sigma^2 := \frac{1}{n} \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty. \tag{4.42}$$

Thus if we interpret the role of n to be that of σ^{-2}, the Gaussian model makes the two maximum likelihood problems equivalent.

Now we will show that our model, defined by the log-likelihood function $L(\tilde{\boldsymbol{\beta}})$ in (4.16) with the parameter space Ω in (4.33), satisfies the condtions of Theorem 4.1, and as a result, that $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ exists and is consistent. To do this, let us first note that Ω is *locally compact* and *Hausdorff* so that we can add a new point, ∞, to it and get a new space Ω^* that is compact ([Royden 68, p. 168]). In Ω^*, a subset is open if it is either an open subset of Ω or the complement of a compact subset of Ω. Define the subdistribution at $\infty \in \Omega^*$ by taking

$$f(\mathbf{x}; \infty) := 0 \quad \text{for any } \mathbf{x} \in \mathbf{R}^{4m_0}. \tag{4.43}$$

Then condition (i) of Theorem 4.1 follows from the fact that

$$\lim_{\tilde{\boldsymbol{\beta}} \to \infty} f(\mathbf{x};\tilde{\boldsymbol{\beta}}) = 0 = f(\mathbf{x};\infty).$$

Note that for any $\tilde{\boldsymbol{\beta}} \in \Omega$,

$$\begin{aligned}
\log \frac{f(\mathbf{x};\boldsymbol{\beta})}{f(\mathbf{x};\tilde{\boldsymbol{\beta}})} &= -\frac{1}{2\sigma^2}\left[(\mathbf{x}-G(\boldsymbol{\beta}))^T(\mathbf{x}-G(\boldsymbol{\beta})) - (\mathbf{x}-G(\tilde{\boldsymbol{\beta}}))^T(\mathbf{x}-G(\tilde{\boldsymbol{\beta}}))\right] \\
&\geq -\frac{1}{2\sigma^2}\left[(\mathbf{x}-G(\boldsymbol{\beta}))^T(\mathbf{x}-G(\boldsymbol{\beta}))\right].
\end{aligned}$$

For $\tilde{\boldsymbol{\beta}} = \infty$, the above log is intrerpreted to be $+\infty$ and the inequality still holds. thus

$$\mathrm{E}\left[\inf_{\tilde{\boldsymbol{\beta}} \in \Omega^*} \log \frac{f(\mathbf{X};\boldsymbol{\beta})}{f(\mathbf{X};\tilde{\boldsymbol{\beta}})}\right] \geq -\frac{1}{2\sigma^2}\mathrm{E}\left[\sigma^2\chi^2_{4m_0}\right] = -2m_0 > -\infty, \tag{4.44}$$

and condition (ii) follows. Finally condition (iii) is satisfied if $\boldsymbol{\beta}$ corresponds to at least two distinct prominent points, since it is obvious that $P_\infty \neq P_{\boldsymbol{\beta}}$, and for every $\tilde{\boldsymbol{\beta}} \in \Omega \setminus \{\boldsymbol{\beta}\}$, the mean $G(\tilde{\boldsymbol{\beta}})$ uniquely determines $P_{\tilde{\boldsymbol{\beta}}}$, and thus by Lemma 4.2, if $\tilde{\boldsymbol{\beta}} \neq \boldsymbol{\beta}$, $G(\tilde{\boldsymbol{\beta}}) \neq G(\boldsymbol{\beta})$ and $P_{\tilde{\boldsymbol{\beta}}} \neq P_{\boldsymbol{\beta}}$. Therefore all the conditions of Theorem 4.1 are satisfied and hence we have the existence and consistency of the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$, namely,

Theorem 4.2 *If $\boldsymbol{\beta}$ corresponds to at least two distinct prominent points* ($m_0 \geq 2$), *then $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ exists, and*

$$\hat{\boldsymbol{\beta}}_{\mathrm{ML}} \to \boldsymbol{\beta} \quad \textit{in probability} \quad \textit{as} \quad \sigma \to 0. \tag{4.45}$$

Note that $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ cannot be ∞, since

$$f(\mathbf{X};\tilde{\boldsymbol{\beta}}) > 0 = f(\mathbf{X};\infty)$$

for any $\tilde{\boldsymbol{\beta}} \in \Omega$.

Now we need the following theorem to derive the asymptotic normality of $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$. For a proof, again see [Chernoff 79, pp. 19–21].

Theorem 4.3 *Suppose that the following assumptions hold.*

(i) $\hat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}$ *in probability as* $n \to \infty$,

(ii) $\mathrm{E}_{\boldsymbol{\beta}}\left[\left.\frac{\partial \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}}\right] = 0$,

(iii) $\mathrm{E}_{\boldsymbol{\beta}}\left[\left.f^{-1}(\mathbf{X};\tilde{\boldsymbol{\beta}})\frac{\partial^2 f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}^2}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}}\right] = 0$,

(iv) $J_1(\boldsymbol{\beta})$ *is positive definite,*

and

(v) $\mathrm{E}_{\boldsymbol{\beta}}[W(\delta)] \to 0$ *as* $\delta \to 0$,

where -

$$J_1(\boldsymbol{\beta}) := \mathrm{E}_{\boldsymbol{\beta}}\left[\left.\frac{\partial \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \left.\frac{\partial \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}}\right], \tag{4.46}$$

$$W(\delta) := \sup_{\left\|\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|<\delta} \left\| \frac{\partial^2 \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}^2} - \left.\frac{\partial^2 \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}^2}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \right\|_{\sup} \tag{4.47}$$

and $\|\cdot\|_{\sup}$ *represents the largest component of the matrix.*

Then we have

$$\mathcal{L}\left[\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\right] \to \mathcal{N}\left(0, J_1^{-1}(\boldsymbol{\beta})\right) \quad \text{as } n \to \infty. \tag{4.48}$$

Again by the same argument as in the paragraph immediately following Theorem 4.1, we can apply Theorem 4.3 to our small σ situation. The only exception is that the conclusion (4.48) should be translated into

$$\mathcal{L}\left[\frac{\hat{\boldsymbol{\beta}}_{\mathrm{ML}} - \boldsymbol{\beta}}{\sigma}\right] \to \mathcal{N}\left(0, \sigma^{-2}J^{-1}\right) \quad \text{as } \sigma \to 0, \tag{4.49}$$

since $J_1(\boldsymbol{\beta})$ is Fisher's information matrix corresponding to the likelihood function of a sample $\mathbf{X}$ of size 1, and for a sample of size n, the corresponding information matrix $J_n(\boldsymbol{\beta})$ becomes $J_n(\boldsymbol{\beta}) = nJ_1(\boldsymbol{\beta})$ so that $J_1^{-1}(\boldsymbol{\beta}) = nJ_n^{-1}(\boldsymbol{\beta})$, which corresponds to $\sigma^{-2}J^{-1}$.

Now we will show that the conditions of Theorem 4.3 are satisfied by $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ if $\boldsymbol{\beta}$ corresponds to at least two distinct prominent points ($m_0 \geq 2$). First, condition (i) follows from Theorem 4.2. Condition (ii) follows from (4.19) and the fact that $\mathrm{E}_{\boldsymbol{\beta}}[\mathbf{X}] = G(\boldsymbol{\beta})$. To show that condition (iii) is satisfied, let us first note the following identity:

$$\frac{\partial f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}} = f(\mathbf{X};\tilde{\boldsymbol{\beta}}) \frac{\partial \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}. \tag{4.50}$$

By differentiating both sides of (4.50) with respect to $\tilde{\boldsymbol{\beta}}$, we get

$$\begin{aligned} &\left.\frac{\partial^2 f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}^2}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \\ &= f(\mathbf{X};\boldsymbol{\beta}) \left[\left.\frac{\partial \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \left.\frac{\partial \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \right. \\ &\qquad \left. + \left.\frac{\partial^2 \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}^2}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \right]. \end{aligned} \tag{4.51}$$

Thus by dividing both sides of (4.51) by $f(\mathbf{X};\boldsymbol{\beta})$ and then by taking expectations under $\boldsymbol{\beta}$, we get

$$\begin{aligned} &\mathrm{E}_{\boldsymbol{\beta}} \left[f^{-1}(\mathbf{X};\tilde{\boldsymbol{\beta}}) \left.\frac{\partial^2 f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}^2}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \right] \\ &= \mathrm{E}_{\boldsymbol{\beta}} \left[\left.\frac{\partial \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \left.\frac{\partial \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \right] \\ &\qquad + \mathrm{E}_{\boldsymbol{\beta}} \left[\left.\frac{\partial^2 \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}^2}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \right]. \end{aligned}$$

However the first term on the right is J and the second is $-J$ from (4.21) and (4.24) in Lemma 4.1. Therefore condition (iii) is satisfied.

The proof that condition (iv) is satisfied, i.e., that J is positive definite, will be given in Lemma 4.3 in the next section.

Now it only remains to show that condition (v) is satisfied. To do so, it suffices to show that the third derivatives

$$Q(\mathbf{X};\tilde{\boldsymbol{\beta}};i,j,k) := \frac{\partial^3 \log f(\mathbf{X};\tilde{\boldsymbol{\beta}})}{\partial\tilde{\boldsymbol{\beta}}_i \partial\tilde{\boldsymbol{\beta}}_j \partial\tilde{\boldsymbol{\beta}}_k}, \quad i,j,k = 1,\ldots,2m_0+3 \tag{4.52}$$

are uniformly bounded in the region $\left\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| < \delta$ by a function $H(\mathbf{X})$ whose expectation under $\boldsymbol{\beta}$ is finite. Indeed, from Equation (4.27) with $K := G$, $\Sigma := \sigma^2$, it is easy to see that $Q(\mathbf{X};\tilde{\boldsymbol{\beta}};i,j,k)$ in (4.52) is a polynomial in the components of $\mathbf{X}$, $G(\tilde{\boldsymbol{\beta}})$, $\partial G(\tilde{\boldsymbol{\beta}})/\partial\tilde{\boldsymbol{\beta}}_i$, $\partial G(\tilde{\boldsymbol{\beta}})/\partial\tilde{\boldsymbol{\beta}}_j$, $\partial G(\tilde{\boldsymbol{\beta}})/\partial\tilde{\boldsymbol{\beta}}_k$, $\partial^2 G(\tilde{\boldsymbol{\beta}})/\partial\tilde{\boldsymbol{\beta}}_i\partial\tilde{\boldsymbol{\beta}}_j$, $\partial^2 G(\tilde{\boldsymbol{\beta}})/\partial\tilde{\boldsymbol{\beta}}_j\partial\tilde{\boldsymbol{\beta}}_k$, $\partial^2 G(\tilde{\boldsymbol{\beta}})/\partial\tilde{\boldsymbol{\beta}}_i\partial\tilde{\boldsymbol{\beta}}_k$, and $\partial^3 G(\tilde{\boldsymbol{\beta}})/\partial\tilde{\boldsymbol{\beta}}_i\partial\tilde{\boldsymbol{\beta}}_j\partial\tilde{\boldsymbol{\beta}}_k$ of degree two, and hence uniformly bounded in the region $\left\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| < \delta$, for all $i,j,k = 1,\ldots,2m_0+3$, by a function $H(\mathbf{X})$ of the form

$$H(\mathbf{X}) := \|\mathbf{a}(\delta)\|\,\|\mathbf{X}\| + \|\mathbf{b}(\delta)\|$$

for some $\mathbf{a}(\delta)$ and $\mathbf{b}(\delta) \in \mathbf{R}^{4m_0}$. Then

$$\begin{aligned} \mathrm{E}_{\boldsymbol{\beta}}[H(\mathbf{X})] &\le \|\mathbf{a}(\delta)\|\,(\|G(\boldsymbol{\beta})\| + 4m_0\sigma\mathrm{E}[|\varepsilon|]) + \|\mathbf{b}(\delta)\| \quad \text{where} \quad \mathcal{L}[\varepsilon] = \mathcal{N}(0,1) \\ &= \|\mathbf{a}(\delta)\|\,(\|G(\boldsymbol{\beta})\| + \frac{4\sqrt{2}}{\sqrt{\pi}}m_0\sigma) + \|\mathbf{b}(\delta)\| < \infty. \end{aligned}$$

Therefore all the conditions of Theorem 4.3 are satisfied and we have, except for Lemma 4.3, proved the asymptotic normality of the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$.

Summarizing all these, we have proved the following theorem.

Theorem 4.4 *Let $L(\mathbf{X};\tilde{\boldsymbol{\beta}})$ be the log-likelihood function defined by* (4.16) *on* Ω *which is given by* (4.33). *Let* $\boldsymbol{\beta} := \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\theta} \\ \boldsymbol{T} \end{bmatrix}$ *be the true value of* $\tilde{\boldsymbol{\beta}} \in \Omega$ *where* $\boldsymbol{\mu}_1$ *represents at least two distinct prominent points* ($m_0 \ge 2$).

Then the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}(\mathbf{X})$ *which maximizes* $L(\mathbf{X};\tilde{\boldsymbol{\beta}})$ *exists and satisfies*

$$\hat{\boldsymbol{\beta}}_{\mathrm{ML}}(\mathbf{X}) \to \boldsymbol{\beta} \quad \textit{in probability} \quad \textit{as } \sigma \to 0. \tag{4.53}$$

and

$$\mathcal{L}\left[\frac{\hat{\beta}_{\mathrm{ML}}(\mathbf{X})-\beta}{\sigma}\right] \to \mathcal{N}\left(0, \sigma^{-2}J^{-1}\right) \quad \textit{as} \;\; \sigma \to 0. \tag{4.54}$$

Remark: Note that from Equation (4.32)

$$\sigma^{-2}J^{-1} = \left[\left.\frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}}\right|^{T}_{\tilde{\beta}=\beta} \left.\frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}}\right|_{\tilde{\beta}=\beta}\right]^{-1},$$

and is independent of σ.

It is easy to see that with probability $\to 1$, the second derivative of the log-likelihood function $L(\tilde{\beta})$ is continuous and negative definite in a neighborhood of β and hence with probability $\to 1$, the likelihood equation (4.17) has a unique solution in this neighborhood as $\sigma \to 0$. (See, for example, [Foutz 77].) However, as mentioned earlier, in the last section, a stronger result, to the effect that the maximum likelihood estimate is unique wth probability 1 can be proved. That proof will be deferred to Appendix A.1.

Now Theorem 4.4 allows us the approxmation

$$\mathcal{L}\left[\hat{\beta}_{\mathrm{ML}}(\mathbf{X})-\beta\right] \approx \mathcal{N}\left(0, J^{-1}\right) \tag{4.55}$$

for a sufficently small $\sigma > 0$, and this provides us with J^{-1} as a measure of performance with which our estimators $\hat{\theta}$ and $\hat{T}$ can be compared.

4.3 Fisher's Information Matrix J and Its Inverse—Asymptotic Covariance Matrix

In this section, explicit formulae for Fisher's information matrix J and its inverse J^{-1} will be derived. The results will be used to estimate the approximate standard deviations of the maximum likelihood estimate discussed so far, and then we will be able to compare them with the ones from the sample distributions of the proposed estimator obtained by simulation. That task will be done in the next chapter.

First, let us recall from Equation (4.16) that

$$L(\mathbf{X};\tilde{\boldsymbol{\beta}}) = -2m_0\log(2\pi\sigma^2) - \frac{1}{2}\frac{(\mathbf{X}-G(\tilde{\boldsymbol{\beta}}))^T(\mathbf{X}-G(\tilde{\boldsymbol{\beta}}))}{\sigma^2}$$

where

$$\tilde{\boldsymbol{\beta}} := \begin{bmatrix} \tilde{\boldsymbol{\mu}}_1 \\ \tilde{\theta} \\ \tilde{\boldsymbol{T}} \end{bmatrix} \in \Omega \subset \mathbf{R}^{2m_0+3}, \tag{4.56}$$

$$\begin{aligned} G(\tilde{\boldsymbol{\beta}}) &:= \begin{bmatrix} \tilde{\boldsymbol{\mu}}_1 \\ (\mathbf{U}(\tilde{\theta})\otimes \mathbf{I}_{m_0})\tilde{\boldsymbol{\mu}}_1 + \tilde{\boldsymbol{T}}\otimes \mathbf{1}_{m_0} \end{bmatrix} \\ &= \begin{bmatrix} \tilde{\boldsymbol{\mu}}_1 \\ \begin{bmatrix} \mathbf{U}(\tilde{\theta}) & 0 & \cdots & 0 \\ 0 & \mathbf{U}(\tilde{\theta}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{U}(\tilde{\theta}) \end{bmatrix} \tilde{\boldsymbol{\mu}}_1 + \begin{bmatrix} \tilde{\boldsymbol{T}} \\ \vdots \\ \tilde{\boldsymbol{T}} \end{bmatrix} \end{bmatrix} \in \mathbf{R}^{4m_0}. \end{aligned} \tag{4.57}$$

By differentiating both sides of Equation (4.57), we have

$$\frac{\partial G(\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}} = \begin{bmatrix} \mathbf{I}_{2m_0} & 0_{2m_0\times 1} & 0_{2m_0\times 2} \\ \mathbf{U}(\tilde{\theta})\otimes \mathbf{I}_{m_0} & \left(\mathbf{U}(\frac{\pi}{2}+\tilde{\theta})\otimes \mathbf{I}_{m_0}\right)\tilde{\boldsymbol{\mu}}_1 & \mathbf{I}_2\otimes \mathbf{1}_{m_0} \end{bmatrix} \tag{4.58}$$

and hence from Equation (4.32) we have

$$\begin{aligned} J &= \frac{1}{\sigma^2}\left.\frac{\partial G(\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}\right|^T_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \left.\frac{\partial G(\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}}\right|_{\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}} \\ &= \frac{1}{\sigma^2}\begin{bmatrix} \mathbf{I}_{2m_0} & \mathbf{U}(\theta)^T\otimes \mathbf{I}_{m_0} \\ 0_{1\times 2m_0} & \boldsymbol{\mu}_1^T\left(\mathbf{U}(\frac{\pi}{2}+\theta)^T\otimes \mathbf{I}_{m_0}\right) \\ 0_{2\times 2m_0} & \mathbf{I}_2\otimes \mathbf{1}^T_{m_0} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{2m_0} & 0_{2m_0\times 1} & 0_{2m_0\times 2} \\ \mathbf{U}(\theta)\otimes \mathbf{I}_{m_0} & \left(\mathbf{U}(\frac{\pi}{2}+\theta)\otimes \mathbf{I}_{m_0}\right)\boldsymbol{\mu}_1 & \mathbf{I}_2\otimes \mathbf{1}_{m_0} \end{bmatrix} \\ &= \frac{1}{\sigma^2}\begin{bmatrix} \mathbf{I}_{2m_0}+(\mathbf{U}(\theta)^T\mathbf{U}(\theta))\otimes \mathbf{I}_{m_0} & ((\mathbf{U}(\theta)^T\mathbf{U}(\frac{\pi}{2}+\theta))\otimes \mathbf{I}_{m_0})\boldsymbol{\mu}_1 & \mathbf{U}(\theta)^T\otimes \mathbf{1}_{m_0} \\ \boldsymbol{\mu}_1^T((\mathbf{U}(\frac{\pi}{2}+\theta)^T\mathbf{U}(\theta))\otimes \mathbf{I}_{m_0}) & \boldsymbol{\mu}_1^T((\mathbf{U}(\frac{\pi}{2}+\theta)^T\mathbf{U}(\frac{\pi}{2}+\theta))\otimes \mathbf{I}_{m_0})\boldsymbol{\mu}_1 & \boldsymbol{\mu}_1^T(\mathbf{U}(\frac{\pi}{2}+\theta)^T\otimes \mathbf{1}_{m_0}) \\ \mathbf{U}(\theta)\otimes \mathbf{1}^T_{m_0} & (\mathbf{U}(\frac{\pi}{2}+\theta)\otimes \mathbf{1}^T_{m_0})\boldsymbol{\mu}_1 & \mathbf{I}_2\otimes(\mathbf{1}^T_{m_0}\mathbf{1}_{m_0}) \end{bmatrix}. \end{aligned} \tag{4.59}$$

By simplifying the last matrix above, we get

$$J = \frac{1}{\sigma^2}\begin{bmatrix} 2\mathbf{I}_{2m_0} & (\mathbf{U}(\frac{\pi}{2})\otimes \mathbf{I}_{m_0})\boldsymbol{\mu}_1 & \mathbf{U}(\theta)^T\otimes \mathbf{1}_{m_0} \\ \boldsymbol{\mu}_1^T(\mathbf{U}(\frac{\pi}{2})^T\otimes \mathbf{I}_{m_0}) & \boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 & \boldsymbol{\mu}_1^T(\mathbf{U}(\frac{\pi}{2}+\theta)^T\otimes \mathbf{1}_{m_0}) \\ \mathbf{U}(\theta)\otimes \mathbf{1}^T_{m_0} & (\mathbf{U}(\frac{\pi}{2}+\theta)\otimes \mathbf{1}^T_{m_0})\boldsymbol{\mu}_1 & m_0\mathbf{I}_2 \end{bmatrix}, \tag{4.60}$$

and by expanding the submatrices out, we get

$$J = \frac{1}{\sigma^2}\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{bmatrix}, \tag{4.61}$$

where **A**, **B**, and **D** are $2m_0$ by $2m_0$, $2m_0$ by 3, and 3 by 3 matrices respectively, given by

$$\mathbf{A} := \begin{bmatrix} 2 & 0 & \cdots & 0 \\ 0 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 2 \end{bmatrix}, \tag{4.62}$$

$$\mathbf{B} := \begin{bmatrix} -y_1 & \cos\theta & \sin\theta \\ x_1 & -\sin\theta & \cos\theta \\ \vdots & \vdots & \vdots \\ -y_{m_0} & \cos\theta & \sin\theta \\ x_{m_0} & -\sin\theta & \cos\theta \end{bmatrix}, \tag{4.63}$$

and

$$\mathbf{D} := \begin{bmatrix} \sum_{i=1}^{m_0}(x_i^2+y_i^2) & -\sum_{i=1}^{m_0}x_i\sin\theta - \sum_{i=1}^{m_0}y_i\cos\theta & \sum_{i=1}^{m_0}x_i\cos\theta - \sum_{i=1}^{m_0}y_i\sin\theta \\ -\sum_{i=1}^{m_0}x_i\sin\theta - \sum_{i=1}^{m_0}y_i\cos\theta & m_0 & 0 \\ \sum_{i=1}^{m_0}x_i\cos\theta - \sum_{i=1}^{m_0}y_i\sin\theta & 0 & m_0 \end{bmatrix}. \tag{4.64}$$

This relatively simple partitioned structure of the matrix will allow us to compute the inverse without much difficulty. Before we try to invert the information matrix J, let us first prove that it is positive definite and hence invertible (non-singular). We prove it in the following lemma.

Lemma 4.3 *If $\boldsymbol{\beta}$ corresponds to at least two distinct prominent points* ($m_0 \geq 2$), *then Fisher's information matrix defined by* (4.24) *is positive definite, i.e., for any* $\mathbf{z} \in \mathbf{R}^{2m_0+3}$,

$$\mathbf{z}^T J \mathbf{z} \geq 0$$

and

$$\mathbf{z}^T J \mathbf{z} = 0 \quad \textit{if and only if} \quad \mathbf{z} = 0.$$

Proof : First note that for any $\mathbf{z} \in \mathbf{R}^{2m_0+3}$,

$$\begin{aligned}
\mathbf{z}^T J \mathbf{z} &= \mathbf{z}^T \left[\frac{1}{\sigma^2} \frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}} \bigg|_{\tilde{\beta}=\beta}^T \frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}} \bigg|_{\tilde{\beta}=\beta} \right] \mathbf{z} \\
&= \frac{1}{\sigma^2} \left[\frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}} \bigg|_{\tilde{\beta}=\beta} \mathbf{z} \right]^T \left[\frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}} \bigg|_{\tilde{\beta}=\beta} \mathbf{z} \right] \\
&= \frac{1}{\sigma^2} \left\| \frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}} \bigg|_{\tilde{\beta}=\beta} \mathbf{z} \right\|^2 \qquad (4.65) \\
&\geq 0. \qquad (4.66)
\end{aligned}$$

Now assume that

$$\mathbf{z}^T J \mathbf{z} = 0$$

Then from (4.65), we have

$$\frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}} \bigg|_{\tilde{\beta}=\beta} \mathbf{z} = 0$$

where

$$\frac{\partial G(\tilde{\beta})}{\partial \tilde{\beta}} \bigg|_{\tilde{\beta}=\beta} = \begin{bmatrix} \mathbf{I}_{2m_0} & \mathbf{0}_{2m_0 \times 1} & \mathbf{0}_{2m_0 \times 2} \\ \mathbf{U}(\theta) \otimes \mathbf{I}_{m_0} & \left(\mathbf{U}(\frac{\pi}{2} + \theta) \otimes \mathbf{I}_{m_0} \right) \mu_1 & \mathbf{I}_2 \otimes \mathbf{1}_{m_0} \end{bmatrix}$$

from Equation (4.58). Let

$$\mathbf{z} := [z_1, \ldots, z_{2m_0}, a, b, c]^T.$$

Then it follows that

$$\mathbf{0} = \begin{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_{2m_0} \end{bmatrix} \\ (\mathbf{U}(\theta) \otimes \mathbf{I}_{m_0}) \begin{bmatrix} z_1 \\ \vdots \\ z_{2m_0} \end{bmatrix} + (\mathbf{U}(\frac{\pi}{2} + \theta) \otimes \mathbf{I}_{m_0}) \mu_1 a + (\mathbf{I}_2 \otimes \mathbf{1}_{m_0}) \begin{bmatrix} b \\ c \end{bmatrix} \end{bmatrix}. \qquad (4.67)$$

Clearly $[z_1, \ldots, z_{2m_0}] = 0$, and if $a \neq 0$, then we get

$$(\mathbf{U}(\frac{\pi}{2} + \theta) \otimes \mathbf{I}_{m_0}) \mu_1 = -a^{-1} (\mathbf{I}_2 \otimes \mathbf{1}_{m_0}) \begin{bmatrix} b \\ c \end{bmatrix}, \qquad (4.68)$$

and hence

$$\mu_1 = -a^{-1}(\mathbf{U}(\frac{\pi}{2}+\theta)^T \otimes \mathbf{1}_{m_0}) \begin{bmatrix} b \\ c \end{bmatrix} \tag{4.69}$$

which implies that, for all $i = 1, \ldots, m_0$,

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = -a^{-1}\mathbf{U}(\frac{\pi}{2}+\theta)^T \begin{bmatrix} b \\ c \end{bmatrix}. \tag{4.70}$$

i.e., all the (x_i, y_i)'s are the the same and this is a contradiction. So $a = 0$ and it follows from Equation (4.67) that $[b, c] = 0$. Therefore $\mathbf{z} = \mathbf{0}$. This completes the proof.

Now we will use the easily verified fact that, since J is non-singular,

$$J^{-1} = \sigma^2 \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^T & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}^T & \mathbf{E}^{-1} \end{bmatrix}, \tag{4.71}$$

where

$$\mathbf{E} := \mathbf{D} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}, \tag{4.72}$$

$$\mathbf{F} := \mathbf{A}^{-1}\mathbf{B}. \tag{4.73}$$

Since we are only interested in the 3 by 3 portion $J^{-1}(\theta, \boldsymbol{T})$ of the asymptotic covariance matrix J^{-1} corresponding to the parameters θ and $\boldsymbol{T}$, we only have to compute $\sigma^2\mathbf{E}^{-1}$ where $\mathbf{E}$ is defined in (4.72).

By a long but routine calculation, we get the following symmetric matrix, of which only the upper diagonal elements are shown.

$$J^{-1}(\theta, \boldsymbol{T}) := \sigma^2\mathbf{E}^{-1}$$

$$= \frac{2\sigma^2}{m_0(\sigma_x^2+\sigma_y^2)} \begin{bmatrix} 1 & \mu_x \sin\theta + \mu_y \cos\theta & -(\mu_x \cos\theta - \mu_y \sin\theta) \\ & (\sigma_x^2+\sigma_y^2) + (\mu_x \sin\theta + \mu_y \cos\theta)^2 & -(\mu_x \sin\theta + \mu_y \cos\theta)(\mu_x \cos\theta - \mu_y \sin\theta) \\ & & (\sigma_x^2+\sigma_y^2) + (\mu_x \cos\theta - \mu_y \sin\theta)^2 \end{bmatrix} \tag{4.74}$$

where

$$\mu_x := \frac{1}{m_0}\sum_{i=1}^{m_0} x_i \tag{4.75}$$

$$\mu_y := \frac{1}{m_0}\sum_{i=1}^{m_0} y_i \tag{4.76}$$

$$\sigma_x^2 := \frac{1}{m_0}\sum_{i=1}^{m_0} (x_i - \mu_x)^2 \tag{4.77}$$

$$\sigma_y^2 := \frac{1}{m_0}\sum_{i=1}^{m_0} (y_i - \mu_y)^2. \tag{4.78}$$

In particular, when $\theta = 0$, as will be the case in our simulation study, the matrix (4.74) becomes

$$J^{-1}(\theta, T) = \frac{2\sigma^2}{m_0(\sigma_x^2 + \sigma_y^2)} \begin{bmatrix} 1 & \mu_y & -\mu_x \\ \mu_y & (\sigma_x^2 + \sigma_y^2) + \mu_y^2 & -\mu_x\mu_y \\ -\mu_x & -\mu_x\mu_y & (\sigma_x^2 + \sigma_y^2) + \mu_x^2 \end{bmatrix}. \tag{4.79}$$

We will refer to the diagonal entries of this matrix by $\hat{\sigma}^2_{\mathrm{ML}}$ for use in the next chapter.

Now we will discuss the case when σ is incorrectly assumed to be $\sigma_a = k\sigma$ for positive $k \neq 1$ in constructing the likelihood function. We will distinguish this case from the one discussed so far by attaching a subscript "a" to the notation that has been in use. Hence the log-likelihood function of this case will be denoted by $L_a(\tilde{\beta})$ instead of $L(\tilde{\beta})$, where

$$L_a(\tilde{\beta}) := -2m_0 \log(2\pi\sigma_a^2) - \frac{1}{2}\frac{(\mathrm{X} - G(\tilde{\beta}))^T(\mathrm{X} - G(\tilde{\beta}))}{\sigma_a^2}.$$

Then it is clear that, for both $L(\tilde{\beta})$ and $L_a(\tilde{\beta})$, maximizing them is equivalent to minimizing $(\mathrm{X} - G(\tilde{\beta}))^T(\mathrm{X} - G(\tilde{\beta}))$. Hence

$$\hat{\beta}_{\mathrm{ML}a} = \hat{\beta}_{\mathrm{ML}}, \tag{4.80}$$

and all the results derived so far for $\hat{\beta}_{\mathrm{ML}}$ are also valid for $\hat{\beta}_{\mathrm{ML}a}$. In particular, $\hat{\beta}_{\mathrm{ML}a}$ will have the same asymptotic distribution as $\hat{\beta}_{\mathrm{ML}}$, as the true value of $\sigma \to 0$, and J^{-1} will be the asymptotic covariance of it regardless of the value of $k > 0$.

Chapter 5

Simulation Study

To compare the performance of the proposed estimator empirically with that of the maximum likelihood estimate discussed in the last chapter, simulations of two separate cases have been carried out by generating random prominent points, Gaussian noise, and observed locations of points in two image frames.

As the pseudo-random number generator, we used the one developed by Wichmann and Hill [Wichmann 82], and as the Gaussian random number generator, we used the one by Box and Muller [Box 58].

5.1 Method of Simulation

Without loss of generality, we assumed that the true locations of the prominent points are within the open *unit disc*,

$$\{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 < 1\},$$

and also that there was no motion involved, *i.e.*, $\theta = 0$, $\boldsymbol{T} = 0$, for in view of the invariance of the method we are employing, the problem of estimating no motion is the same as the one of estimating a non-trivial motion.

In each of the two cases, we generated $n = 10$ independent random points (x_i, y_i), $i = 1, \ldots, 10$, uniformly distributed on the unit disc, by taking

$$(x_i, y_i) := (\sqrt{U_{i1}} \cos(2\pi U_{i2}), \sqrt{U_{i1}} \sin(2\pi U_{i2})), \tag{5.1}$$

where U_{i1}, U_{i2} are independent uniform random variables on $(0, 1)$. This is legitimate because the density function of the point (x_i, y_i) is then, by change of variables,

$$\begin{aligned}
&\text{the density function of } (U_{i1}, U_{i2}) \Big/ \left| \frac{\partial(x_i, y_i)}{\partial(U_{i1}, U_{i2})} \right| \\
= \;& 1 \Big/ \left| \begin{array}{cc} \frac{1}{2\sqrt{U_{i1}}} \cos(2\pi U_{i2}) & -2\pi\sqrt{U_{i1}} \sin(2\pi U_{i2}) \\ \frac{1}{2\sqrt{U_{i1}}} \sin(2\pi U_{i2}) & 2\pi\sqrt{U_{i1}} \cos(2\pi U_{i2}) \end{array} \right| \\
= \;& \frac{1}{\pi} \quad \text{on the unit disc.}
\end{aligned} \tag{5.2}$$

We fixed these generated points $\{(x_i, y_i) \mid i = 1, \ldots, 10\}$ to serve as our prominent points in the entire simulation.

Then we took $p = .8$ as the probability for each prominent point to be observed in an image frame, and accordingly selected the observed points that were to be fixed throughout the entire simulation yielding $(m_0, m_1, m_2) = (6, 8, 8)$ for one case which we call Case I, and $(m_0, m_1, m_2) = (5, 7, 8)$ for the other which we call Case II.

Three values of the true scale of Gaussian noise σ were chosen representing the ranges of "*low*" ($\sigma = .02$), "*moderate*" ($\sigma = .05$), and "*high*" ($\sigma = .10$) noise images respectively. See Figure 5.1 for examples of the two simulated image frames for each value of σ in Case I, and Figure 5.2 in Case II.

Also the same three values $\{.02, .05, .10\}$ were chosen for σ_a in the estimation procedure so that there were total of 9 choices of (σ, σ_a), namely $\{.02, .05, .10\}^2$.

From now on throughout the entire chapter we will refer to the situation where we have a particular choice of (σ, σ_a) among the ones given above by Situation(σ, σ_a), or just Situation if there will be no confusion.

In each Case, and for each Situation(σ, σ_a), we ran 1000 trials. Each trial consisted of generating the observed locations in the two image frames by adding the independent error vectors according to the Gaussian noise $\mathcal{N}(0, \sigma^2 \mathbf{I}_2)$ to the true locations, and producing $\hat{\theta}$ and $\hat{T}$ as the estimates of θ and T by the steps described in Subsection 3.3.4 with the following exception. In STEP 9, if $\Theta_{\max} = \emptyset$, then we assigned the values alternately between $\hat{\theta} = 4$, $\hat{T}_x = \hat{T}_y = 3$, and $\hat{\theta} = -4$, $\hat{T}_x = \hat{T}_y = -3$ for the purpose of tabulation. Remember that in the actual estimation, $\hat{\theta} \in (-\pi, \pi]$ always, and seldom will we get $|\hat{T}_x| > 2$ or $|\hat{T}_y| > 2$ in practice because all the true locations are within the unit disc.

For a given Case and trial number, the observed locations for different Situations

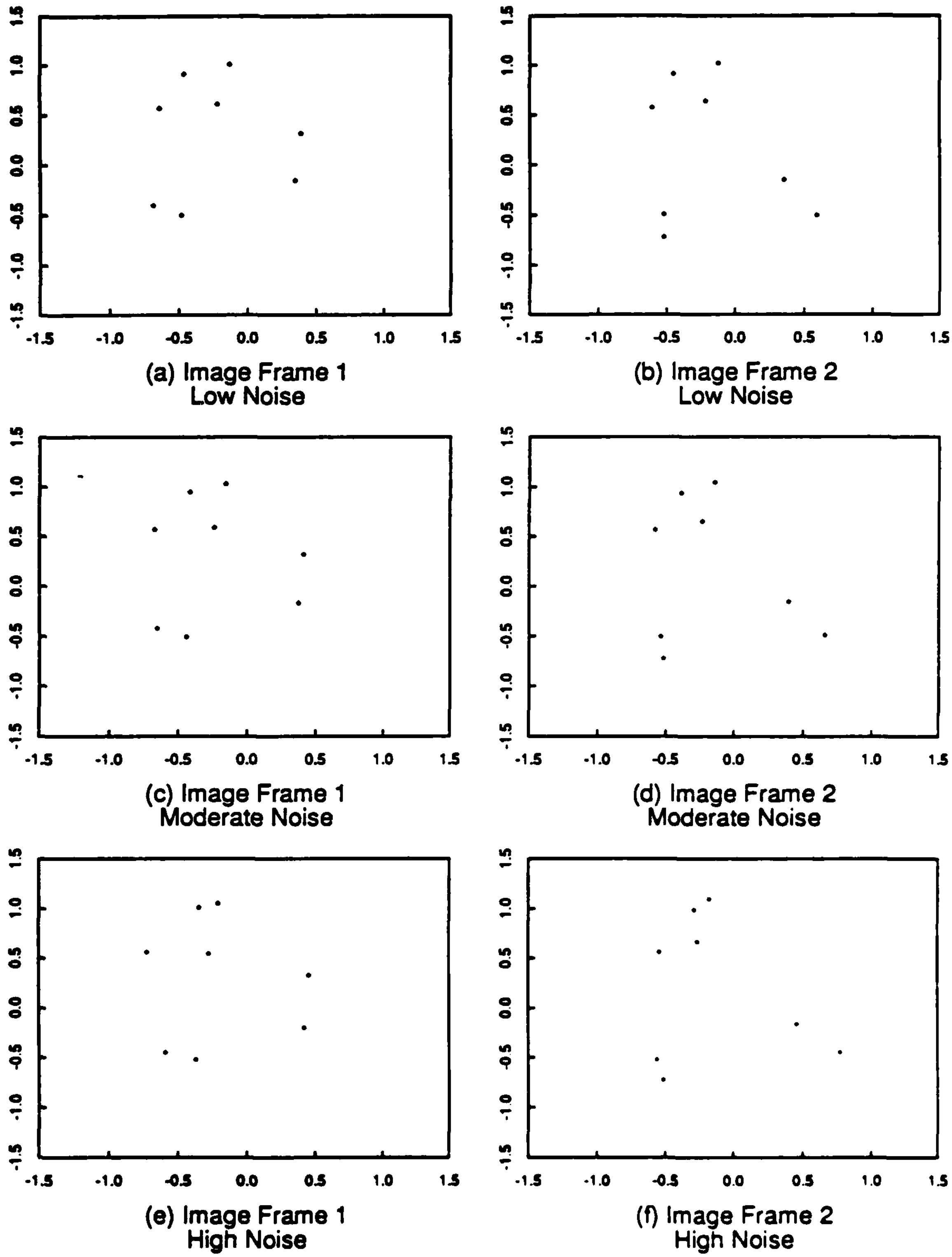
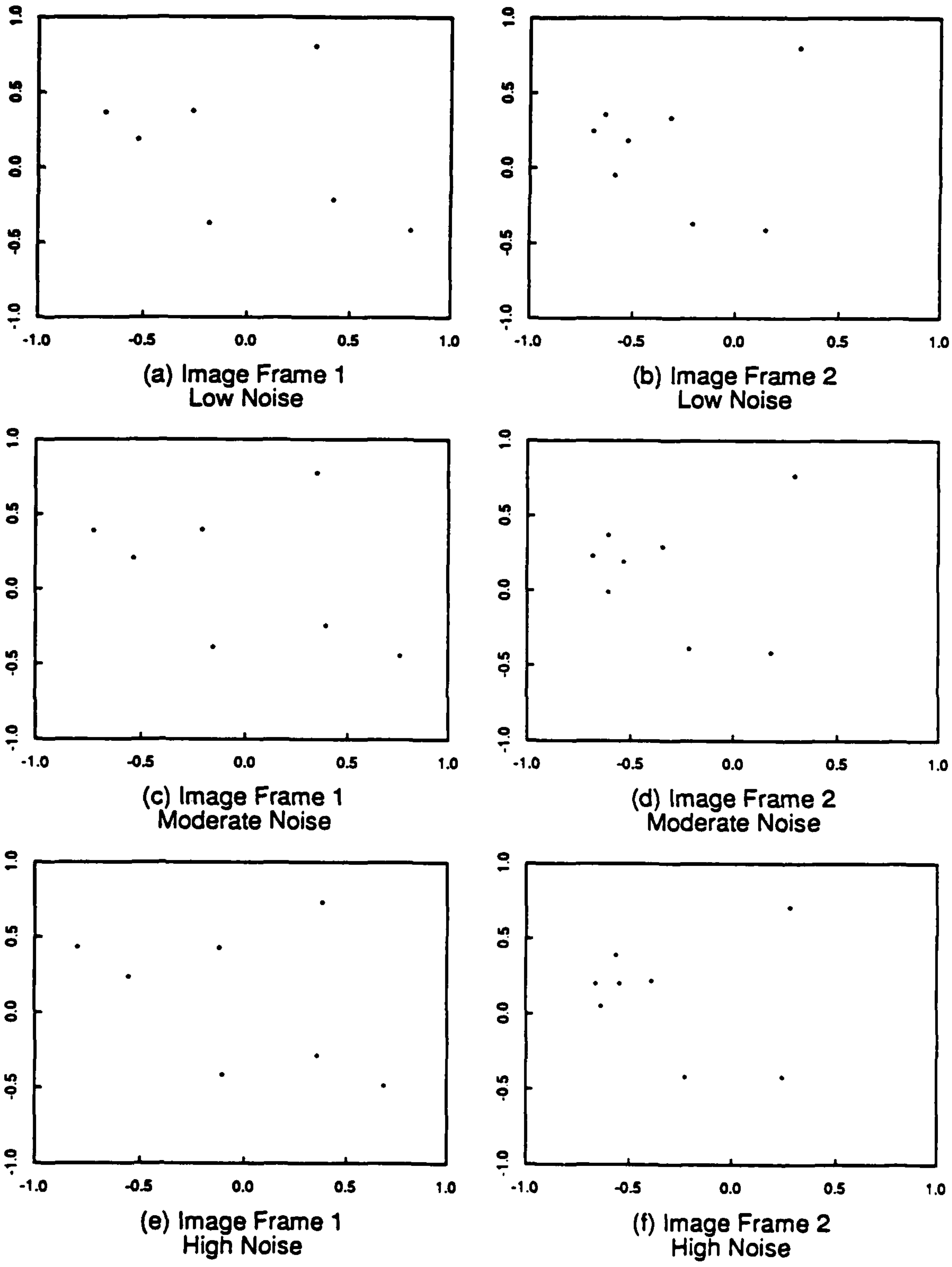


Figure 5.1: Examples of the two simulated image frames for each value of σ, (a)(b) $\sigma = .02$ (low noise), (c)(d) $\sigma = .05$ (moderate noise), (e)(f) $\sigma = .10$ (high noise) in Case I, $(m_0, m_1, m_2) = (6, 8, 8)$.



(a) Image Frame 1
Low Noise

(b) Image Frame 2
Low Noise

(c) Image Frame 1
Moderate Noise

(d) Image Frame 2
Moderate Noise

(e) Image Frame 1
High Noise

(f) Image Frame 2
High Noise

Figure 5.2: Examples of the two simulated image frames for each value of σ, (a)(b) $\sigma = .02$ (low noise), (c)(d) $\sigma = .05$ (moderate noise), (e)(f) $\sigma = .10$ (high noise) in Case II, $(m_0, m_1, m_2) = (5, 7, 8)$.

were related. Thus, in image frame i, $i = 1, 2$,

$$\text{an observed location} = \text{the corresponding true location} + \sigma\varepsilon_i,$$

where ε_i depends on the Case and trial number, but does not change with either σ or σ_a.

5.2 Results of the Simulation

For simplicity we use $\hat{\gamma}$ to represent any one of the estimates $\hat{\theta}$, $\hat{T}_x$, or $\hat{T}_y$, and also use $\hat{\gamma}_{(i)}$ to denote the i-th order statistic of the 1000 $\hat{\gamma}$'s that were produced in each Situation(σ, σ_a).

In each Case, we produced the histograms of the estimates $\hat{\theta}$, $\hat{T}_x$, and $\hat{T}_y$ for each Situation(σ, σ_a). See Figures 5.3 – 5.8.

Then we compared the sample distribution of $\hat{\gamma}$ with the normal distributions by costructing the *normal qq-plot*. With a normal qq-plot one can conveniently get the estimates $\widehat{m}$, $\hat{s}$ of the mean and the standard deviation of the normal distribution $\mathcal{N}(m, s^2)$ by which the sample distribution of $\hat{\gamma}$ is approximated, by fitting a straight line to them and finding the y-intercept and the slope. We fitted the straight line by using Krasker and Welsch's robust regression technique [Krasker 79] to the points on the plot that belong to the [40, 60]-percentiles interval, or equivalently the [−.253, .253]-normal scores interval in order to capture the behaviour of the ones in the middle. After fitting the straight line, we estimated the standard deviation of $\hat{\gamma}_{(i)}$ by the formula [Chambers 83,Kendall 77a],

$$\widehat{sd}(\hat{\gamma}_{(i)}) = \frac{\hat{s}}{\phi(\Phi^{-1}(p_i))}\sqrt{\frac{p_i(1-p_i)}{1000}}, \tag{5.3}$$

where $p_i = (i - .5)/1000$ for $i = 1, \ldots, 1000$ and ϕ, Φ are the probability density function, and the cumulative distribution function of the standard normal distribution, respectively. Then we showed, for each plotted point, $\pm 4\widehat{sd}(\hat{\gamma}_{(i)})$ bounds

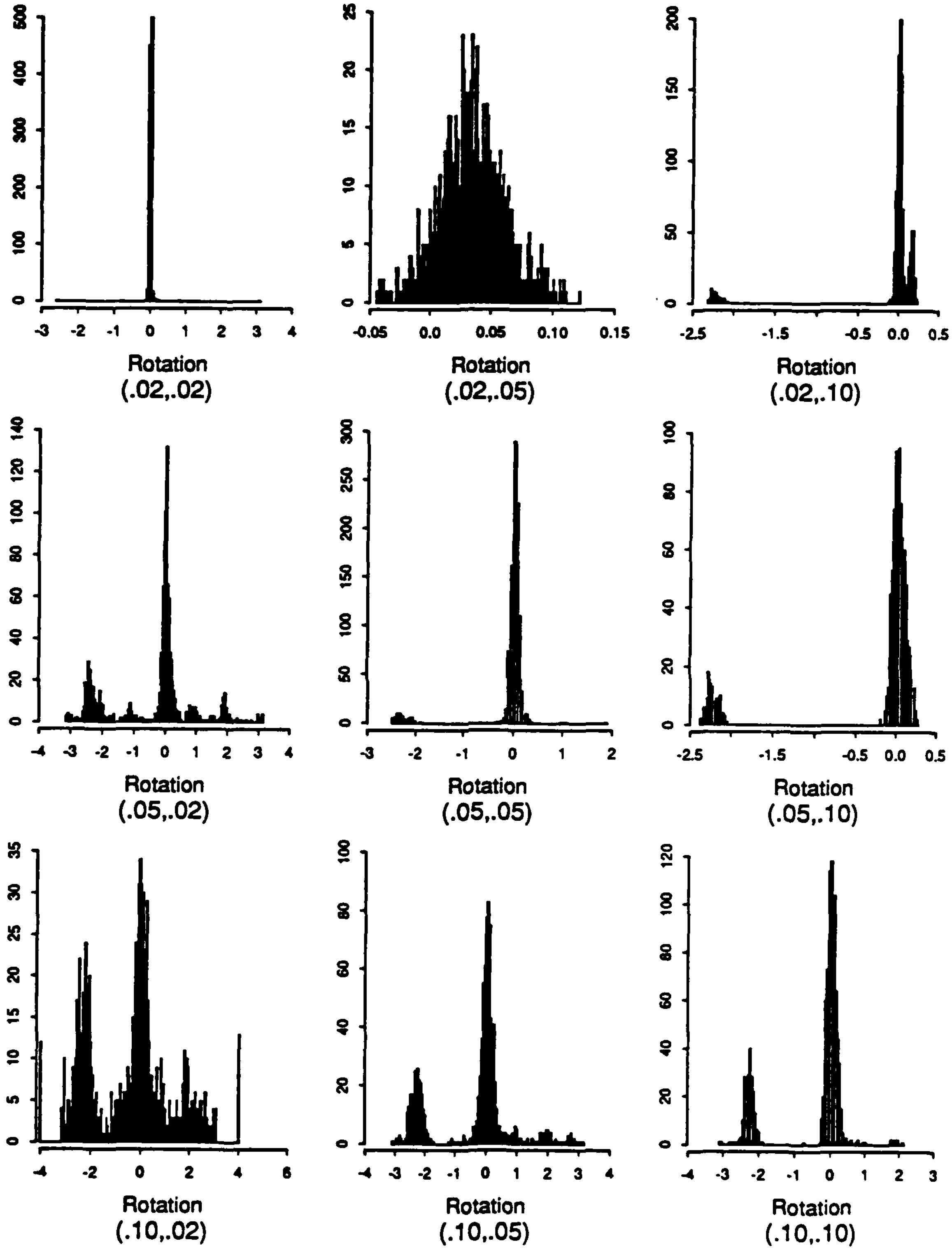


Figure 5.3: Histograms of $\hat{\theta}$ for each Situation in Case I, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

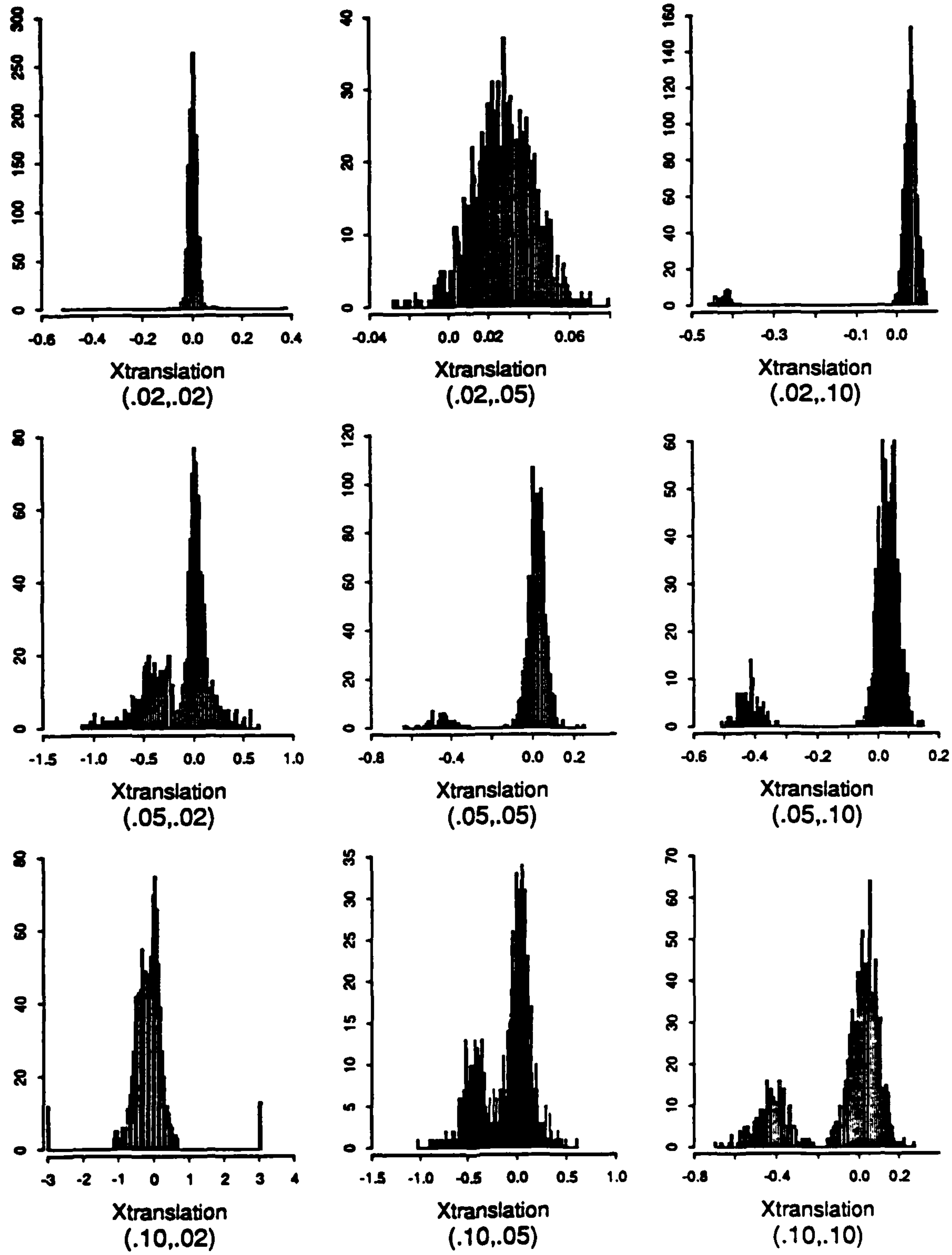


Figure 5.4: Histograms of $\hat{T}_x$ for each Situation in Case I, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

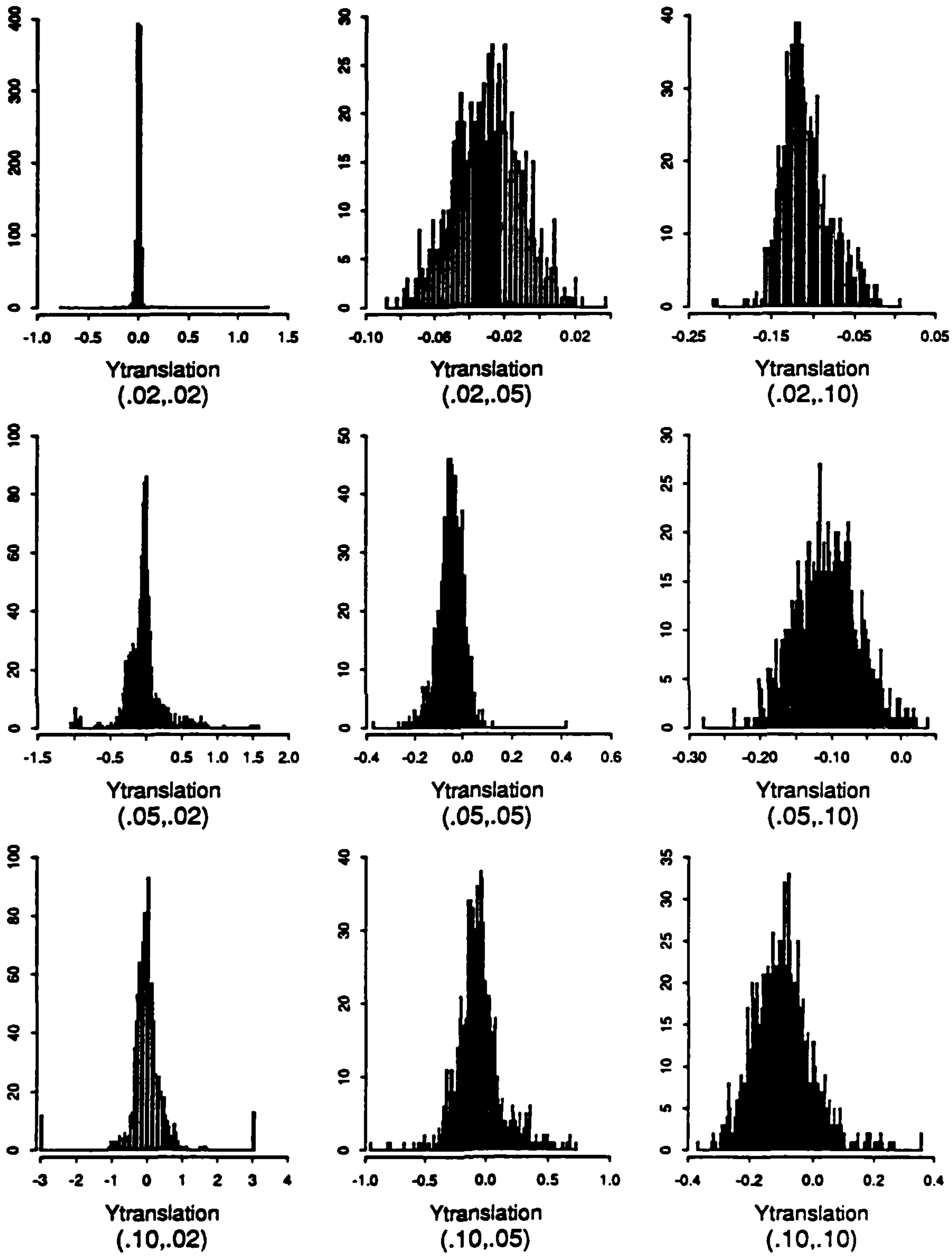


Figure 5.5: Histograms of $\widehat{T}_y$ for each Situation in Case I, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

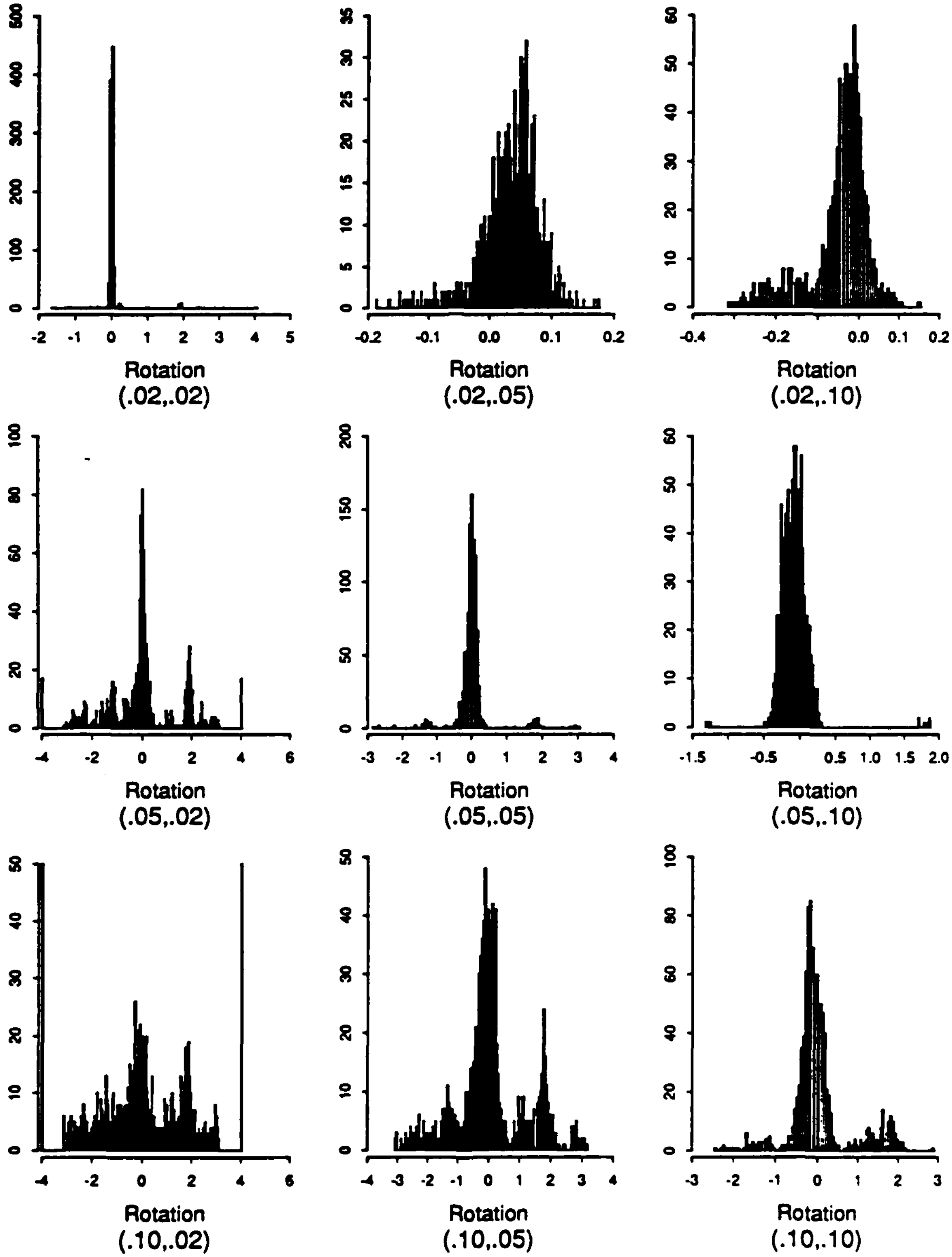


Figure 5.6: Histograms of $\hat{\theta}$ for each Situation in Case II, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

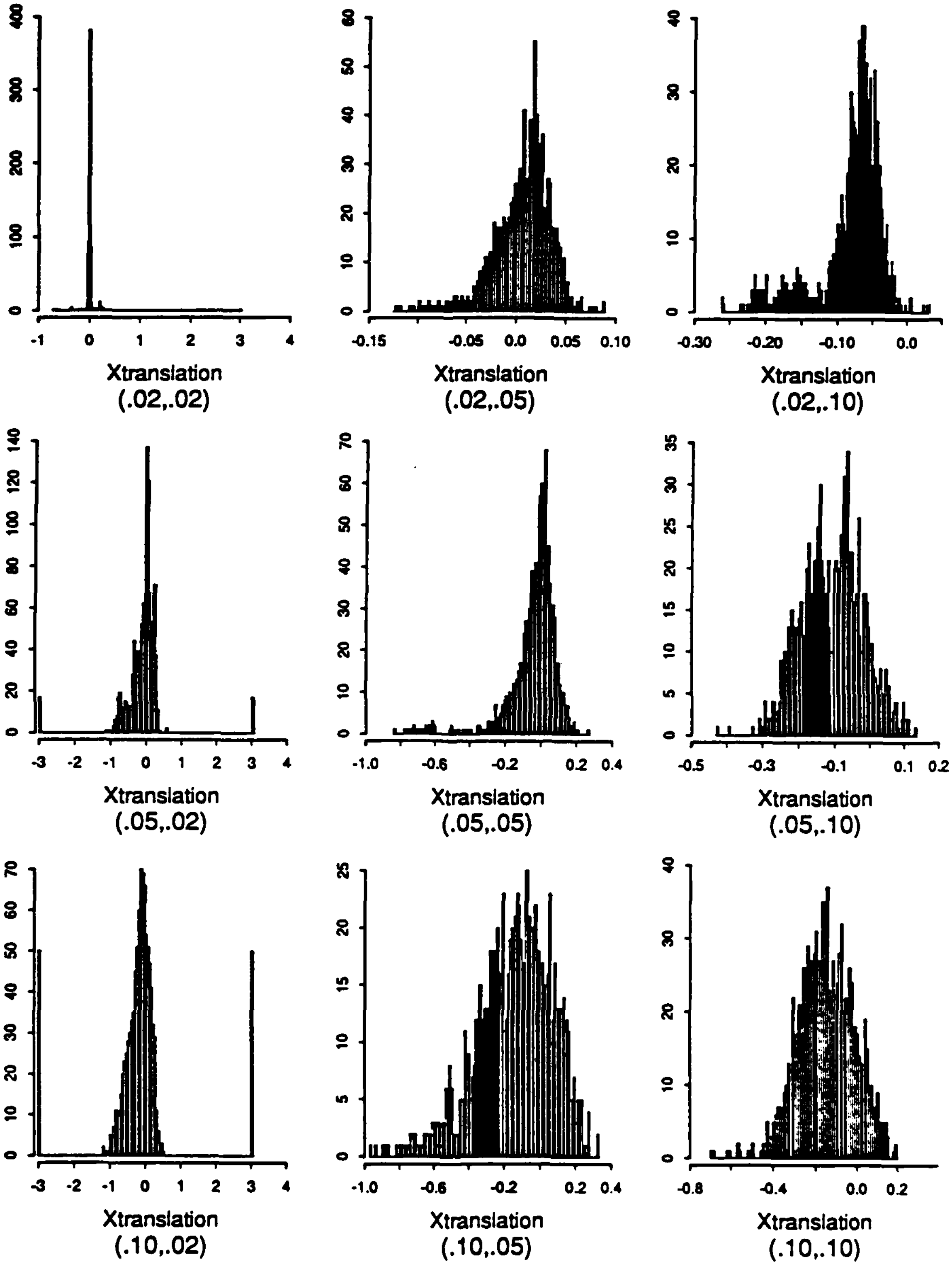


Figure 5.7: Histograms of $\widehat{T}_x$ for each Situation in Case II, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

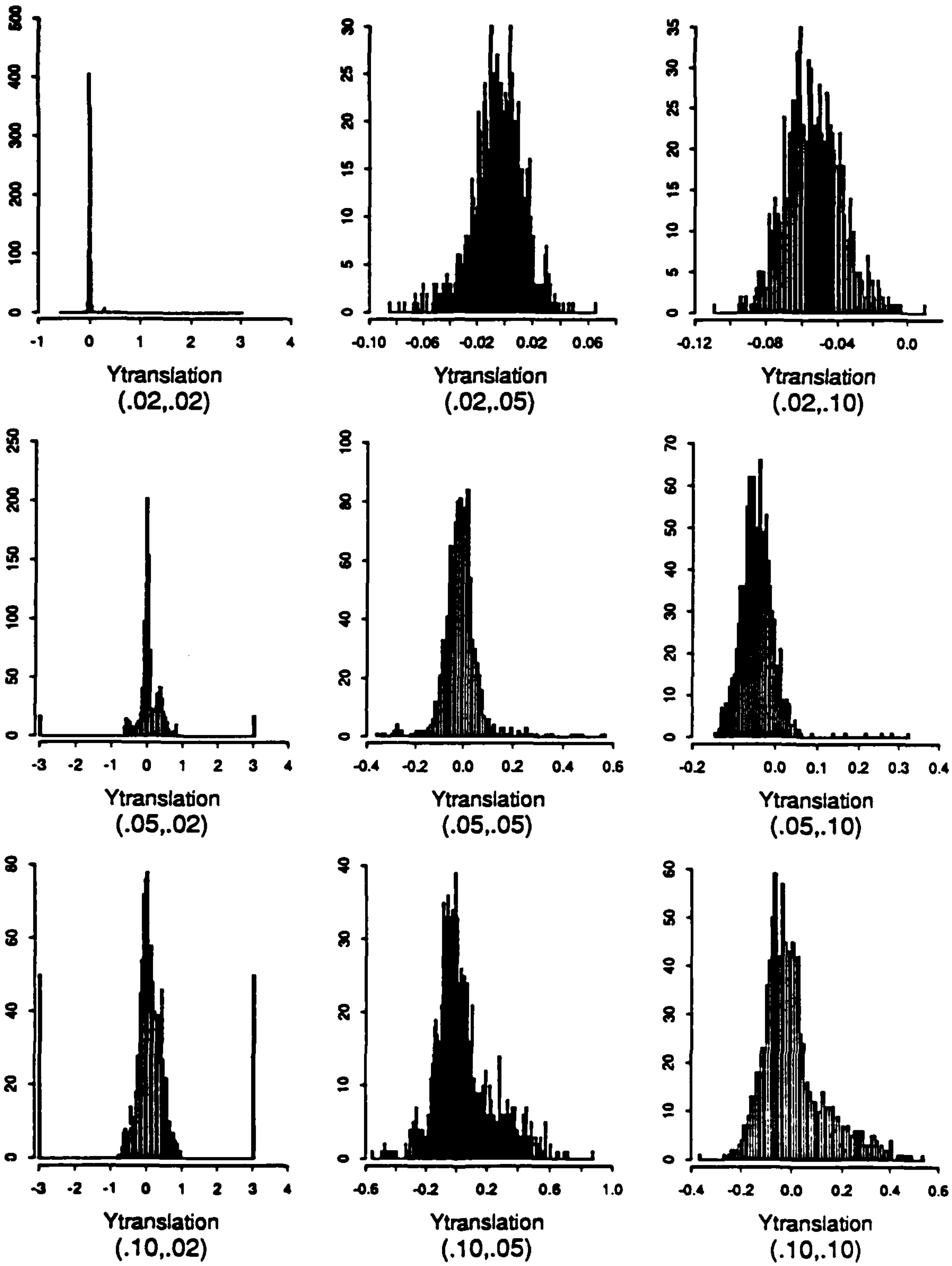


Figure 5.8: Histograms of $\hat{T}_y$ for each Situation in Case II, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

around the line fitted to the plot. This may give us some sense of how well the sample distribution of $\hat{\gamma}$ is approximated by the normal distributon $\mathcal{N}(\widehat{m}, \hat{s}^2)$.

We have presented the normal qq-plots of $\hat{\theta}$, $\widehat{T}_x$, and $\widehat{T}_y$ with the bounds in Figures 5.9 – 5.11, for each Situation in Case I, and in Figures 5.12 – 5.14, for each Situation in Case II.

As expected, every Situation(.02,σ_a) gave us a satisfactory result except that when $\sigma_a = .10$, $\hat{\theta}$ and $\widehat{T}_x$ seem to have long left tailed distributions in both Cases.

For Situations(.05,σ_a), all but when $\sigma_a = .10$ in Case II, $\hat{\theta}$ and $\widehat{T}_x$ show long left tailed distributions while $\widehat{T}_y$ consistently shows satisfactory results.

For Situations(.10,σ_a) every thing seems to fall apart. When $\sigma_a = .10$ our approach seems to confuse nearby points. When $\sigma_a = .05$ or .02, the standard for matching difference vectors is difficult to meet, and our method frequently fails to find satisfactory matches.

We have summarized the information obtained from the normal qq-plots in Tables 5.1 – 5.6. They show $\widehat{m}$, $\hat{s}$ of $\hat{\gamma}$ and corresponding $\hat{\sigma}_{\mathrm{ML}}$, the ratio $\hat{s}/\hat{\sigma}_{\mathrm{ML}}$ along with the number of points outside the bounds of ± 4 standard deviations from the fitted line for each situation.

In Table 5.7, we have collected all the ratios $\hat{s}/\hat{\sigma}_{\mathrm{ML}}$ in Case I, and in Table 5.8, the same thing in Case II. From them you can easily see that the ratios $\hat{s}/\hat{\sigma}_{\mathrm{ML}}$ are the smallest when $\sigma = \sigma_a$ and the largest occur when $\sigma = .10$ and $\sigma_a = .02$, namely when we assumed too low a value for the standard deviation.

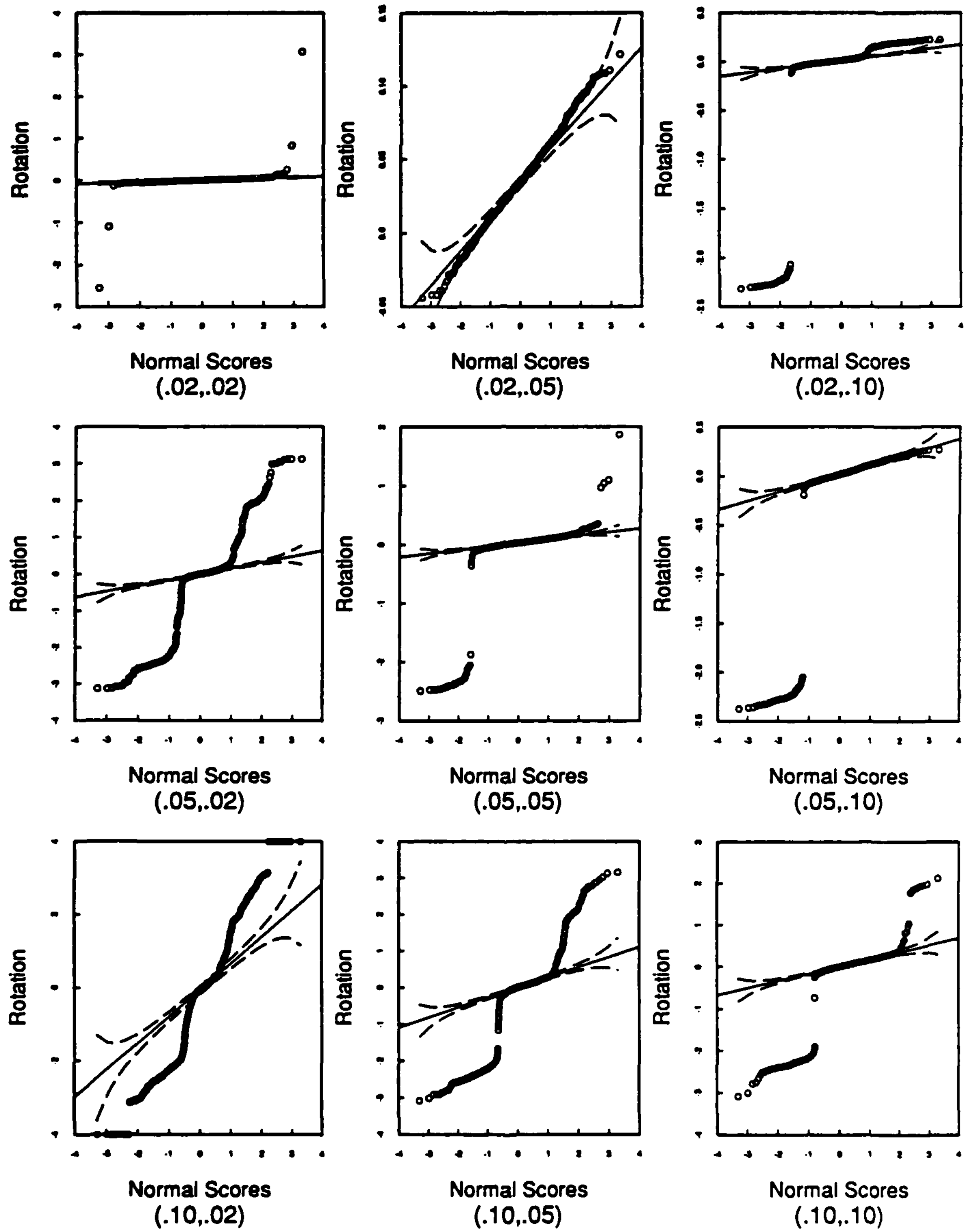


Figure 5.9: Normal qq-plots of $\hat{\theta}$ for each Situation in Case I, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

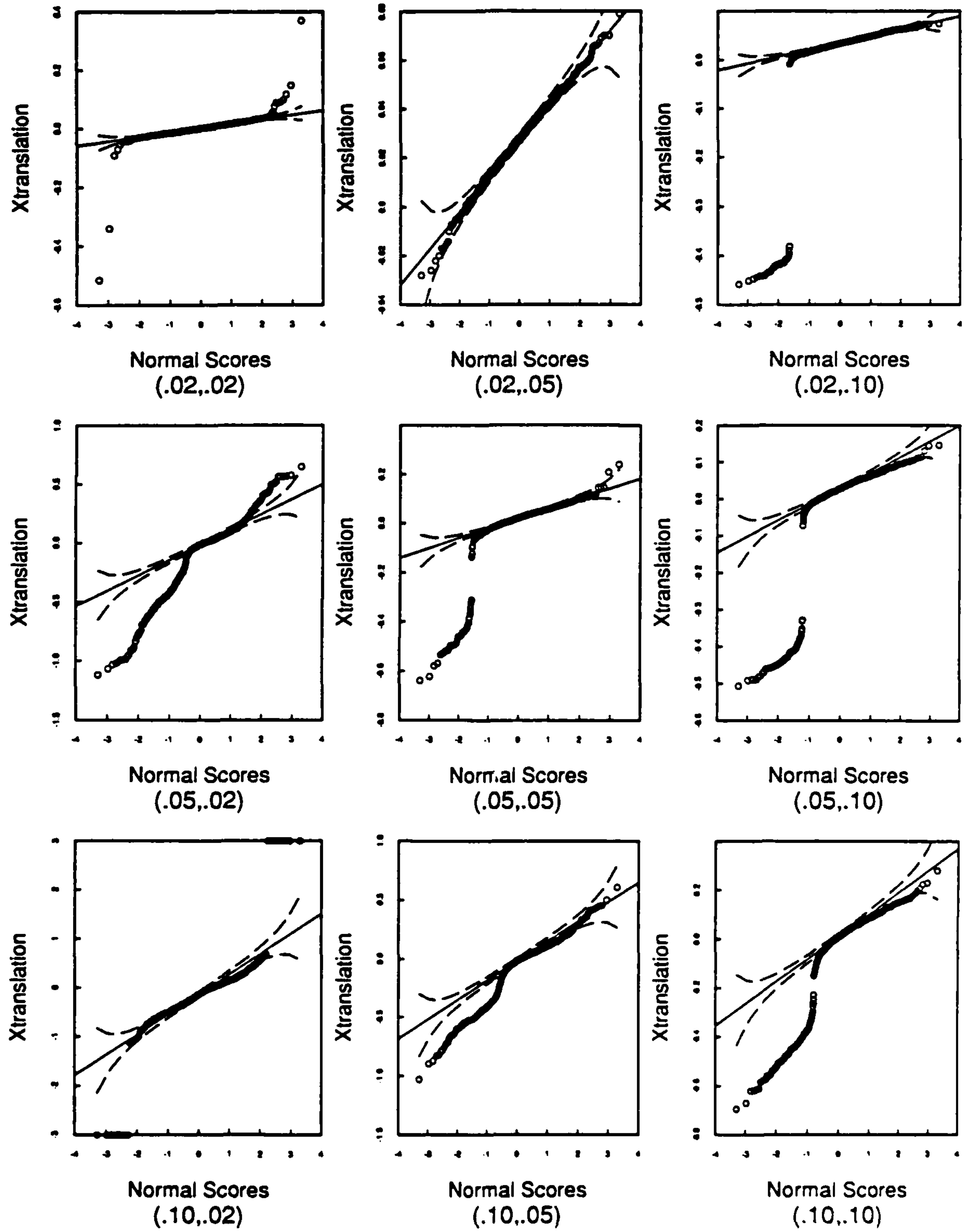


Figure 5.10: Normal qq-plots of $\hat{T}_x$ for each Situation in Case I, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

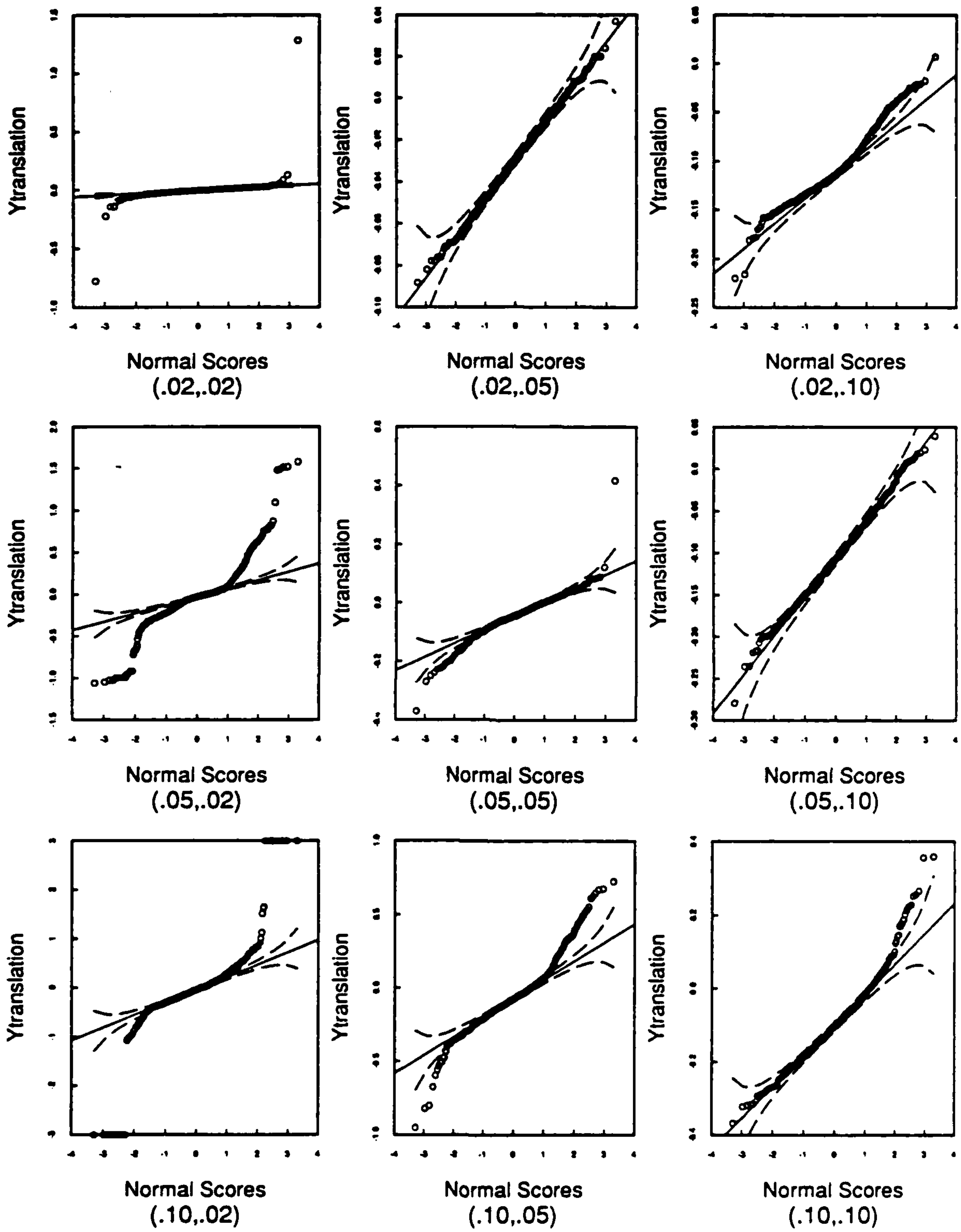


Figure 5.11: Normal qq-plots of $\widehat{T}_y$ for each Situation in Case I, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

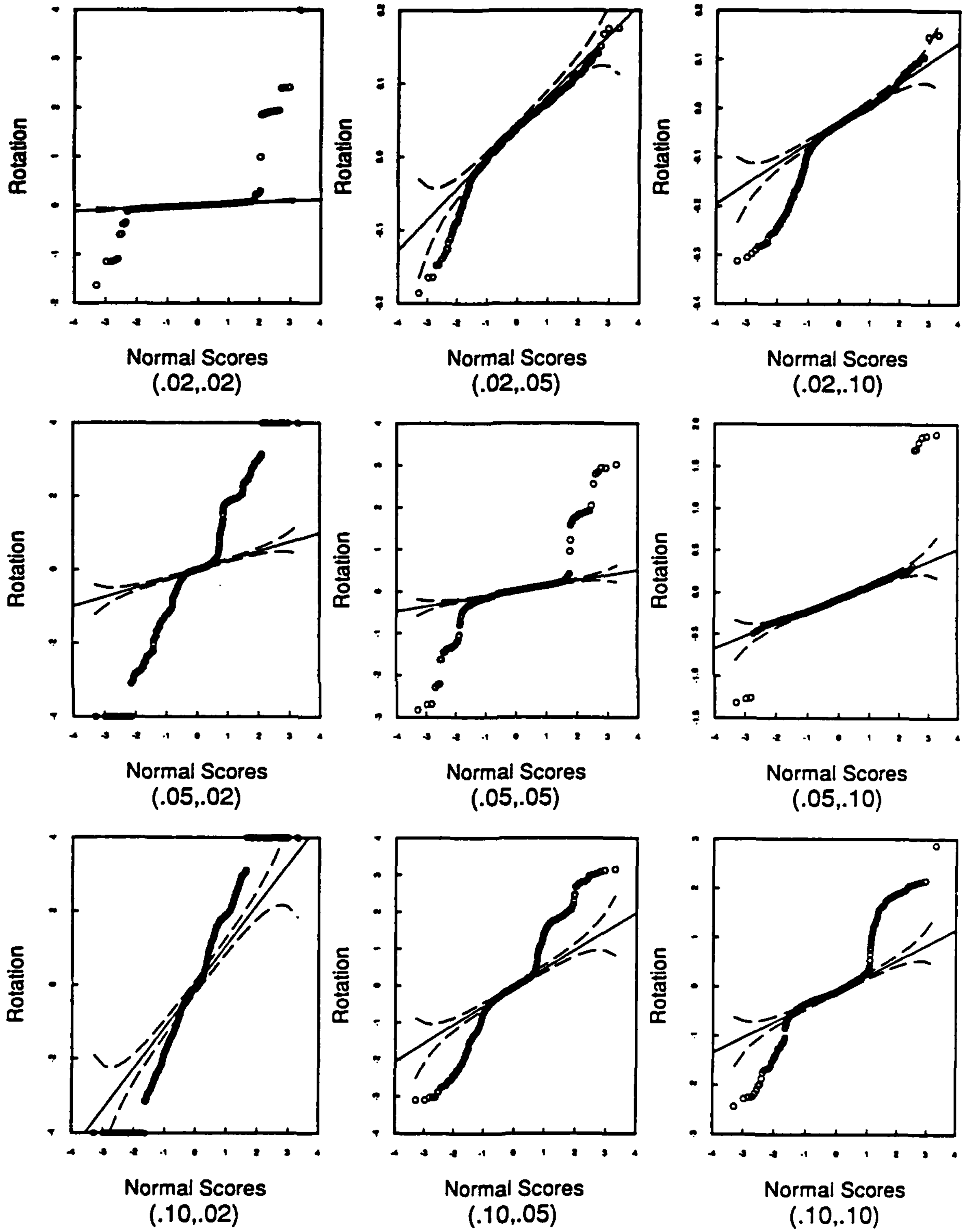


Figure 5.12: Normal qq-plots of $\hat{\theta}$ for each Situation in Case II, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

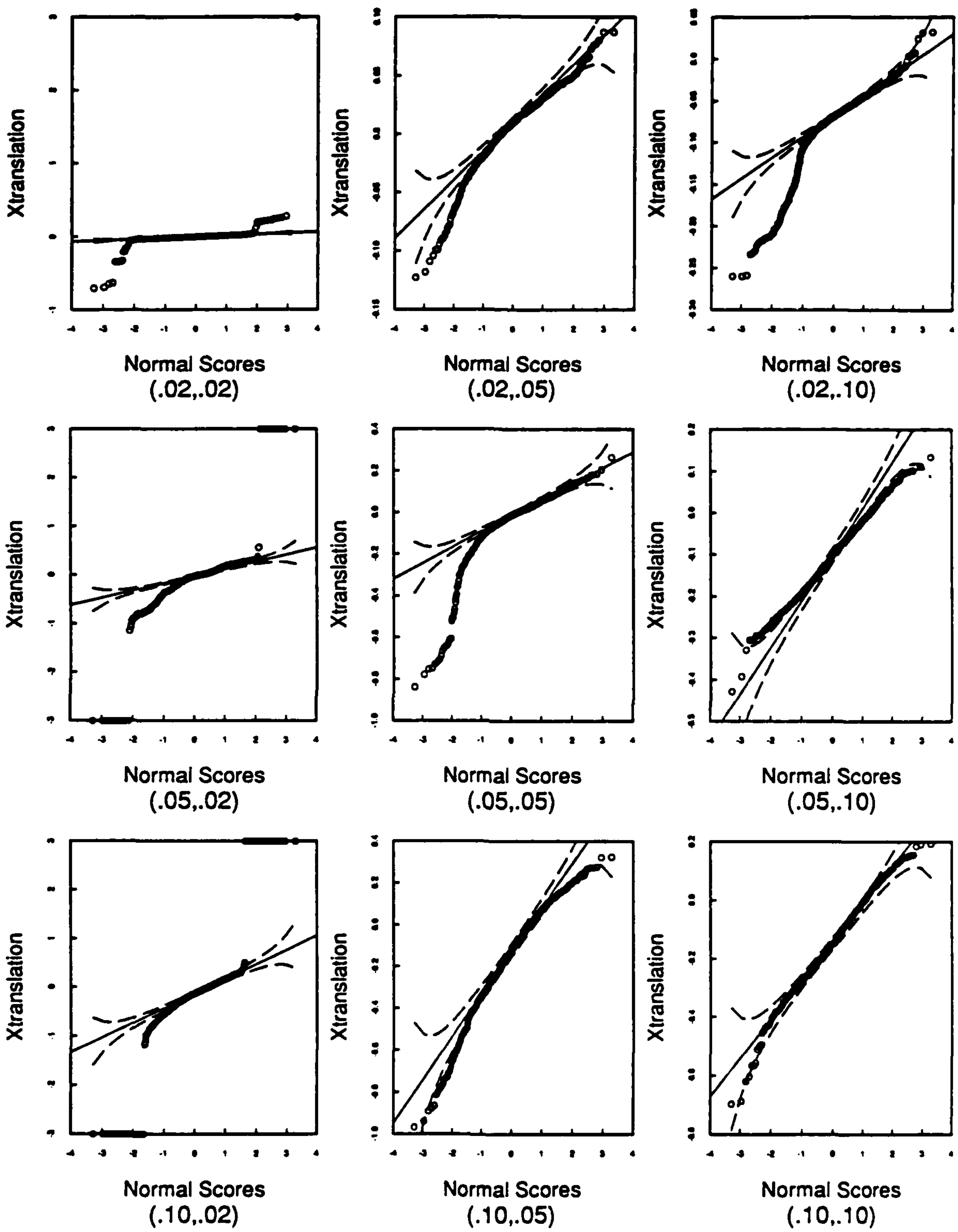


Figure 5.13: Normal qq-plots of $\widehat{T}_x$ for each Situation in Case II, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

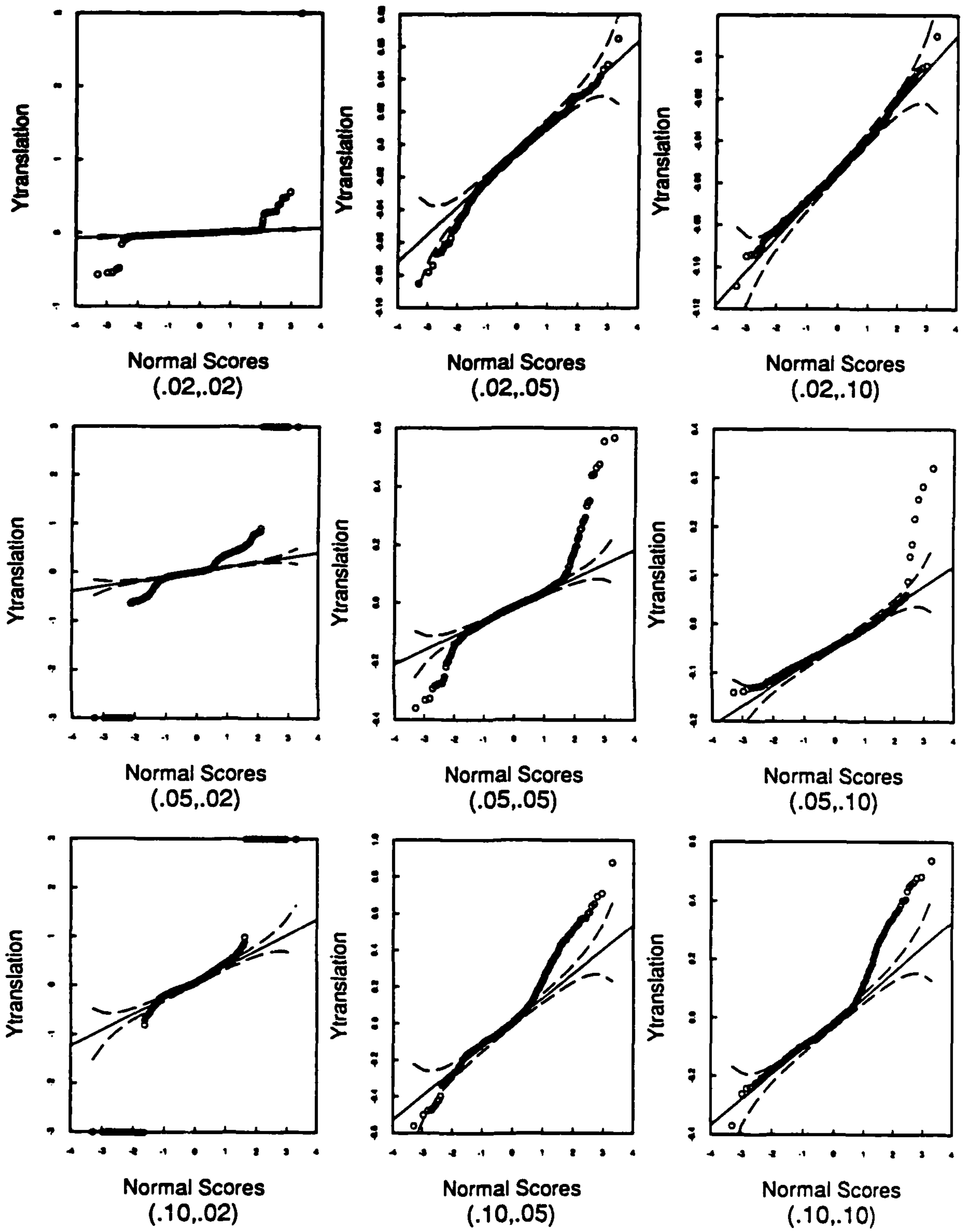


Figure 5.14: Normal qq-plots of $\hat{T}_y$ for each Situation in Case II, 1000 trials each. The two numbers in parentheses represent σ and σ_a respectively.

σ	Summary Statistics	σ_a		
		.02	.05	.10
.02	$\widehat{m}$	.0007	.0340	.0146
	$\hat{s}$	.0224	.0232	.0411
	$\hat{\sigma}_{\mathrm{ML}}$	.0182	.0182	.0182
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	1.2308	1.2747	2.2582
	# Outside ± 4 $\widehat{sd}$ Bounds	32	79	320
.05	$\widehat{m}$	−.0016	.0331	.0211
	$\hat{s}$	.1567	.0612	.0895
	$\hat{\sigma}_{\mathrm{ML}}$	.0456	.0456	.0456
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	3.4364	1.3421	1.9247
	# Outside ± 4 $\widehat{sd}$ Bounds	605	312	135
.10	$\widehat{m}$	−.0861	.0063	.0027
	$\hat{s}$	.7240	.2741	.1694
	$\hat{\sigma}_{\mathrm{ML}}$	.0912	.0912	.0912
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	7.9386	3.0055	1.8575
	# Outside ± 4 $\widehat{sd}$ Bounds	(25) 650	453	295

Table 5.1: Summary from the normal qq-plots of $\hat{\theta}$ in Case I, 1000 trials each. The numbers in the parentheses represent the numbers of trials with "No Matches" if there are any.

σ	Summary Statistics	σ_a .02	.05	.10
.02	$\widehat{m}$	.0016	.0276	.0346
	$\hat{s}$	.0152	.0149	.0138
	$\hat{\sigma}_{\mathrm{ML}}$	.0135	.0135	.0135
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	1.1259	1.1037	1.0222
	# Outside $\pm 4\ \widehat{sd}$ Bounds	23	0	181
.05	$\widehat{m}$	−.0161	.0210	.0279
	$\hat{s}$	.1294	.0404	.0431
	$\hat{\sigma}_{\mathrm{ML}}$	.0337	.0337	.0337
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	3.8398	1.1988	1.2789
	# Outside $\pm 4\ \widehat{sd}$ Bounds	423	144	312
.10	$\widehat{m}$	−.1367	−.0198	.0078
	$\hat{s}$	.4112	.1663	.0906
	$\hat{\sigma}_{\mathrm{ML}}$	.0673	.0673	.0673
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	6.1100	2.4710	1.3462
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(25) 255	523	420

Table 5.2: Summary from the normal *qq*-plots of $\widehat{T}_x$ in Case I, 1000 trials each. The numbers in the parentheses represent the numbers of trials with "No Matches" if there are any.

σ	Summary Statistics	σ_a		
		.02	.05	.10
.02	$\widehat{m}$	−.0012	−.0293	−.1132
	$\hat{s}$	.0143	.0188	.0255
	$\hat{\sigma}_{\mathrm{ML}}$	.0126	.0126	.0126
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	1.1349	1.4921	2.0238
	# Outside $\pm 4\ \widehat{sd}$ Bounds	137	0	365
.05	$\widehat{m}$	−.0256	−.0429	−.1064
	$\hat{s}$	.0997	.0467	.0463
	$\hat{\sigma}_{\mathrm{ML}}$	.0314	.0314	.0314
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	3.1752	1.4873	1.4745
	# Outside $\pm 4\ \widehat{sd}$ Bounds	501	103	0
.10	$\widehat{m}$	−.0407	−.0719	−.1025
	$\hat{s}$	.2560	.1276	.0838
	$\hat{\sigma}_{\mathrm{ML}}$	.0628	.0628	.0628
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	4.0764	2.0318	1.3344
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(25) 239	138	38

Table 5.3: Summary from the normal *qq*-plots of $\hat{T}_y$ in Case I, 1000 trials each. The numbers in the parentheses represent the numbers of trials with "No Matches" if there are any.

σ	Summary Statistics	σ_a .02	.05	.10
.02	$\widehat{m}$	.0040	.0409	−.0300
	$\hat{s}$	.0304	.0421	.0414
	$\hat{\sigma}_{\mathrm{ML}}$	.0250	.0250	.0250
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	1.2160	1.6840	1.6560
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(1) 163	209	220
.05	$\widehat{m}$	.0070	.0189	−.0794
	$\hat{s}$	.2494	.1222	.1496
	$\hat{\sigma}_{\mathrm{ML}}$	.0625	.0625	.0625
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	3.9904	1.9552	2.3936
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(34) 684	301	9
.10	$\widehat{m}$	−.0381	−.0350	−.0976
	$\hat{s}$	1.1089	.5037	.3121
	$\hat{\sigma}_{\mathrm{ML}}$	.1250	.1250	.1250
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	8.8712	4.0296	2.4968
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(100) 668	442	273

Table 5.4: Summary from the normal qq-plots of $\hat{\theta}$ in Case II, 1000 trials each. The numbers in the parentheses represent the numbers of trials with "No Matches" if there are any.

σ	Summary Statistics	σ_a .02	.05	.10
	$\widehat{m}$	−.0004	.0105	−.0686
	$\hat{s}$	.0172	.0248	.0247
.02	$\hat{\sigma}_{\mathrm{ML}}$	.0141	.0141	.0141
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	1.2199	1.7589	1.7518
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(1) 71	250	254
	$\widehat{m}$	−.0270	−.0138	−.1011
	$\hat{s}$	.1476	.0763	.1108
.05	$\hat{\sigma}_{\mathrm{ML}}$	.0352	.0352	.0352
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	4.1932	2.1676	3.1477
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(34) 610	239	393
	$\widehat{m}$	−.1290	−.1216	−.1472
	$\hat{s}$	.2998	.2067	.1309
.10	$\hat{\sigma}_{\mathrm{ML}}$	.0705	.0705	.0705
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	4.2525	2.9319	1.8567
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(100) 312	119	4

Table 5.5: Summary from the normal qq-plots of $\widehat{T}_x$ in Case II, 1000 trials each. The numbers in the parentheses represent the numbers of trials with "No Matches" if there are any.

σ	Summary Statistics	σ_a		
		.02	.05	.10
.02	$\widehat{m}$	−.0011	−.0039	−.0540
	$\hat{s}$	.0179	.0170	.0161
	$\hat{\sigma}_{\mathrm{ML}}$	.0144	.0144	.0144
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	1.2431	1.1806	1.1181
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(1) 61	50	0
.05	$\widehat{m}$	.0073	−.0141	−.0454
	$\hat{s}$	.0999	.0492	.0410
	$\hat{\sigma}_{\mathrm{ML}}$	.0359	.0359	.0359
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	2.7827	1.3705	1.1421
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(34) 502	105	105
.10	$\widehat{m}$	.0650	.0053	−.0204
	$\hat{s}$	.3237	.1334	.0869
	$\hat{\sigma}_{\mathrm{ML}}$	.0718	.0718	.0718
	$\hat{s}/\hat{\sigma}_{\mathrm{ML}}$	4.5084	1.8579	1.2103
	# Outside $\pm 4\ \widehat{sd}$ Bounds	(100) 283	271	227

Table 5.6: Summary from the normal qq-plots of $\widehat{T}_y$ in Case II, 1000 trials each. The numbers in the parentheses represent the numbers of trials with "No Matches" if there are any.

$\hat{\gamma}$	σ	σ_a .02	.05	.10
$\hat{\theta}$	.02	1.2308	1.2747	2.2582
	.05	3.4364	1.3421	1.9247
	.10	7.9386	3.0055	1.8575
$\hat{T}_x$	.02	1.1259	1.1037	1.0222
	.05	3.8398	1.1988	1.2789
	.10	6.1100	2.4710	1.3462
$\hat{T}_y$	.02	1.1349	1.4921	2.0238
	.05	3.1752	1.4873	1.4745
	.10	4.0764	2.0318	1.3344

Table 5.7: The ratios $\hat{s}/\hat{\sigma}_{\mathrm{ML}}$ in Case I, 1000 trials each.

$\hat{\gamma}$	σ	σ_a .02	.05	.10
$\hat{\theta}$	.02	1.2160	1.6840	1.6560
	.05	3.9904	1.9552	2.3936
	.10	8.8712	4.0296	2.4968
$\hat{T}_x$	.02	1.2199	1.7589	1.7518
	.05	4.1932	2.1676	3.1477
	.10	4.2525	2.9319	1.8567
$\hat{T}_y$	.02	1.2431	1.1806	1.1181
	.05	2.7827	1.3705	1.1421
	.10	4.5084	1.8579	1.2103

Table 5.8: The ratios $\hat{s}/\hat{\sigma}_{\mathrm{ML}}$ in Case II, 1000 trials each.

Chapter 6

Conclusions

The problem of estimating a rigid body motion from two noisy images of an object taken at two different times has been studied.

The available data consist of the unordered locations of some of the prominent points of the object. Because it is assumed here that these points are not individually recognized and some of them may be missed in one or both of the images, it is not obvious which points of the two images correspond to one another. Moreover the observed locations are subject to error.

A computational procedure, capitalizing on the rigidity of the object, has been proposed for estimating the motion parameters of the object in the presence of $\mathcal{N}(0, \sigma^2 I_2)$ Gaussian noise independently added to the positions in the images of the points observed.

In principle, one might apply maximum likelihood to the estimation problem, but the difficulty in formulating and calculating the likelihood function under the above mentioned assumptions is formidable. However one ought to expect to do better when the observed points are recognized and the common points among them are matched without error in identification.

For this favorable situation, where the likelihood function can be easily formulated, the asymptotic normality and consistency of the maximum likelihood estimate as $\sigma \to 0$, have been proved. Also the asymptotic covariance matrix for that has been derived explicitly. Similar results hold even when σ is incorrectly assumed to be $\sigma_a = k\sigma$ for positive $k \neq 1$.

Simulation results have shown that in most cases where we either know the correct σ (i.e., $\sigma_a = \sigma$), or assumed a larger value for that ($\sigma_a > \sigma$), the proposed estimator does reasonably well compared with the maximum likelihood estimate, considering the fact that the latter is under a favorable situation.

Computationally, there is still much to be desired and it will be an interesting problem to produce an algorithm which reduces the cost of computing for this method.

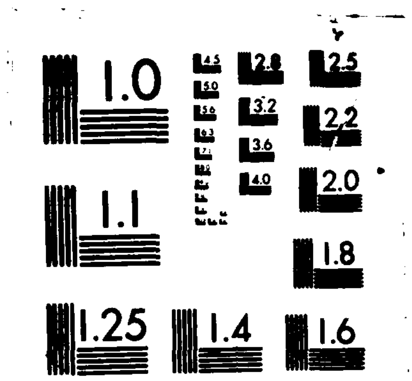
One important generalization of this method will be to allow a noise structure where the nearby observed locations are correlated, which is often the case in real world data.

Appendix A

AD-A183 034 THE ESTIMATION OF A RIGID BODY MOTION IN THE PRESENCE OF NOISE(U) HARVARD UNIV CAMBRIDGE MASS DEPT OF STATISTICS C H PARK 31 JUL 87 TR-ONR-C-1 2/2

UNCLASSIFIED N00014-86-K-0246 F/G 14/4 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

A.1 The Almost Sure Uniqueness of $\hat{\beta}_{\mathrm{ML}}(\mathbf{X})$

In the following discussion, we will continue to use the same notation as in Chapter 4, and assume $m_0 \geq 2$.

From the definition of $\hat{\beta}_{\mathrm{ML}}(\mathbf{X})$ and Equation (4.16), it is clear that $\hat{\beta}_{\mathrm{ML}}(\mathbf{X})$ can be defined to be the value of $\tilde{\beta} \in \Omega$ which minimizes $\left\|\mathbf{X} - G(\tilde{\beta})\right\|^2$, or equivalently $\left\|\mathbf{X} - G(\tilde{\beta})\right\|$.

First, let us define the subset Ω_0 of Ω by

$$\Omega_0 := \left\{ \tilde{\beta} = \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\theta} \\ \tilde{T} \end{bmatrix} \in \Omega \,\middle|\, \tilde{\mu}_1 = \tilde{\mu}_{11} \otimes \mathbf{1}_{m_0}, \tilde{\mu}_{11} \in \mathbf{R}^2, \tilde{\theta} \in (-\pi, \pi], \tilde{T} \in \mathbf{R}^2 \right\}, \quad \text{(A.1)}$$

that is, Ω_0 is the subset of Ω consisting of $\tilde{\beta}$'s which correspond to m_0 identical points. Let $G(\Omega)$ be the *range* of the function G:

$$G(\Omega) := \{G(\tilde{\beta}) \mid \tilde{\beta} \in \Omega\} \subset \mathbf{R}^{4m_0}. \quad \text{(A.2)}$$

Similarly let

$$G(\Omega_0) := \{G(\tilde{\beta}) \mid \tilde{\beta} \in \Omega_0\} \subset G(\Omega) \subset \mathbf{R}^{4m_0}. \quad \text{(A.3)}$$

Then we can rephrase Lemma 4.2 as the following lemma.

Lemma A.1 *The function G restricted to $\Omega \setminus \Omega_0$ is one-to-one onto $G(\Omega) \setminus G(\Omega_0)$.*

We need the following lemma to proceed.

Lemma A.2 *$G(\Omega)$ is a closed subset of $\mathbf{R}^{4m_0}$.*

Proof : Let $\mathbf{y}^{(i)} := G(\beta^{(i)})$ with $\beta^{(i)} \in \Omega$, $i = 1, 2, \ldots$ be a sequece in $G(\Omega)$ that converges to a point $\mathbf{y}^{(0)} \in \mathbf{R}^{4m_0}$. For each $i = 1, 2, \ldots$, let

$$\beta^{(i)} := \begin{bmatrix} \mu^{(i)} \\ \theta^{(i)} \\ T^{(i)} \end{bmatrix}, \ \mu^{(i)} \in \mathbf{R}^{2m_0}, \ \theta^{(i)} \in (-\pi, \pi], \ T^{(i)} \in \mathbf{R}^2,$$

so that

$$\mathbf{y}^{(i)} = \begin{bmatrix} \boldsymbol{\mu}^{(i)} \\ (\mathbf{U}(\theta^{(i)}) \otimes \mathbf{I}_{m_0})\boldsymbol{\mu}^{(i)} + \boldsymbol{T}^{(i)} \otimes \mathbf{1}_{m_0} \end{bmatrix}.$$

Then clearly the $\boldsymbol{\mu}^{(i)}$ converges to some $\boldsymbol{\mu}^{(0)} \in \mathbf{R}^{2m_0}$.

Since $[-\pi, \pi]$ is compact and $(-\pi, \pi] \subset [-\pi, \pi]$, by taking a subsequence, we can assume that the $\theta^{(i)}$ converge in $[-\pi, \pi]$. Then the $\mathbf{U}(\theta^{(i)})$ converge to some $\mathbf{U}(\theta^{(0)})$, $\theta^{(0)} \in (-\pi, \pi]$, and it follows that the $\boldsymbol{T}^{(i)}$ converge to some $\boldsymbol{T}^{(0)} \in \mathbf{R}^2$. Therefore

$$\mathbf{y}^{(0)} = \lim_{n \to \infty} \mathbf{y}^{(i)} = G(\boldsymbol{\beta}^{(0)}),$$

where

$$\boldsymbol{\beta}^{(0)} := \begin{bmatrix} \boldsymbol{\mu}^{(0)} \\ \theta^{(0)} \\ \boldsymbol{T}^{(0)} \end{bmatrix} \in \Omega.$$

Hence $\mathbf{y}^{(0)} \in G(\Omega)$. So $G(\Omega)$ is closed. This completes the proof.

Let λ be the Lebesgue measure on $\mathbf{R}^{4m_0}$. Then the following arguments will show that $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}(\mathbf{X})$ is unique for λ-almost all $\mathbf{X} \in \mathbf{R}^{4m_0}$, and hence almost surely, because the Gaussian probability measure is absolutely continuous with respect to λ.

For any $\mathbf{x} \in \mathbf{R}^m$, and any set $F \subset \mathbf{R}^m$, let

$$\delta_F(\mathbf{x}) := \inf\{\|\mathbf{x} - \mathbf{y}\| \mid \mathbf{y} \in F\}. \tag{A.4}$$

Note that if $\mathbf{x} \in F$, then $\delta_F(\mathbf{x}) = 0$. We have the following propositions.

Proposition A.1 *For any non-empty closed set $F \subset \mathbf{R}^m$ and any $\mathbf{x} \in \mathbf{R}^m$, $\mathbf{x} \in F$ if and only if $\delta_F(\mathbf{x}) = 0$.*

Proof : We only need to prove the "if" part. Suppose $\delta_F(\mathbf{x}) = 0$. Then, for each $n = 1, 2, \ldots$, there exists a $\mathbf{y}_n \in F$ such that

$$\|\mathbf{x} - \mathbf{y}_n\| < \delta_F(\mathbf{x}) + \frac{1}{n} = \frac{1}{n}.$$

Hence the $\mathbf{y}_n$ converge to $\mathbf{x}$, and since F is closed, $\mathbf{x} \in F$.

Proposition A.2 *For any non-empty closed set $F \subset \mathbf{R}^m$, and any $\mathbf{x} \in \mathbf{R}^m$, there is a $\mathbf{y} \in F$ with $\|\mathbf{x} - \mathbf{y}\| = \delta_F(\mathbf{x})$.*

Proof : Without loss of generality, we can restrict $\mathbf{y}$ to the set

$$\mathcal{B}_{\mathbf{x}} := \{\mathbf{y} \in F \mid \|\mathbf{x} - \mathbf{y}\| \leq 2\delta_F(\mathbf{x})\},$$

which is compact, and on which the function $\mathbf{y} \longmapsto \|\mathbf{x} - \mathbf{y}\|$ is continuous. Thus it attains its minimum on $\mathcal{B}_{\mathbf{x}}$ and therefore on F.

Proposition A.3 *For any non-empty set $F \subset \mathbf{R}^m$, δ_F is a Lipschitz function.*

Proof : For any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{R}^m$, and any $\varepsilon > 0$, we can find a $\mathbf{y}_\varepsilon \in F$ such that

$$\|\mathbf{x}_2 - \mathbf{y}_\varepsilon\| < \delta_F(\mathbf{x}_2) + \varepsilon.$$

Then

$$\begin{aligned} \delta_F(\mathbf{x}_1) - \delta_F(\mathbf{x}_2) &\leq \|\mathbf{x}_1 - \mathbf{y}_\varepsilon\| - \|\mathbf{x}_2 - \mathbf{y}_\varepsilon\| + \varepsilon \\ &\leq \|\mathbf{x}_1 - \mathbf{x}_2\| + \varepsilon. \end{aligned}$$

Hence, by letting $\varepsilon \to 0$ and then by symmetry, we have

$$|\delta_F(\mathbf{x}_1) - \delta_F(\mathbf{x}_2)| \leq \|\mathbf{x}_1 - \mathbf{x}_2\|. \tag{A.5}$$

This completes the proof.

A function $t : \mathbf{R}^m \longrightarrow \mathbf{R}^1$ is called *totally* (*Fréchet*) *differentiable* at $\mathbf{x}_0 \in \mathbf{R}^m$ if the (row) gradient vector $\partial t(\mathbf{x})/\partial \mathbf{x}|_{\mathbf{x}=\mathbf{x}_0}$ exists and $t(\mathbf{x})$ admits the first order Taylor Expansion at $\mathbf{x} = \mathbf{x}_0$:

$$t(\mathbf{x}_0 + \boldsymbol{\varepsilon}) = t(\mathbf{x}_0) + \left.\frac{\partial t(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0} \boldsymbol{\varepsilon} + o(\|\boldsymbol{\varepsilon}\|) \quad \text{as } \boldsymbol{\varepsilon} \to 0. \tag{A.6}$$

Let λ_m be the Lebesgue measure on $\mathbf{R}^m$. We have the following theorem due to H. Rademacher ([Rademacher 19]).

Theorem A.1 (Rademacher's Theorem) *Any real-valued Lipschitz function on an open set O in $\mathbf{R}^m$ is totally differentiable at λ_m-almost all $\mathbf{x} \in O$.*

Proof : See, for example, [Federer 69, pp. 216–217].

Federer [Federer 59, 4.8. Theorem] has proved the following theorems.

Theorem A.2 *For any non-empty closed set $F \subset \mathbf{R}^m$, and $\mathbf{y} \in F$, the set*

$$Q(\mathbf{y}) := \{\mathbf{z} \in \mathbf{R}^m \mid \delta_F(\mathbf{y} + \mathbf{z}) = \|\mathbf{z}\|\} \tag{A.7}$$

is convex.

Proof : By translating the set F by $-\mathbf{y}$, we can assume that $\mathbf{y} = \mathbf{0}$. Then

$$\mathbf{z} \in Q(\mathbf{y}) \quad \text{if and only if} \quad \|\mathbf{b} - \mathbf{z}\| \geq \|\mathbf{z}\| \quad \text{for all} \quad \mathbf{b} \in F.$$

Furthermore

$$\|\mathbf{b} - \mathbf{z}\|^2 - \|\mathbf{z}\|^2 = \mathbf{b}^T(\mathbf{b} - 2\mathbf{z}) \quad \text{for any} \quad \mathbf{b}, \mathbf{z} \in \mathbf{R}^m,$$

and consequently, if $\mathbf{b}$, $\mathbf{z}$, $\mathbf{w} \in \mathbf{R}^m$, $\alpha, \gamma \geq 0$, $\alpha + \gamma = 1$, then

$$\begin{aligned}
\|\mathbf{b} - (\alpha\mathbf{z} + \gamma\mathbf{w})\|^2 - \|\alpha\mathbf{z} + \gamma\mathbf{w}\|^2 &= \mathbf{b}^T(\mathbf{b} - 2\alpha\mathbf{z} - 2\gamma\mathbf{w}) \\
&= \mathbf{b}^T[\alpha(\mathbf{b} - 2\mathbf{z}) + \gamma(\mathbf{b} - 2\mathbf{w})] \\
&= \alpha\mathbf{b}^T(\mathbf{b} - 2\mathbf{z}) + \gamma\mathbf{b}^T(\mathbf{b} - 2\mathbf{w}) \\
&= \alpha(\|\mathbf{b} - \mathbf{z}\|^2 - \|\mathbf{z}\|^2) + \gamma(\|\mathbf{b} - \mathbf{w}\|^2 - \|\mathbf{w}\|^2).
\end{aligned}$$

It follows that $Q(\mathbf{y})$ is convex.

Theorem A.3 *For any non-empty closed set $F \subset \mathbf{R}^m$, if $\mathbf{x} \in \mathbf{R}^m \setminus F$ and δ_F is totally differentiable at $\mathbf{x}$, then there exists a unique $\mathbf{y} \in F$ such that $\|\mathbf{x} - \mathbf{y}\| = \delta_F(\mathbf{x})$.*

Proof : By Propositions A.1 and A.2, there is a $\mathbf{y} \in F$ such that $\|\mathbf{x}-\mathbf{y}\| = \delta_F(\mathbf{x}) > 0$. Then $\mathbf{x}-\mathbf{y}, 0 \in Q(\mathbf{y})$ by definition of $Q(\mathbf{y})$, and hence by Theorem A.2 $(1-\alpha)(\mathbf{x}-\mathbf{y}) \in Q(\mathbf{y})$ for $0 \leq \alpha \leq 1$. Let

$$\boldsymbol{\varepsilon} := -\alpha(\mathbf{x}-\mathbf{y}). \tag{A.8}$$

Then we have

$$\begin{aligned}
\delta_F(\mathbf{x}+\boldsymbol{\varepsilon}) &= \delta_F[\mathbf{y}+(1-\alpha)(\mathbf{x}-\mathbf{y})] \\
&= \|(1-\alpha)(\mathbf{x}-\mathbf{y})\| \\
&= \delta_F(\mathbf{x}) - \alpha\delta_F(\mathbf{x}) \quad \text{for } 0 \leq \alpha \leq 1.
\end{aligned}$$

Thus from the total differentiability of δ_F at $\mathbf{x}$, as $\alpha \downarrow 0$, we must have

$$\begin{aligned}
-\frac{\partial\delta_F(\mathbf{x})}{\partial\mathbf{x}}\frac{\mathbf{x}-\mathbf{y}}{\delta_F(\mathbf{x})} &= \frac{\partial\delta_F(\mathbf{x})}{\partial\mathbf{x}}\frac{\boldsymbol{\varepsilon}}{\|\boldsymbol{\varepsilon}\|} \\
&= \frac{\delta_F(\mathbf{x}+\boldsymbol{\varepsilon})-\delta_F(\mathbf{x})}{\|\boldsymbol{\varepsilon}\|} + \frac{o(\|\boldsymbol{\varepsilon}\|)}{\|\boldsymbol{\varepsilon}\|} \\
&= -1 + \frac{o(\|\boldsymbol{\varepsilon}\|)}{\|\boldsymbol{\varepsilon}\|},
\end{aligned}$$

and hence by taking the limit as $\alpha \downarrow 0$, we get

$$\frac{\partial\delta_F(\mathbf{x})}{\partial\mathbf{x}}\frac{\mathbf{x}-\mathbf{y}}{\delta_F(\mathbf{x})} = 1. \tag{A.9}$$

Since $\|\partial\delta_F(\mathbf{x})/\partial\mathbf{x}\| \leq 1$ from (A.5), it follows that

$$\frac{\partial\delta_F(\mathbf{x})}{\partial\mathbf{x}} = \frac{\mathbf{x}-\mathbf{y}}{\delta_F(\mathbf{x})}. \tag{A.10}$$

Therefore such $\mathbf{y}$ must be unique.

Applying Theorems A.1 and A.3 with $O = \mathbf{R}^{4m_0} \setminus G(\Omega)$, $F = G(\Omega)$, along with the fact that for any $\mathbf{x} \in F$, $\mathbf{x}$ itself is the unique $\mathbf{y} \in F$ satisfying $\|\mathbf{x}-\mathbf{y}\| = \delta_F(\mathbf{x})$, we have proved the following theorem.

Theorem A.4 *There exists a set $N \subset \mathbf{R}^{4m_0}$ with $\lambda(N) = 0$ such that for any $\mathbf{X} \in \mathbf{R}^{4m_0} \setminus N$, there exists a unique $\mathbf{y}(\mathbf{X}) \in G(\Omega)$ that satisfies*

$$\|\mathbf{X} - \mathbf{y}(\mathbf{X})\| = \delta_{G(\Omega)}(\mathbf{X}). \tag{A.11}$$

A similar theorem has been proved by A. Pázman ([Pázman 84]).

Let N be a set in $\mathbf{R}^{4m_0}$ chosen to satisfy the properties of Theorem A.4. If $\mathbf{X} \in \mathbf{R}^{4m_0} \setminus N$, $\mathbf{y}(\mathbf{X}) \in G(\Omega)$ of Theorem A.4 is unique. Suppose that $\mathbf{y}(\mathbf{X}) \notin G(\Omega_0)$. Then by Lemma A.1, $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}(\mathbf{X})$ is unique. Hence, to show that $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}(\mathbf{X})$ is almost surely unique, it suffices to show that $\lambda(M) = 0$ where

$$M := \{\mathbf{X} \in \mathbf{R}^{4m_0} \setminus N \mid \mathbf{y}(\mathbf{X}) \in G(\Omega_0)\}. \tag{A.12}$$

Theorem A.5 $\lambda(M) = 0$.

Proof : Let

$$\mathbf{X} := \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \in M, \ \mathbf{X}_1, \mathbf{X}_2 \in \mathbf{R}^{2m_0}.$$

Then $\mathbf{y}(\mathbf{X}) \in G(\Omega_0)$.

For any $\tilde{\boldsymbol{\beta}} := \begin{bmatrix} \tilde{\boldsymbol{\mu}}_1 \\ \tilde{\theta} \\ \tilde{\boldsymbol{T}} \end{bmatrix} \in \Omega, \quad \tilde{\boldsymbol{\mu}}_1 \in \mathbf{R}^{2m_0}, \quad \tilde{\theta} \in (-\pi, \pi], \quad \tilde{\boldsymbol{T}} \in \mathbf{R}^2,$

$$\begin{aligned} \left\|\mathbf{X} - G(\tilde{\boldsymbol{\beta}})\right\|^2 &= \|\mathbf{X}_1 - \tilde{\boldsymbol{\mu}}_1\|^2 + \left\|\mathbf{X}_2 - (\mathbf{U}(\tilde{\theta}) \otimes \mathbf{I}_{m_0})\tilde{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{T}} \otimes \mathbf{1}_{m_0}\right\|^2 \\ &= \|\mathbf{X}_1 - \tilde{\boldsymbol{\mu}}_1\|^2 + \left\|(\mathbf{U}(\tilde{\theta})^T \otimes \mathbf{I}_{m_0})\mathbf{X}_2 - \mathbf{U}(\tilde{\theta})^T \tilde{\boldsymbol{T}} \otimes \mathbf{1}_{m_0} - \tilde{\boldsymbol{\mu}}_1\right\|^2, \end{aligned} \tag{A.13}$$

since left multiplication by $\mathbf{U}(\tilde{\theta})^T$ preserves the norm, and so does the one by $\mathbf{U}(\tilde{\theta})^T \otimes \mathbf{I}_{m_0}$.

Now let us choose a $\boldsymbol{\beta}_0 := \begin{bmatrix} \boldsymbol{\mu}_0 \\ \theta_0 \\ \boldsymbol{T}_0 \end{bmatrix} \in \Omega_0$, $\boldsymbol{\mu}_0 \in \mathbf{R}^{2m_0}$, $\theta_0 \in (-\pi, \pi]$, $\boldsymbol{T}_0 \in \mathbf{R}^2$ with $\mathbf{y}(\mathbf{X}) = G(\boldsymbol{\beta}_0)$. Then

$$\|\mathbf{X} - G(\boldsymbol{\beta}_0)\|^2 = \min_{\tilde{\boldsymbol{\beta}} \in \Omega} \left\|\mathbf{X} - G(\tilde{\boldsymbol{\beta}})\right\|^2$$

$$
\begin{aligned}
&= \min_{\tilde{\boldsymbol{\mu}}_1 \in \mathbf{R}^{2m_0}} \left\| \mathbf{X} - G\left(\begin{bmatrix} \tilde{\boldsymbol{\mu}}_1 \\ \boldsymbol{\theta}_0 \\ \boldsymbol{T}_0 \end{bmatrix} \right) \right\|^2 \\
&= \min_{\tilde{\boldsymbol{\mu}}_1 \in \mathbf{R}^{2m_0}} \left[\|\mathbf{X}_1 - \tilde{\boldsymbol{\mu}}_1\|^2 + \|\mathbf{X}_1' - \tilde{\boldsymbol{\mu}}_1\|^2 \right] \qquad \text{(A.14)}
\end{aligned}
$$

where

$$\mathbf{X}_1' := (\mathbf{U}(\theta_0)^T \otimes \mathbf{I}_{m_0})\mathbf{X}_2 - \mathbf{U}(\theta_0)^T T_0 \otimes \mathbf{1}_{m_0}. \qquad \text{(A.15)}$$

Now the last minimization problem in (A.14) has a unique solution for $\tilde{\boldsymbol{\mu}}_1$,

$$\tilde{\boldsymbol{\mu}}_1 = \frac{\mathbf{X}_1 + \mathbf{X}_1'}{2}, \qquad \text{(A.16)}$$

which, by Lemma A.1, must be of the form $\tilde{\boldsymbol{\mu}}_{11} \otimes \mathbf{1}_{m_0}$ for some $\tilde{\boldsymbol{\mu}}_{11} \in \mathbf{R}^2$ because $\mathbf{y}(\mathbf{X}) \in G(\Omega_0)$. Then it follows that

$$\mathbf{X}_1 + (\mathbf{U}(\theta_0)^T \otimes \mathbf{I}_{m_0})\mathbf{X}_2 = \mathbf{z} \otimes \mathbf{1}_{m_0} \text{ for some } \mathbf{z} \in \mathbf{R}^2,$$

or

$$\mathbf{X}_2 = -(\mathbf{U}(\theta_0) \otimes \mathbf{I}_{m_0})\mathbf{X}_1 + \mathbf{w} \otimes \mathbf{1}_{m_0} \text{ for some } \mathbf{w} \in \mathbf{R}^2.$$

Thus we have

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ -(\mathbf{U}(\theta_0) \otimes \mathbf{I}_{m_0})\mathbf{X}_1 + \mathbf{w} \otimes \mathbf{1}_{m_0} \end{bmatrix}. \qquad \text{(A.17)}$$

Therefore

$$M \subset \left\{ \mathbf{X} \middle| \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ -(\mathbf{U}(\theta_0) \otimes \mathbf{I}_{m_0})\mathbf{X}_1 + \mathbf{w} \otimes \mathbf{1}_{m_0} \end{bmatrix}, \mathbf{X}_1 \in \mathbf{R}^{2m_0}, \theta_0 \in (-\pi, \pi], \mathbf{w} \in \mathbf{R}^2 \right\}$$

and hence M is contained in a smooth manifold of dimension $\leq 2m_0 + 3 < 4m_0$. It then follows that

$$\lambda(M) = 0.$$

In summary, we have

Theorem A.6 *The maximum likelihood estimate $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}(\mathbf{X})$ is almost surely unique.*

Bibliography

[Baird 84] Baird, H. S. (1984). Model-Based Image Matching Using Location. Ph.D. Dissertation, Priceton University, Princeton, New Jersey.

[Bishop 75] Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, Massachusetts.

[Bolles 79] Bolles, R. C. (1979). Robust Feature Matching through Maximal Cliques. In *Proceedings of the Society of Photo-Optical and Instrumentation Engineers*, Vol. 182, pp. 140–149.

[Box 58] Box, G. E. P. and Muller, M. E. (1958). A Note on the Generation of Random Normal Deviates. *Annals of Mathematical Statistics*, Vol. 29, No. 2, pp. 610–611.

[Chambers 83] Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983). *Graphical Methods for Data Analysis.* Wadsworth International Group, Belmont, California, pp. 227–233.

[Chernoff 79] Chernoff, H. (1979). *Sequential Analysis and Optimal Design.* Regional Conference Series in Applied Mathematics No. 8, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

[Choquet 69] Choquet, G. (1969). *Geometry in a Modern Setting.* Hermann, Paris; Houghton Mifflin, Boston, Massachusetts.

[Fang 83a] Fang, J, Q. and Huang, T. S. (1983). Solving Three-Dimensional Small Rotation Motion Equations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, pp. 253–258.

[Fang 83b] Fang, J, Q. and Huang, T. S. (1983). Estimating 3-D Movement of a Rigid Object: Experimental Results. In *Proceedings of the International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, pp. 1035–1037.

[Federer 59] Federer, H. (1959). *Geometric Measure Theory.* Springer-Verlag, New York.

[Federer 69] Federer, H. (1969). Curvature Measures. *Transactions of the American Mathematical Society*, Vol. 93, pp. 418–491.

[Foutz 77] Foutz, R. V. (1977). On the Unique Consistent Solution to the Likelihood Equations. *Journal of the American Statistical Association*, Vol. **72**, No. 357, pp. 147–148.

[Hu 69] Hu, S. T. (1969). *Introduction to General Topology*. Holden-Day, San Francisco, California.

[Huang 81] Huang, T. S. (1981). *Image Sequence Analysis*. Springer-Verlag, New York.

[Huzurbazar 49] Huzurbazar, V. S. (1949). On a Property of Distributions Admitting Sufficient Statistics. *Biometrika*, Vol. **36**, pp. 71–74.

[Johansen 84] Johansen, S. (1984). *Functional Relations, Random Coefficients, and Nonlinear Regression with Application to Kinetic Data*. Lecture Notes in Statistics, Vol. **22**. Springer-Verlag, New York.

[Johnson 70] Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics*, Vol. **3**. Continuous Univariate Distributions-2. Houghton Mifflin, Boston, Massachusetts.

[Kahl 80] Kahl, D. J., Rosenfeld, A. and Danker, A. (1980). Some Experiments in Point Pattern Matching. *IEEE Transactions SMC*, Vol. **SMC-10**, 1980, pp. 105–116.

[Kendall 77a] Kendall, M. G. and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. **1**. Distribution Theory, 4th ed. MacMillan, New York.

[Kendall 77b] Kendall, M. G. and Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. **2**. Inference and Relationship, 4th ed. MacMillan, New York.

[Krasker 79] Krasker, W. S. and Welsch, R. E. (1982). Efficient Bounded-Influence Regression Estimation. *Journal of the American Statistical Association*, Vol. **77**, No. 379, pp. 595– 604.

[Lavine 81] Lavine, D., Lambird, B. A. and Kanal, L. N. (1981). Recognition of Spatial Point Patterns. In *Proceedings of the IEEE Conference on Pattern Recognition and Image Processing*, Dallas, Texas, August 1981, pp. 49–53.

[Nagel 81a] Nagel, H. H. (1981). On the Derivation of 3-D Rigid Point Configurations from Image Sequences. In *Proceedings of the IEEE Conference on Pattern Recognition and Image Processing*, Dalls, Texas, August 1981, pp. 103–108.

[Nagel 81b] Nagel, H. H. and Neumann, B. (1981). On 3-D Reconstruction from Two Perspective Views. In *Proceedings of the International Joint Conference on Artificial Intelligence*, Vancouver, Canada, pp. 661–663.

[Ogawa 86] Ogawa, H. (1986). Labeled Point Pattern Matching by Delaunay Triangulazation and Maximal Cliques. *Pattern Recognition*, Vol. **19**, No. 1, pp. 35–40.

[Pázman 84] Pázman, A. (1984). Nonlinear Least Squares—Uniqueness versus Ambiguity. *Mathematische Operationsforschung und Statistik, Series Statistics*, Vol. **15**, pp. 323–336.

[Rademacher 19] Rademacher, H. (1919). Über Partielle und Totale Differenzierbarkeit von Funktionen Mehrerer Variabeln und über die Transformation der Doppelintegrale. *Mathematische Annalen*, Vol. **79**, pp. 340–359.

[Ranade 80] Ranade, S. and Rosenfeld, A. (1980). Point Pattern Matching by Relaxation. *Pattern Recognition*, Vol. **12**, pp. 269–275.

[Roach 79] Roach, J. W. and Aggarwal, J. K. (1979). Computer Tracking of Objects Moving in Space. *IEEE Transactions PAMI*, Vol. **PAMI-1**, No. 2, April 1979, pp. 127–135.

[Roach 80] Roach, J. W. and Aggarwal, J. K. (1980). Determining the Movement of Objects from a Sequence of Images. *IEEE Transactions PAMI*, Vol. **PAMI-2**, No. 6, November 1980, pp. 554–562.

[Royden 68] Royden, H. L. (1968). *Real Analysis*. 2nd ed. MacMillan, New York.

[Seidl 74] Seidl, R. A. (1974). A Theory of Structure and Encoding of Visual patterns with Applications to Character Recognition. Ph.D. Dissertation, University of Newcastle, Australia.

[Simon 72] Simon, J. C., Checroun, A. and Roche, C. (1972). A Method of Comparing Two Patterns Independent of Possible Transformations and Small Distortions. *Pattern Recognition*, No. 4, pp. 73–81.

[Tsai 81] Tsai, R. Y. and Huang, T. S. (1981). Estimating 3-D Motion Parameters of a Rigid Planar Patch, I. *IEEE Transactions ASSP*, Vol. **ASSP-29**, No. 6, December 1981, pp. 1147–1152.

[Tsai 82] Tsai, R. Y., Huang, T. S. and Zhu, W. L. (1982). Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch, II: Singular Value Decomposition. *IEEE Transactions ASSP*, Vol. **ASSP-30**, No. 4, August 1982, pp. 525–534.

[Tsai 84] Tsai, R. Y. and Huang, T. S. (1984). Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surface. *IEEE Transactions PAMI*, Vol. **PAMI-6**, No. 1, January 1984, pp. 13–26.

[Ullman 79] Ullman, S. (1979). *The Interpretation of Visual Motion*. MIT Press, Cambridge, Massachusetts.

[Wichmann 82] Wichmann, B. A. and Hill, I. D. (1982). Algorithm AS 183. An Efficient and Portable Pseudo-Random Number Generator. *Applied Statistics*, Vol. **31**, No. 2, pp. 188–190.

[Zahn 74] Zahn, C. T. (1974). An Algorithm for Noisy Template Matching. In *Proceedings of the International Federation for Information Processing*, pp. 727–732.

END
9-87
DTIC