Statistical properties of the Hough transform estimator

Itai Dattner

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE MASTER DEGREE

University of Haifa
Faculty of Social Sciences
Department of Statistics

June 2007
Statistical properties of the Hough transform estimator

By: Itai Dattner

Supervised by: Prof. Alexander Goldenshluger

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE MASTER DEGREE

University of Haifa
Faculty of Social Sciences
Department of Statistics

June 2007

Approved by: ___________________________ Date: ___________________________
(Supervisor)

Approved by: ___________________________ Date: ___________________________
(Chairperson of M.A. Committee)
I am deeply grateful to my advisor, Professor Alexander Goldenshluger, for his tremendous support throughout this project. Without his professional and financial support this thesis could not have been completed. Working with Alex has been inspiring; the calm and friendly atmosphere that he projects has made this a wonderful experience. As for Alex’s professional support, his in-depth knowledge in a broad area of subjects and his ability to quickly understand the issues and guide me towards fruitful solutions was both useful and astounding.

I would also like to thank Hana Avraham and Offra Blum for their moral and technical support. Hana and Offra helped in making my time at the Statistics Department a wonderful period.
Contents

Acknowledgment ii

Abstract v

List of figures vi

Notation vii

1 Introduction 1
   1.1 The Hough transform estimator .......................................... 1
   1.2 Literature survey .................................................................. 6
      1.2.1 Algorithm enhancement .................................................. 6
      1.2.2 Detection of general shapes .......................................... 7
      1.2.3 Applications of the HT ............................................... 7
      1.2.4 Statistical procedures related to the HT ......................... 8
      1.2.5 Statistical concepts relevant to our model ...................... 12
   1.3 Summary .............................................................................. 14

2 Preliminaries 16
   2.1 Consistency ......................................................................... 16
      2.1.1 VC-classes (Vapnik-Cervonenkis-classes) ...................... 17
   2.2 Rate of convergence and limit distribution .......................... 18
      2.2.1 Rate of convergence ..................................................... 18
2.2.2 Limit distribution .............................................. 20

3 Statistical properties of the HT estimator 22
  3.1 Observation model ............................................. 22
    3.1.1 Asymptotic results ......................................... 23
    3.1.2 Equivariance properties ................................... 25
    3.1.3 Robustness .................................................. 25
  3.2 Proofs ................................................................ 26
    3.2.1 Proof of Theorem 3.1 .......................................... 26
    3.2.2 Proof of Theorem 3.2 .......................................... 27
    3.2.3 Proof of Theorem 3.3 .......................................... 31
    3.2.4 Proof of Theorem 3.4 .......................................... 34

4 Simulations results .................................................... 36
  4.1 Choice of the parameter $r$ ....................................... 36
  4.2 Vertical line ....................................................... 39
  4.3 Robustness .......................................................... 43
    4.3.1 Breakdown point .............................................. 43
    4.3.2 Equivariance .................................................. 45

Appendix ................................................................. 47

References ............................................................. 50
Statistical properties of the Hough transform estimator

Itai Dattner

Abstract

The Hough transform is a common computer vision algorithm used to detect shapes in a noisy image. Originally the Hough transform was proposed as a technique for detection of straight lines in images. The objective of this algorithm is to find the line that "best" fits a set of planar points. In this work we study the statistical properties of the Hough transform estimator for polar parameterization of the straight line. We consider the simple case of one line detection.

We show that the estimator is consistent, and establish its rate of convergence and limiting distribution. The rate of convergence turns out to be $n^{1/3}$, which is slower than the standard parametric rate $n^{1/2}$. However, we study the break down properties of the Hough transform estimator and conclude that it is very robust. Numerical results of the properties of the Hough transform estimator are discussed as well. In particular, we performed an extensive experiment in order to define a "rule of thumb" for determination of the optimal width parameter of the template used in the algorithm.
## List of Figures

1.1 The Hough transform: (a) Original domain. (b) Hough domain. .................. 1

1.2 A test image. ......................................................................................................... 2

1.3 (a) Relevant data from the image of Figure 1.2. (b) The data after edge detection. ................................................................................................................................. 3

1.4 The Hough transform for the data points in Figure 1.3(b). ......................... 4

1.5 The accumulator array corresponding to Figure 1.4. ................................. 4

4.1 An illustration of the dual plot for different choices of $r$: (a) large value of $r$; (b) ”optimal” $r$; (c) small value of $r$. ................................................................. 37

4.2 Histogram of $\hat{r}_j$, $j = 1, ..., M$ for $M = 192$ experiments. In each experiment $r \in \mathcal{R} = [0.276 : (0.025) : 0.976]$, i.e., 29 potential values. For each value of $r$ (4.2) is calculated based on a sample size $n = 200$, and $N = 200$ simulations. 38

4.3 Boxplot for the experiments data. ................................................................. 38

4.4 The empirical total variance (4.2). ................................................................. 39

4.5 Distribution of $W_{0.276,300}$ based on 250 simulations. ......................... 40

4.6 Histogram of $W_{0.276,300}$................................................................. 40

4.7 Distribution of $W_{0.6,300}$ based on 250 simulations. ......................... 41

4.8 Histogram of $W_{0.6,300}$................................................................. 41

4.9 Distribution of $W_{0.976,300}$ based on 250 simulations. ......................... 42

4.10 Histogram of $W_{0.976,300}$................................................................. 42
4.11 The accumulator array for vertical line detection. The estimated parameters are \( \hat{\phi} = 0, \hat{t} = 0.59596 \).

4.12 (a) An illustration of the breakdown properties of the HT estimator. The cluster on the right is the contamination of 20 ”bad” observations. (b) The accumulator array.

4.13 (a) An illustration of the breakdown properties of the HT estimator. The cluster on the right is the contamination of 24 ”bad” observations. (b) The accumulator array.

4.14 Equivariance property of the HT. The transformation is rotation of \( \pi/8 \) of the data set. (a) The blue cloud is the original data set while the red is the transformed data. (b) The blue and red sinusoids are defined respectively to the data sets.
Notation

\( \mathbb{R}^1 \) real line
\( \mathbb{R}^+_1 \) positive real line
\( \mathbb{R}^d \) \( d \) dimensional space
\( \lfloor x \rfloor \) the largest integer \( n \leq x \)
\( \lceil x \rceil \) the smallest integer \( n \geq x \)
\( x \land y \) minimum of \( x \) and \( y \)
\( x \lor y \) maximum of \( x \) and \( y \)
\( \| x \| \) Euclidean norm of \( x \)
\( x_n \to x \) \( \lim_{n \to \infty} x_n = x \)
\( \approx \) approximation
\( \xrightarrow{a.s.} \) convergence almost surely
\( \xrightarrow{p} \) convergence in probability
\( \xrightarrow{d} \) convergence in distribution
\( \Theta \) parameter space
\( Z \) observations space
\( 1_A \) the indicator function of the set \( A \)
\( P \) probability measure
\( P_n \) empirical probability measure of observations,
\( P_n(A) = \frac{1}{n} \sum_{i=1}^{n} 1 \{ Z_i \in A \} \)
\( E \) expectation with respect to \( P \)
\( Z_n = O_p(1) \) if for all \( \epsilon > 0 \) there exist a constant \( M_\epsilon \) for which \( P(|Z_n| \geq M_\epsilon) < \epsilon \)
\( Z_n = o_p(1) \) if \( Z_n \xrightarrow{p} 0 \)
Chapter 1

Introduction

1.1 The Hough transform estimator

In the area of computer vision, the Hough transform [Hough (1959)] is a standard tool of image analysis which allows recognition of patterns in an image space. The recognition of patterns is done by detecting parameterized curves in the image (straight lines, polynomials, circles, etc.). In order to describe the main idea underlying the Hough transform (HT) let us consider the set of points $(X_1, Y_1), ..., (X_5, Y_5)$ displayed in Figure 1.1(a). Four points out of the five are collinear and the straight line underneath them can be described in the polar form as

$$X \cos \phi_0 + Y \sin \phi_0 = t_0,$$

![Figure 1.1: The Hough transform: (a) Original domain. (b) Hough domain.](image)
where $t_0 = 0.5$ is the perpendicular distance from the origin to the line, and $\phi_0 = \frac{3\pi}{4}$ is the angle between the normal to the line and the $X$-axis. The Hough transform maps each point $(X_i, Y_i)$ in the original $(X, Y)$-plane to the sinusoidal curve

$$C_i(\phi, t) = \{(\phi, t) : X_i \cos \phi + Y_i \sin \phi = t\}$$

(1.1)

in the $(\phi, t)$-plane which is usually referred to as the Hough domain. Figure 1.1(b) displays the HT for the point set shown in Figure 1.1(a). Curves $C_i(\phi, t)$ corresponding to the four collinear points intersect at a single point with coordinates $(\phi_0, t_0)$ given above. Thus, a problem of line detection in the original domain can be reduced to a search of the intersection point in the Hough domain. This property of the Hough transform is the main contribution of the method.

In practice, the search for intersection point is performed as follows. The Hough domain is quantized into cells, and for each cell the number of sine curves crossing is counted. The cell with the maximum number of crossing curves is a natural estimator of the parameters of the line in the image domain.

Let us illustrate an implementation of the HT on the real image given in Figure 1.2. The red arrow in Figure 1.2 points at a fragment (straight line) of the image we want to investigate while Figure 1.3(a) displays a close-up view of this fragment.

Usually, the HT for image points is evaluated after a preprocessing step which called
Figure 1.3: (a) Relevant data from the image of Figure 1.2. (b) The data after edge detection.

*Edge detection.* Edge detection of an image reduces significantly the amount of data and filters out information that may be regarded as less relevant, preserving the important structural properties of an image. Figure 1.3(b) displays the data after edge detection, while in Figure 1.4 the Hough transform for the data points in Figure 1.3(b) is shown. Figure 1.4 displays 83 sine curves corresponding to 83 data points. The sine curves do not intersect at a single point because of the noisy data. However, as mentioned above, the Hough domain is typically quantized into cells, and a cell with maximum number of sine curves crossing is found.

The quantization process is as follows. Divide the Hough domain into $m \times k$ rectangular cells, each of size $\delta \phi \times \delta t$. Let $S_{i,j}, i = 1, ..., m, j = 1, ..., k$ denote the cell centered at $(\phi_i, t_j)$, and let $H(\phi_i, t_j)$ be the number of sine curves crossing this cell. The coordinate $(\phi_i, t_j)$ for which the value of $H(\phi, t)$ is maximized is a natural estimator for the parameters of the line and denoted by $(\hat{\phi}, \hat{t})$. This procedure results in a 3D array which referred to as the *accumulator array*. The accumulator array corresponding to the graph in Figure 1.4 is displayed in Figure 1.5. A counting process in the example above showed that maximum number of curves cross the cell $S_{24,79}$. In particular $H(\phi_{24}, t_{79}) = 21$. Thus, the HT estimator is $(\hat{\phi}, \hat{t}) = (\phi_{24}, t_{79}) = (\frac{\pi}{4}, 23.0057)$ and we conclude that the data points in the fragment under investigation correspond to a straight line with parameters $(\hat{\phi}, \hat{t})$ given above.

Formally, the HT estimator is defined as follows. Let $\theta = (\phi, t)$, and let data points $(X_1, Y_1), ..., (X_n, Y_n)$ be given on the plane. Each observation pair $(X_i, Y_i)$ defines a sine curve in the Hough domain, $C_i(\theta)$ as given by (1.1). Now, let $S(\theta)$ be the rectangular
Figure 1.4: The Hough transform for the data points in Figure 1.3(b).

Figure 1.5: The accumulator array corresponding to Figure 1.4.
parameter cell in the Hough domain

\[ S(\theta) = \{ \theta' = (\phi', t') : |\phi' - \phi| \leq \frac{\delta_\phi}{2}, |t' - t| \leq \frac{\delta_t}{2} \} \] (1.2)

centered at \( \theta = (\phi, t) \), where \( \delta_\phi > 0, \delta_t > 0 \) are the quantization parameters. Then, the HT estimator is defined as

\[ \hat{\theta}_n = (\hat{\phi}, \hat{t}) = \arg \max_{\theta \in [0, 2\pi) \times R^1_+} \frac{1}{n} \sum_{i=1}^{n} 1\{ S(\theta) \cap C_i(\theta) \neq \emptyset \}. \] (1.3)

Equation (1.3) implies that in defining an HT we have a few important choices to make. After choosing a parameterization for the problem, a cell size and a cell shape need to be determined. Princen, Illingworth, and Kittler (1992) studied different types of parameterization and cell shapes such as the rectangular cell, a disk with a finite radius and a line-segment cell. They suggest that an HT based on the polar parameterization and the line-segment cell is the appropriate choice in most cases.

In this thesis we study the following version of the HT with line segment cell. In particular, let \( \delta_\phi = 0 \) and \( \delta_t = r > 0 \), then \( S(\theta) \) given in (1.2) is a line-segment-cell centered at \( \theta = (\phi, t) \). Note that \( S(\theta) \cap C_i(\theta) \neq \emptyset \) if and only if the distance between the curve \( C_i(\theta) \) and the line-segment center is less than or equal to \( r/2 \). Let

\[ M_{r,n}(\theta) = \frac{1}{n} \sum_{i=1}^{n} 1\{ |X_i \cos \phi + Y_i \sin \phi - t| \leq \frac{r}{2} \}. \]

Then, the HT estimator takes the following form:

\[ \hat{\theta}_{r,n} = (\hat{\phi}, \hat{t}) = \arg \max_{\theta \in [0, 2\pi) \times R^1_+} M_{r,n}(\theta). \] (1.4)

For each specific parameterization and a specific cell size and shape, the HT estimator admits the following geometrical interpretation in the original domain. Let

\[ D_\theta = \{(x, y) : |x \cos \phi + y \sin \phi - t| \leq \frac{r}{2} \}, \quad \theta = (\phi, t) \in [0, 2\pi) \times R^1_+; \] (1.5)

then \( D_\theta \) is the set of all points of the plane lying between two lines with parameters \( (\phi, t + \frac{r}{2}) \) and \( (\phi, t - \frac{r}{2}) \). Hence, the HT estimator given by (1.4) seeks the value \( \theta \) such that the corresponding set \( D_\theta \) covers the maximal number of data points. The set \( D_\theta \) is referred to as the template of the HT in the observation space. Different choices of parameterization, a cell size and a cell shape in the Hough domain result in a different template in the original domain.
Although the HT is a computer vision tool, it can be considered as a statistical technique for line estimation. However, while inference using regression method is well understood, there are only a few papers studying the statistical properties of the HT. In particular, Goldenshluger and Zeevi (2004) studied the HT for the Cartesian parameterization and a circular cell shape. In this work we investigate the statistical properties of the HT estimator for the polar parameterization and a line-segment cell shape. We prove consistency of the HT estimator and derive its asymptotic distribution.

In the next section we present a literature survey. The objective of this survey is merely to point to relevant literature and to indicate the place of our research within it.

1.2 Literature survey

The vast research on the Hough transform can be classified into four categories: algorithm enhancement, detection of general shapes, applications, and statistical properties of the HT. Our survey is organized accordingly. Since our main interest is statistical in nature, we focus on the statistical properties of the HT.

The Hough transform was first introduced as a method of detecting complex patterns of points in a binary image data [Hough (1959)]. It was developed in connection with the study of particle tracks through the viewing field of a bubble chamber and was patented in 1962 [Hough (1962)]. The original Hough method was based on the Cartesian parameterization of a line. The use of this method led to the practical difficulty of an unbounded parameter space as both the slope and the intercept are unbounded. Duda and Hart (1972) suggested the polar parameterization that eliminates this problem. Moreover, they showed how the method can be extended to find more general classes of curves in picture. The works of Hough and Duda and Hart laid the foundations of the HT technique.

1.2.1 Algorithm enhancement

Studies in the area of the HT enhancement are motivated by problems inherent in the HT method. These are its numerical complexity and large storage requirements. The former is due to the fact that the HT increments the cells in the accumulator array corresponding to all curves that pass through all points in the image. Thus, much of the computation time is spent storing counts for parameter cells which are not relevant to the shape under
detection. The storage requirements of the HT are a bigger problem though, since the size of the accumulator array is exponential in the number of parameters. When the number of parameters is greater than two, the storage space requirements become excessive. These problems led to a new class of HT called the randomized/probabilistic Hough transform (RHT/PHT). It is based on the observation that, in many cases, accumulation procedure for all parameter cells is not necessary for the reliable detection of shapes. Significant computational savings are obtained by limiting the counting process, i.e., by using just a subset of the curves obtained from the HT. The RHT [Xu et al. (1990), Xu and Oja (1993)] is a many-to-one algorithm in which pairs of points are randomly selected. Each of the $n(n - 1)/2$ pairs of data points defines a line and thus votes for a single accumulator. The PHT [Kiryati et al. (1991)] is a one-to-many algorithm (in the sense that one data point votes for many accumulators). It is similar to the standard HT but uses just a small random subset of the data points. Kiryati et al. (2000) describe merits and drawbacks of the RHT and PHT. They set a unified framework for the analysis of the RHT and PHT and provide guidelines for using the best algorithm in a given application.

1.2.2 Detection of general shapes

We discussed implementation of the HT for the estimation of line patterns. The same idea, however, can be used for the detection of other shapes. Kimme, Ballard and Sklansky (1975) showed how circular arcs could be detected and applied this technique to medical image processing. Essential to their method was the use of edge direction information to constrain the range of parameters which had to be addressed. Detection of parabolas and ellipses was investigated by Tsukune and Goto (1983). Ballard (1981) showed how generalized HT can efficiently find arbitrary shapes for any orientation or any scale, including shapes which cannot be represented analytically.

1.2.3 Applications of the HT

The HT is a valuable method in a large range of machine vision problems. One of the major reasons is that straight lines and other simple shapes occur in most natural and man made scenes [Illingworth and Kittler (1988)]. Examples of the use of the HT are numerous. Focusing on recent applications we can mention the following references. Dobes et al (2006) developed successful computer method of eyes and eyelids localization using a mod-
ified Hough transform. Application of the Hough transform for seed row localization was studied by Leemans and Destain (2006). Rover, Klefenz, and Weihs (2004) showed that application of the Hough transform to digitized sounds yields a useful sound characterization. The generated data allows to distinguish between sounds played by different instruments. Machine vision inspection of rice seed based on Hough transform was conducted by Cheng and Ying (2004). Reed and Parker (1996) provided an overview to the implementation of Lemon, a complete optical music recognition system. The Hough transform is one of the techniques employed by the implementation.

1.2.4 Statistical procedures related to the HT

In this subsection we describe statistical techniques closely related to the HT.

**Dual plot.** As mentioned above, the Hough transform in its polar form maps each point \((X_i, Y_i)\) in the original \((X, Y)\)-plane to a sinusoidal curve in the \((\phi, t)\)-plane. In the statistical literature this concepts is referred to as the **dual plot**. The dual plot interchanges the roles of points and lines. This technique appears in the literature for the Cartesian parameterization of a line.

Daniels (1954) used the dual plot to develop a distribution-free test for the hypothesis that all regression parameters have specified values. In particular, observations \((X_i, Y_i)\) are assumed to follow the linear regression model

\[
Y_i = a_0 + b_0 X_i + \epsilon_i, \quad i = 1, \ldots, n,
\]  

(1.6)

where \(b_0, a_0\) are unknown parameters, and \(\epsilon_i\) are zero mean random variables with a continuous and symmetric distribution. The aim is to test the hypothesis that \(b_0 = b^*\), and \(a_0 = a^*\). Under the assumption that there are no repeated values of the independent variable, the \(n\) lines over the \((b, a)\)-plane

\[
a = -X_i b + Y_i,
\]  

(1.7)

partition the plane into \(\frac{1}{2}(n^2 + n + 2)\) polygonal regions with probability one. Some of these regions, \(2n\), are open and extending to infinity and the remaining \(\frac{1}{2}(n - 1)(n - 2)\) form a set of contiguous closed regions. Each line passes above or below the point \((b_0, a_0)\) according as the corresponding \(\epsilon_i\) is positive or negative. Under the null hypothesis that either event is equally likely, we should expect that the point \((b_0, a_0)\) will be situated somewhere near
the middle of the set of closed regions rather than in or near one of the open regions. This idea motivates the $m$ test developed in Daniels (1954). The test assigns a score to each region equal to the minimum number $m$ of lines which have to be crossed to escape from it into one of the open regions. For example, for an open region, $m = 0$. The hypothesis $b_0 = b^*, a_0 = a^*$ is rejected if the score $m$ for the region containing $(b_0, a_0)$ is significantly low. The distribution of $m$ is given by

$$P(m \leq m_0) = \frac{(n - 2m_0)}{2^{n-1}} \sum_j \left(\frac{n}{n - m_0} + j(n - 2m_0)\right), \quad j = 0, \ldots, \left\lfloor \frac{m_0}{n - 2m_0} \right\rfloor,$$  \hspace{1cm} (1.8)

where $\lfloor x \rfloor$ stands for the largest integer that is less than or equal to $x$. The maximum possible value of $m$ is $\lfloor \frac{1}{2}(n - 1) \rfloor$. In Daniels (1954) a table of $P(m \leq m_0)$ for $n$ ranging from 3 to 30 is presented and asymptotic results are developed. Based on the results above, confidence regions for $(b_0, a_0)$ can be derived. For example, by (1.8) the probability that $(b_0, a_0)$ lies in an open region is $P(m = 0) = \frac{n}{2^{n-1}}$. So the largest closed polygon formed by the lines is a confidence region for $(b_0, a_0)$ with probability $1 - n/2^{n-1}$. Generally, let $CR$ be the union of all regions for which $m > m_0$. Then the confidence region is given by

$$P((b_0, a_0) \in CR) = P(m > m_0).$$

A robust line estimation technique (called resistant line) based on the dual plot was proposed by Johnstone and Velleman (1985). The resistant line is an exploratory method resistant to outliers in $x$ and $y$. Here we describe the main idea underlying this method.

Consider the model (1.6) where the $X$-values are ordered such that $X_1 \leq X_2 \leq \cdots \leq X_n$, and divide them into three groups, $L$, $M$, and $R$ containing $n_L$, $n_M$, and $n_R$ data points respectively ($n_L + n_M + n_R = n$). Assume that $n_M = 0$, and denote by

$$e_i(b) = Y_i - X_i b$$

the residual in the point $(X_i, Y_i)$ (ignoring $a$). Mapping the $n$ residuals from the original plane onto the parameter plane results in $n$ lines, each with slope $-X_i$ and intercept $Y_i$ corresponding to the original data point $(X_i, Y_i)$. Note that the $n$ lines are ordered with respect to their slopes as a result of the ordered $X$-values in the original plane. Thus, the horizontal grouping in the original plane is now vertical. If we plot the pointwise median of the lines for each group then we obtain two piecewise-linear curves which intersect only once. The $b$ value of this intersection is the slope of the resistant line. Thus, the resistant line slope estimate $\hat{b}$ is defined as the solution to

$$\text{med}_{i \in L} e_i(b) = \text{med}_{i \in R} e_i(b).$$
The intercept estimate \( \hat{a} \) may be chosen to make the median residuals of both groups equal to zero.

The authors conclude that if \( n_L \) and \( n_R \) are both odd, then \( \hat{b} \) is median unbiased for \( b_0 \) without any assumption of symmetry on the \( \epsilon_i \). If the errors are symmetric about 0 and independent (not necessarily identically distributed), then \( \hat{b} \) and \( \hat{a} \) are symmetrically distributed about \( b_0 \) and \( a_0 \) respectively.

Rousseeuw and Hubert (1999) proposed a line estimation technique (called regression depth) with a dual plot interpretation. Here we describe the method for the case of simple regression. Consider the model (1.6), and denote by \( \hat{\theta} = (\hat{b}, \hat{a}) \) the vector of coefficients of a line we fit to the data points. Assume we fit a line which lies above the whole data points. Clearly, this is a bad fit. Note that rotating this line until it is vertical can be done without passing any observation. Now, assume we fit a line which crosses the observations cloud. By rotating this line to a vertical position we cross some observations. The regression depth of an estimator \( \hat{\theta} \) is the minimum number of observations we cross while rotating the line until it is vertical. The best estimator in that sense will have the highest depth.

Consider the dual plot representation of (1.6) given by (1.7). Here, the depth of \( \hat{\theta} \) is the smallest number of lines that need to be removed to set \( \hat{\theta} \) in one of the open regions of the line’s arrangement. For any set of \( n \) lines in the parameter plane, there exists a point \( \hat{\theta} \) that can be set in one of the open regions only by removing at least \( \lfloor n/3 \rfloor \) lines. For a simple regression model (under regular assumptions) the regression depth of \( \hat{\theta} \) corresponds to the test statistic \( m \) given in Daniels (1954).

**M-estimators.** In the previous paragraphs we presented statistical techniques related to the HT in the sense of the dual plot. Here we describe a different point of view. The HT estimator (1.4) is defined as the solution to a maximization problem. In the statistical literature this problem is related to the family of M-estimators. For example, consider the classical linear model given in (1.6). The aim of linear regression is to estimate \( \theta_0 = (b_0, a_0) \) from the data \( Z_i = (X_i, Y_i) \). A family of M-estimators is defined as

\[
\hat{\theta} = \arg \max_{\theta} \sum_{i=1}^{n} m_{\theta}(Z_i),
\]

where \( m_{\theta} \) is a user-chosen function. If the observations \( Z_i = (X_i, Y_i) \) are i.i.d. random variables, and \( f \) denotes a joint probability density of observations \( Z \), and if \( m_{\theta}(Z) = \log f(Z, \theta) \), then \( \hat{\theta} \) is the maximum likelihood estimator of \( \theta_0 \). In the case of the HT given
by (1.4), the function $m_\theta(Z)$ takes the form

$$m_\theta(Z) = 1\{|X_i \cos \phi + Y_i \sin \phi - t| \leq \frac{r}{2}\}.$$ 

There is a well-developed theory of M-estimators [e.g. Huber (1981), Van der Vaart and Wellner (1996), Chapter 3.2]. In most M-estimators the function $m_\theta$ is smooth and therefore the estimators converge at $\sqrt{n}$-rate. In the case of HT, the function $m_\theta$ is not smooth, thus leads to a different asymptotics behavior. This phenomenon is described in detail in the next paragraph.

Mode estimation and cube root asymptotics. The HT estimator is closely related to the Chernoff mode estimator. Chernoff (1964) introduced the following nonparametric estimator of the mode of a univariate unimodal probability density $f$. Let $a > 0$, and $X = (X_1, \ldots, X_n)$ be an independent sample from $f$. For a given point $\theta \in \mathbb{R}$ define

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \in [\theta - a, \theta + a]\},$$

and

$$\hat{\theta}_n = \arg \max_{\theta \in \mathbb{R}} M_n(\theta).$$

In words, $M_n(\theta)$ is the proportion of observations $X_i$ belonging to the interval $[\theta - a, \theta + a]$, and $\hat{\theta}_n$ is the center of the interval of length $2a$ containing the maximal number of data points. Chernoff (1964) studied properties of the mode estimator (1.9) and showed that $\hat{\theta}_n$ has a non-standard asymptotic behavior. In particular, $\hat{\theta}_n$ converges to the mode $\theta_0$ of $f$ at the rate $n^{1/3}$:

$$(nC)^{1/3}(\hat{\theta}_n - \theta_0) \xrightarrow{d} W, \quad n \to \infty,$$

where $C = [f'(\theta_0 - a) - f'(\theta_0 + a)]^2 [8f(\theta_0 + a)]^{-1}$, $W$ is a maximizer of the random process $Z(z) - z^2$, and $Z(\cdot)$ is the standard two-sided Brownian motion. The cube root asymptotics indicated in (1.10) is a consequence of the fact that $M_n(\theta)$ is discontinuous as a function of $\theta$. Kim and Pollard (1990) explain in detail the nature of the cube root asymptotics phenomenon and generalize the above results to higher dimensions. Groeneboom and Wellner (2001) present algorithms and programs for computing the distribution of $W$ and its quantiles.

It turns out that the HT estimator also has the cube root asymptotics. The rate of convergence of the HT estimator was rigorously described in Goldenshluger and Zeevi (2004).
They studied the statistical properties of the HT estimator in the Cartesian parameteriza-
tion, and a circular cell shape. We briefly review their main results here.

Consider the model (1.6) where $X$ is independent of $\epsilon$, and $\epsilon$ is a random variable with
bounded, symmetric and strictly unimodal density $f_\epsilon$ which is continuously differentiable
with bounded first derivative. Define

$$\hat{\theta}_{r,n} = \arg \max_{\theta \in \mathbb{R}^2} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{|X_i b + a - Y_i|^2 \leq r^2(X_i^2 + 1)\}$$

to be the HT estimator. It turns out that for every fixed $r > 0$,

$$n^{1/3}(\hat{\theta}_{r,n} - \theta_0) \xrightarrow{d} W,$$

where $W$ has the distribution of the (almost surely unique) maximizer of the process
$\theta \mapsto \frac{1}{2} \theta^T V_0 \theta + G(\theta)$. Here

$$V_0 = \mathbb{E}\{[f_\epsilon'(r\|Z\|) - f_\epsilon'(-r\|Z\|)]ZZ^T\}$$

(1.11)

$Z = (X, 1)^T$, and $G$ is a zero-mean Gaussian process with continuous sample paths and
stationary increments such that for any $\xi, \eta \in \mathbb{R}^2$,

$$\mathbb{E}[G(\xi) - G(\eta)]^2 = 2\mathbb{E}\{f_\epsilon(r\|Z\|)|Z^T(\xi - \eta)|\}. \quad (1.12)$$

The cube-root rate of convergence is due to the discontinuous nature of the indicator
function involved in the HT estimator.

1.2.5 Statistical concepts relevant to our model

Measurement error model. Measurement error (ME) models are a generalization of
the standard regression model. Let $(\tilde{X}, \tilde{Y})$, $i = 1, \ldots, n$ be random variables lying on the
unknown line with parameters $(b_0, a_0)$, i.e.

$$\tilde{Y}_i = a_0 + b_0 \tilde{X}_i.$$

The ME model assumes that we cannot observe $(\tilde{X}, \tilde{Y})$ directly. Instead we observe the random variables

$$X_i = \tilde{X}_i + \gamma_i,$$

$$Y_i = \tilde{Y}_i + \eta_i,$$
where $\gamma_i$ and $\eta_i$ are noise variables. The independent variable $X_i$ is either fixed or random and is assumed to be uncorrelated with the errors given above. This model is called also error-in-variables model.

There are several types of ME models. If the $X_i$ are assumed to be deterministic, then the model is known as the functional model; whereas, if the $X_i$ are i.i.d. random variables and independent of the errors, the model is known as the structural model. A generalization of these two models is the ultrastructural model which assumes that the $X_i$ are independent but not identically distributed.

One of the most important differences between ME models and ordinary regression models concerns model identifiability. Generally speaking, this means that different sets of parameters can lead to the same joint distribution of $(X, Y)$. For example, if we assume that all the random variables are normal, then the structural model is not identifiable [Cheng and Van Ness (1999)]. In this situation it is impossible to estimate consistently the parameters from the data, because the limit of a consistent estimator has to be unique.

The ME models discussed so far are linear where in this work we consider a model in the polar parameterization. However, the arguments given above are still relevant, and are taken into account in the theoretical study of Chapter 3.

Robustness and breakdown point. The HT estimator shares an important statistical property called robustness. Generally speaking, the goal of robust statistics is to develop data analytical methods which are resistant to outlying observations in the data, and which are also able to detect these outliers. An extensive theory of robustness has been developed [Huber (1981), Donoho and Huber (1983), Rousseeuw and Leroy (1987)].

For example, consider the known least squares (LS) estimator. This technique is criticized for its dramatic lack of robustness. Indeed, one single outlier can have an arbitrarily large effect on the estimate. Thus, a robust estimator should be resistant to contamination in the data.

One formalization of robustness properties is given by the breakdown point (BP) of an estimator. Here we present two finite-sample versions of the breakdown point, presented by Donoho and Huber (1983).

Let $Z_n = \{Z_1, ..., Z_n\}$ be a random sample, then the addition breakdown point is given
by
\[ \epsilon_{\text{add}}(\hat{\theta}; Z_n) = \min \left\{ \frac{k}{n+k} : \sup_{Z'_k} \| \hat{\theta}(Z_n \cup Z'_k) - \hat{\theta}(Z_n) \| = \infty \right\}. \]

In words, the addition BP is the minimal contamination fraction that should be added to the sample such that the difference between the estimator of the contaminated sample to the estimator of the original sample is infinite.

The replacement breakdown point of \( \hat{\theta} \) is defined by
\[ \epsilon_{\text{rep}}(\hat{\theta}; Z_n) = \min \left\{ \frac{k}{n} : \sup_{Z'_k} \| \hat{\theta}(Z'_k) - \hat{\theta}(Z_n) \| = \infty \right\}, \]

where \( Z'_k \) denotes the corrupted sample obtained from \( Z_n \) by replacing \( k \) data points of \( Z_n \) with arbitrary values. Here again, the replacement BP is the least fraction of contamination that can move the value of the estimator to infinity.

For example, for the least square estimator \( \epsilon_{\text{rep}}(\text{LS}; Z_n) = 1/n \) which tends to zero for increasing sample size \( n \). Therefore it can be said that LS has a breakdown point of 0%, which is bad in the sense of robustness. Goldenshluger and Zeevi (2004) showed that for large \( r \), the HT estimator (in the Cartesian parameterization) has a BP close to 50%. In Chapter 3 we use the notion of breakdown point in order to characterize the robustness of the HT estimator in its polar form.

1.3 Summary

In this work we study the HT estimator for the polar parameterization of the straight line. In particular, we consider the following error-in-variables model.

Let \((\tilde{X}_i, \tilde{Y}_i), i = 1, \ldots, n\) be i.i.d. points lying on the unknown line with parameters \((\phi_0, t_0)\)
\[ \tilde{X}_i \cos \phi_0 + \tilde{Y}_i \sin \phi_0 = t_0. \]

We assume that we observe \((X_i, Y_i), i = 1, \ldots, n\) i.i.d. random variables given by
\[ X_i = \tilde{X}_i + \gamma_i, \]
\[ Y_i = \tilde{Y}_i + \eta_i, \]

where \( \gamma_i \) and \( \eta_i \) are noise variables independent of \((\tilde{X}_i, \tilde{Y}_i)\). Our goal is to study properties of the HT algorithm defined in (1.4) under these circumstances.
The rest of this thesis is organized as follows. In Chapter 2 we review technical tools for the study of asymptotic properties of M-estimators. Chapter 3 presents the main results of this work. We prove consistency of the HT estimator, study its limit distribution, and discuss equivariance and robustness properties. In particular, the HT estimator is consistent, and has the rate of convergence of $n^{1/3}$ as $n \to \infty$. The limit distribution is complicated, and depends on the unknown parameter $\phi_0$. Moreover, we show that the HT estimator is rotation equivariant, and that its breakdown point is close to 50% for large $r$.

Chapter 4 presents numerical results. We discuss the choice of the parameter $r$, and introduce result of several extensive experiments we performed. Our numerical experience suggests that a good choice of this parameter, in the sense of minimizing the variance of the HT estimator, is $r = 3\sigma$, where $\sigma$ is the standard deviation of the noise variables $\gamma$ and $\eta$. This choice was verified in many different simulations setups. We illustrate an example of detection of vertical line, and show that the HT estimator yields very good results. We illustrate the behavior of the HT under contamination in the data, and we present an example of its equivariance property.
Chapter 2

Preliminaries

In this chapter we present several results on asymptotic properties of M-estimators. These results will be used in Chapter 3 while we study the HT estimator.

Let $Z_1, \ldots, Z_n$ be i.i.d. random variables in sample space $Z \subseteq \mathbb{R}^d$, having a common distribution $P$ that depends on unknown parameter $\theta_0 \in \Theta \subseteq \mathbb{R}^k$. In what follows we assume that $\Theta$ is an open set. Let $m_\theta : Z \to \mathbb{R}^1$ be a known function parameterized by $\theta \in \Theta$. Denote the criterion function by

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} m_\theta(Z_i). \quad (2.1)$$

An M-estimator of $\theta_0$ is defined by

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} M_n(\theta). \quad (2.2)$$

2.1 Consistency

By the law of large numbers (LLN), for any fixed $\theta \in \Theta$, the sequence of random variables $M_n(\theta)$ converges in probability to its deterministic counterpart $M(\theta) := E M_n(\theta)$. Therefore it is reasonable to expect that the sequence $\hat{\theta}_n$ converges in probability to $\theta_0 := \arg \max_{\theta \in \Theta} M(\theta)$, as $n \to \infty$. The next theorem [Van der Vaart (1998), Theorem 5.7] establishes this result rigorously.

**Theorem 2.1** Let $M_n(\theta)$ be given in (2.1), $M(\theta) = EM_n(\theta)$, and let $\hat{\theta}_n$ be defined by (2.2).
Assume that for every $\epsilon > 0$,
\[
\sup_{\theta : \|\theta - \theta_0\| \geq \epsilon} M(\theta) < M(\theta_0), \quad (2.3)
\]
\[
\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{p} 0. \quad (2.4)
\]
Then $\hat{\theta}_n \xrightarrow{p} \theta_0$, $n \to \infty$.

According to Theorem 2.1, consistency of $\hat{\theta}_n$ is ensured by two conditions.

The deterministic condition (2.3) requires that $\theta_0$ is a well separated point of maximum of function $M(\theta)$. The stochastic condition (2.4) requires uniform (over $\theta \in \Theta$) convergence of $M_n(\theta)$ to $M(\theta)$. Recall that pointwise convergence of $M_n(\theta)$ to $M(\theta)$ holds by the LLN. Thus, (2.4) is a uniform version of the LLN.

There is vast literature dealing with uniform LLN [Pollard (1984), Van der Vaart and Wellner (1996), Van der Vaart (1998)]. It turns out that the fulfillment of (2.4) depends on the complexity of the class of functions $\{m_{\theta}, \theta \in \Theta\}$. There are different ways to measure complexity of classes of functions. Here we review relevant notions and concepts that will be used in Chapter 3.

For our purpose we restrict ourselves to functions $\{m_{\theta} = 1_{D_{\theta}}, \theta \in \Theta\}$, where $D_{\theta} \in \mathcal{D} \subseteq \mathbb{R}^d$. Important classes of functions for which uniform convergence of type (2.4) holds are the so-called VC-classes.

### 2.1.1 VC-classes (Vapnik-Cervonenkis-classes)

The following definition of VC-class can be found in Pollard (1984). Consider an arbitrary collection $(z_1, ..., z_n)$ of points in $\mathcal{Z}$ and a collection $\mathcal{D}$ of subsets of $\mathbb{R}^d$. We say that $\mathcal{D}$ picks out a certain subset $A$ of $(z_1, ..., z_n)$ if $A = D \cap (z_1, ..., z_n)$ for some $D \in \mathcal{D}$. We say that $\mathcal{D}$ shatters $(z_1, ..., z_n)$ if all of the $2^n$ possible subsets of $(z_1, ..., z_n)$ are picked out by the sets in $\mathcal{D}$. The VC-index $V(\mathcal{D})$ of the class $\mathcal{D}$ is the smallest $n$ for which no set of size $n$, $(z_1, ..., z_n) \subset \mathbb{R}^d$ is shattered by $\mathcal{D}$. If $\mathcal{D}$ shatters all sets $(z_1, ..., z_n)$ for all $n \geq 1$, we set $V(\mathcal{D}) = \infty$. Clearly, the more rich $\mathcal{D}$ is, the higher the VC-index. We say that $\mathcal{D}$ is a VC-class if $V(\mathcal{D}) < \infty$. Alternatively, VC-class is said to have polynomial discrimination. We illustrate the above definition in the following example.

**Example 1** Let $(z_1, ..., z_n) \subset \mathbb{R}^1$, and consider the class of sets $\mathcal{D} = \{D_t, t \in \mathbb{R}^1\}$, where $D_t = \{z : -\infty < z \leq t\}$. In that case $V(\mathcal{D}) = 2$. To see that, note that every one-point
set \{z_1\} is shattered, but no two-point set \{z_1, z_2\} is shattered. If \(z_1 < z_2\), then the sets \(D_t\) cannot pick out \{z_2\}.

The notion of VC-classes of functions is defined as follows. Let \(Z\) be the sample space, and define the subgraph of a function \(m : Z \to \mathbb{R}^1\) to be a subset of \(Z \times \mathbb{R}^1\) given by \\((z, u) : u < m(z)\). A collection \(\mathcal{M}\) of measurable functions on a sample space is called VC-class, if the collection of all subgraphs of the functions in \(\mathcal{M}\) forms a VC-class of sets (in \(Z \times \mathbb{R}^1\)).

**Example 2** (continuation) The set \(\mathcal{M}\) of all indicator functions \(\{m_t = 1_{D_t}, t \in \mathbb{R}^1\}\), where \(D_t \in \mathcal{D} \in \mathbb{R}^1\), is a VC-class of functions. To see that, note that no collection of \\((z, u) : u < 0\)\ can be shattered by \(\mathcal{M}\), and clearly no collection of \\((z, u) : u \geq 1\)\ can be shattered by \(\mathcal{M}\). In fact, only one-point sets can be shattered by \(\mathcal{M}\), thus, \(V(\mathcal{M}) = V(\mathcal{D}) = 2\).

In particular, it can be shown that any set of indicator functions of VC-classes of sets is a VC-class. In Chapter 3 we verify conditions of Theorem 2.1 for the HT estimator defined in (1.4). Observe that the HT estimator is of the form (2.2) with \(\{m_\theta = 1_{D_\theta}, \theta = (\phi, t) \in [0, 2\pi) \times \mathbb{R}^1_+\}\), where \(D_\theta\) [given in (1.5)] is the set of all points of the \((X, Y)\)-plane lying between two lines with parameters \((\phi, t + \frac{r}{2})\) and \((\phi, t - \frac{r}{2})\). In order to verify (2.4) we will show that \(\mathcal{D} = \{D_\theta, \theta = (\phi, t) \in [0, 2\pi) \times \mathbb{R}^1_+\}\) is a VC-class of sets.

### 2.2 Rate of convergence and limit distribution

As already mentioned in Chapter 1, the rate of convergence of the HT estimator is different from the standard \(\sqrt{n}\)-rate, and it is of the cube-root type. In this section we discuss this phenomenon, and present general results on the limit distribution of M-estimators.

#### 2.2.1 Rate of convergence

Here we follow the discussion in Kim and Pollard (1990). Among other problems, they derive general results on cube-root asymptotics and, as an example, they consider the problem of estimating the mode of a symmetric, unimodal, one-dimensional density. Suppose that \(Z_i\) are i.i.d. random variables on the real line. Let \(M_n(\theta)\) be given in (2.1)
with \( m_\theta(Z_i) = 1\{\theta - 1 \leq Z_i \leq \theta + 1\} \). Thus, the criterion function is the proportion of observations in an interval of length 2 centered at \( \theta \), and \( \hat{\theta}_n \) given in (2.2) is the estimator of the mode, \( \theta_0 \). We describe informally the derivation of the rate of convergence of \( \hat{\theta}_n \).

Generally speaking, the rate of convergence of \( \hat{\theta}_n \) depends on the combined behavior of the criterion function and its deterministic counterpart. Let \( \hat{\theta}_n = \arg \max_{\theta \in \Theta} (M_n(\theta) - M_n(\theta_0)) \), \( \theta_0 = \arg \max_{\theta \in \Theta} (M(\theta) - M(\theta_0)) \), and suppose that \( \hat{\theta}_n \) converges in probability to \( \theta_0 \). Here \( M(\theta) = \mathbb{E}M_n(\theta) = \mathbb{P}(\theta - 1 \leq Z \leq \theta + 1) \), where \( \mathbb{P} \) is the common distribution of \( Z_i \). If \( \mathbb{P} \) has a smooth, unimodal density \( f \), then \( M(\theta) \) is approximately parabolic in a neighborhood of its maximizing value \( \theta_0 \). Thus, the deterministic trend \( M(\theta) - M(\theta_0) \) can be approximated by \(-c_1(\theta - \theta_0)^2\) in a neighborhood of \( \theta_0 \), where \( c_1 \) is some positive constant. Define
\[
L_n(\theta) = [M_n(\theta) - M_n(\theta_0)] - [M(\theta) - M(\theta_0)].
\]

One can argue that for fixed \( \theta \), in a neighborhood of \( \theta_0 \), \( L_n(\theta) \) is approximately \( N(0, \sigma_n^2/n) \), where \( \sigma_n^2 \approx c_2|\theta - \theta_0| \), and \( c_2 \) is a constant. To this end we first note that
\[
M_n(\theta) - M_n(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} 1(\theta - 1 \leq Z_i \leq \theta + 1) - \frac{1}{n} \sum_{i=1}^{n} 1(\theta_0 - 1 \leq Z_i \leq \theta_0 + 1)
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} 1(\theta - 1 \leq Z_i \leq \theta_0 - 1) - \frac{1}{n} \sum_{i=1}^{n} 1(\theta + 1 \leq Z_i \leq \theta_0 + 1).
\]

The first sum on the RHS is a Binomial random variable with parameters \( n \), and \( p = \int_{\theta-1}^{\theta_0-1} f(z)dz \), while the second sum on the RHS is Binomial with \( n \), and \( p = \int_{\theta_0+1}^{\theta+1} f(z)dz \).

Thus, the variance of the expression on the RHS is approximately \( \int_{\theta-1}^{\theta_0-1} f(z)dz + \int_{\theta_0+1}^{\theta+1} f(z)dz \approx c_2|\theta - \theta_0| \), because \( 1 - \int_{\theta-1}^{\theta_0-1} f(z)dz \approx 1 - \int_{\theta_0+1}^{\theta+1} f(z)dz \approx 1 \). Thus informally, we may write that
\[
M_n(\theta) - M_n(\theta_0) \approx c_2n^{-1/2}|\theta - \theta_0|^{1/2} - c_1(\theta - \theta_0)^2.
\]

Since \( M_n(\theta) \) is maximized over \( \theta \), it is greater than \( M_n(\theta_0) \). Therefore, \( n^{-1/2}|\theta - \theta_0|^{1/2} \) should be greater than \( (\theta - \theta_0)^2 \). Thus, the maximum is likely to occur in the range where \( |\theta - \theta_0| \) is of order \( n^{-1/3} \) or smaller. This explains the \( n^{1/3} \)-rate of convergence.

Note that if the variance of the noise, \( \sigma_n^2 \), is of the order of \( |\theta - \theta_0|^2 \), \( |\theta - \theta_0| \to 0 \), then the maximizing value is likely to occur in the range where \( (\theta - \theta_0)^2 \) is of the same order as, or smaller than \( n^{-1/2}|\theta - \theta_0| \). This would yield the standard \( \sqrt{n} \)-rate of convergence.

The following theorem [Van der Vaart and Wellner (1996), Corollary 3.2.6] makes this heuristic argument rigorous.
Theorem 2.2 Let $M_n(\theta)$ be given in (2.1), $M(\theta) = \mathbb{E}M_n(\theta)$, and $\hat{\theta}_n$ be defined by (2.2). Suppose that $\theta_0 = \arg\max_{\theta \in \Theta} M(\theta)$, and assume that for every $\theta$ in a neighborhood of $\theta_0$,

$$M(\theta) - M(\theta_0) \leq -c_1\|\theta - \theta_0\|^2,$$

for some positive constant $c_1$. Furthermore, assume that there exists a function $\phi$ such that $\delta \mapsto \phi(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$. Let $R_n$ be a sequence satisfying $R_n^2\phi(1/R_n) \leq \sqrt{n}$ for all $n$.

If $\hat{\theta}_n \overset{p}{\to} \theta_0$, and if for some constant $c_2 > 0$, and for all $n$

$$\mathbb{E}\sqrt{n} \sup_{\|\theta - \theta_0\| < \delta} |[M_n(\theta) - M_n(\theta_0)] - [M(\theta) - M(\theta_0)]| \leq c_2\phi(\delta),$$

then $R_n\|\hat{\theta}_n - \theta_0\| = O_p(1), \ n \to \infty$.

The theorem establishes an upper bound on the rate of convergence $R_n$. Similarly to the theorem of consistency of an M-estimator, this theorem includes two conditions, deterministic and stochastic.

The deterministic condition (2.5) ensures a parabolic behavior of $M(\theta)$ in a neighborhood of $\theta_0$. The stochastic condition (2.6) is similar in spirit to (2.4). This condition can be verified by showing that the classes of functions

$$\mathcal{M}_\delta = \{m_\theta - m_{\theta_0} : \|\theta - \theta_0\| < \delta\}, \ \delta > 0,$$

are VC-classes. It turns out that the LHS of (2.6) can be bounded in terms of the second moment of the so-called envelope function

$$B_\delta = \sup_{\|\theta - \theta_0\| < \delta} |m_\theta - m_{\theta_0}|.$$

[see Van der Vaart and Wellner (1996), page 292].

2.2.2 Limit distribution

The estimators $\hat{\theta}_n$ converge to some constant $\theta_0$, thus it is necessary to rescale them before studying distributional limit properties. If $\hat{\theta}_n$ maximizes the function $\theta \mapsto M_n(\theta)$, then the rescaled estimators $\hat{h}_n := R_n(\hat{\theta}_n - \theta)$ are maximizers of the local criterion functions $h \mapsto M_n(\theta + \frac{h}{R_n}) - M_n(\theta_0)$. Thus, if these local criterion functions converge to a limit
process $h \mapsto M(h)$, we expect that the sequence $\hat{h}_n$ converges in distribution to $\hat{h}$, the maximizer of this limit process.

The next theorem [Van der Vaart and Wellner (1996), Theorem 3.2.10] will be used in Chapter 3 for developing the limit distribution of the HT estimator. We formulate this theorem in the form suitable for our purposes.

**Theorem 2.3** Let an M-estimator be defined by (2.1)-(2.2). For each $\theta \in \Theta$, let $m_{\theta} : Z \rightarrow \mathbb{R}$. Suppose $\theta \mapsto M(\theta)$ is twice continuously differentiable at a point of maximum $\theta_0$, with nonsingular second-derivative matrix $V$.

Let the class $\mathcal{M}_\delta$ defined in (2.7) be a VC-class with envelope function (2.8). Assume that for some continuous function $\phi$, $E B^2_\delta \leq \phi^2(\delta)$, and $\delta \mapsto \phi(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$.

Assume also that for every $\zeta > 0$,

$$
\lim_{\delta \downarrow 0} \frac{\mathbb{E} B^2_\delta 1\{B_\delta > \zeta \delta^{-2}\phi^2(\delta)\}}{\phi^2(\delta)} = 0,
$$

(2.9)

and for all $K$

$$
\lim_{\xi \downarrow 0} \lim_{\delta \downarrow 0} \sup_{\|h-g\|<\xi \atop \|h\|,\|g\| \leq K} \frac{\mathbb{E}(m_{\theta_0+\delta g} - m_{\theta_0+\delta h})^2}{\phi^2(\delta)} = 0.
$$

(2.10)

Let $\hat{\theta}_n$ be given by (2.2) and assume that $\hat{\theta}_n \overset{p}{\rightarrow} \theta_0$. Then

$$
R_n(\hat{\theta}_n - \theta_0) \overset{d}{\rightarrow} \hat{h}, \ n \rightarrow \infty,
$$

where $R_n$ is the solution to the equation $R_n^2 \phi(1/R_n) = \sqrt{n}$, $\hat{h}$ is the (almost surely) unique maximizer of the process $h \mapsto G(h) + \frac{1}{2} h V h$, and $G$ is some zero-mean Gaussian process with continuous sample paths and

$$
\mathbb{E}(G(g) - G(h))^2 = \lim_{\delta \downarrow 0} \frac{\mathbb{E}(m_{\theta_0+\delta g} - m_{\theta_0+\delta h})^2}{\phi^2(\delta)},
$$

(2.11)

for all $g, h \in \Theta$. 

21
Chapter 3

Statistical properties of the HT estimator

In this chapter we investigate the statistical properties of the Hough transform estimator in the polar parameterization of a line. We consider the error-in-variables observation model.

3.1 Observation model

Let $(\tilde{X}_i, \tilde{Y}_i), i = 1, ..., n$ be i.i.d. random variables lying on the unknown line with parameters $(\phi_0, t_0)$, i.e.

$$\tilde{X}_i \cos \phi_0 + \tilde{Y}_i \sin \phi_0 = t_0. \quad (3.1)$$

Assume that we are given observations $(X_i, Y_i), i = 1, ..., n$ such that

$$X_i = \tilde{X}_i + \gamma_i,$$
$$Y_i = \tilde{Y}_i + \eta_i, \quad (3.2)$$

where $\gamma_i$ and $\eta_i$ are noise variables independent of $(\tilde{X}_i, \tilde{Y}_i)$. This assumption corresponds to adding noise in the horizontal direction, and in the vertical direction. The objective is to estimate $\theta_0 = (\phi_0, t_0)$ from the observed data.

It is well known that $\theta_0$ is not identifiable unless some conditions are satisfied. In particular, in what follows we assume that the joint distribution of $(\tilde{X}_i, \tilde{Y}_i)$ is non-Gaussian. This assumption is sufficient for identifiability [Cheng and Van Ness (1999)].
In the sequel we use the following matrix notation: 
\[ \tilde{Z} = (\bar{X}, \bar{Y})^T, \quad Z = (X, Y)^T, \quad \theta = (\phi, t)^T, \quad \lambda = (\gamma, \eta)^T, \quad U_\phi = (\cos \phi, \sin \phi)^T. \]

Thus (3.1) takes the form
\[ \tilde{Z}_i^T U_\phi = t_0, \quad i = 1, \ldots, n, \]
and we observe the data \( Z_n := \{Z_1, \ldots, Z_n\}. \)

Here we present the following assumption on the noise variables \( \lambda \).

**Assumption 1**

(a) \( \lambda_i = (\gamma_i, \eta_i), i = 1, \ldots, n \) are i.i.d. random vectors independent of \((\bar{X}_i, \bar{Y}_i), i = 1, \ldots, n\).

(b) for any \( \phi \in [0, 2\pi) \), the random variable \( \lambda^T U_\phi \) has density \( f \) which is symmetric and strictly unimodal. By strict unimodality we mean that \( f(x) \) has a maximum at a unique point, \( x = 0 \), and decreases in either directions as \( x \) decreases or increases away from zero.

Note that if \( \gamma_i \) and \( \eta_i \) are independent, and have a Gaussian density \( f \), then the distribution of \( \lambda^T U_\phi \) is \( f \) as well. We call this property rotation invariant. Moreover, Assumption 1 does not require existence of the expectation of the noise variables. This means that \( f \) can be a density with "heavy tails" such as the Cauchy density.

The HT estimator is defined as follows. Let \( \theta = (\phi, t) \in \Theta = [0, 2\pi) \times \mathbb{R}_+^1 \), and define
\[ m_\theta(Z_i) = 1\{|Z_i^T U_\phi - t| \leq \frac{r}{2}\}. \quad (3.3) \]

Denote the criterion function by
\[ M_{r,n}(\theta) = \frac{1}{n} \sum_{i=1}^{n} m_\theta(Z_i), \quad (3.4) \]

thus, the HT estimator is given by
\[ \hat{\theta}_{r,n} = \arg \max_{\theta \in \Theta} M_{r,n}(\theta). \quad (3.5) \]

**3.1.1 Asymptotic results**

We start by presenting the results on consistency of the HT estimator.

**Theorem 3.1** Let \((\bar{X}_i, \bar{Y}_i), i = 1, \ldots, n\) be given by (3.1)-(3.2), and the HT estimator \( \hat{\theta}_{r,n} \) be defined in (3.3)-(3.5). If Assumption 1 holds, then for any fixed \( r > 0 \),
\[ \hat{\theta}_{r,n} \xrightarrow{p} \theta_0, \quad \text{as } n \to \infty. \]
The next theorem establishes the rate of convergence of the HT estimator. In order to obtain this result we assume that

**Assumption 2** \( \lambda_i \sim N_2(0, \sigma^2 I), i = 1, ..., n \) and \( \mathbb{E} \tilde{X}^2, \mathbb{E} \tilde{Y}^2 < \infty \).

**Theorem 3.2** Let \((\tilde{X}_i, \tilde{Y}_i), i = 1, ..., n\) be given by (3.1)-(3.2), and the HT estimator \( \hat{\theta}_{r,n} \) be defined in (3.3)-(3.5). Suppose that Assumption 2 holds. Then for any fixed \( r > 0 \),

\[
n^{1/3} \| \hat{\theta}_{r,n} - \theta_0 \| = O_p(1), \quad n \to \infty.
\]

The above theorem suggests that the rate of convergence of the HT estimator, which we denote by \( R_n \), is of order \( n^{1/3} \). This is a slow rate of convergence which makes the HT procedure statistically inefficient in comparison with the least squares estimator. However, in the sequel we show that the HT has very good robustness properties. Thus, it is reasonable to consider this technique more like a data analysis tool. For example, we can apply it to data in order to distinguish outliers, and then perform standard regression analysis.

The next theorem establishes asymptotic distribution of the HT estimator.

**Theorem 3.3** Let \((\tilde{X}_i, \tilde{Y}_i), i = 1, ..., n\) be given by (3.1)-(3.2), and the HT estimator \( \hat{\theta}_{r,n} \) be defined in (3.3)-(3.5). If Assumption 2 holds, then for every fixed \( r > 0 \),

\[
n^{1/3} (\hat{\theta}_{r,n} - \theta_0) \xrightarrow{d} W,
\]

where \( W \) has the distribution of the maximizer of the process \( \theta \mapsto G(\theta) + \frac{1}{2} \theta^T V_0 \theta \). Here

\[
V_0 = [f'(\frac{r}{2}) - f'(-\frac{r}{2})] \mathbb{E} A(\tilde{Z}) A^T(\tilde{Z}),
\]

\[
A(\tilde{Z}) = (\tilde{Z}^T U_{\phi_0 - \pi/2}, 1)^T,
\]

and \( G \) is a zero-mean Gaussian process with continuous sample paths and stationary increments \( g, h \) defined on \( \Theta \), where

\[
\mathbb{E}[G(g) - G(h)]^2 = 2 f \left( \frac{r}{2} \right) \mathbb{E}[(h - g)^T A(\tilde{Z})].
\]

The limit distribution given above is quite complicated and depends on the unknown parameter \( \phi_0 \). Note that the limit distribution of the HT estimator in the Cartesian parameterization studied in Goldenshluger and Zeevi (2004) does not depend on the unknown parameter \( \theta_0 = (b_0, a_0) \). However, in the Cartesian case, the straight lines in the Hough domain corresponding to the observations with large \( X_i \) values are very steep; i.e., if most
of the observations have large $X$ -- coordinates and the standard deviation of the noise is small, then the corresponding straight lines are nearly parallel. In this case behavior of the HT estimator may be very poor. This follows because the matrix $V_0$ given in (1.11) is nearly singular for large $X_i$, since $f'_r$ is a function of $Z = (X, 1)^T$. Thus, the asymptotic distribution of $\hat{\theta}_{r,n}$ is close to the distribution of the point of maximum of a zero mean Gaussian process given by (1.12). In our case $f'$ is a function of $r$ only, thus, for any reasonable value of $r$, $V_0$ given in (3.6) is non-singular.

3.1.2 Equivariance properties

When studying the equivariance properties of an estimator there are several types to consider, such as regression, scale, and affine equivariance. For instance, scale equivariance implies that the fit is essentially independent of the choice of measurements unit for the response variable $Y$. Affine equivariance means that a linear transformation of the $X_i$ should transform the estimator accordingly. This allows us to use another coordinate system for the explanatory variables, without affecting the estimated $\hat{Y}_i$.

The version of the HT considered in Goldenshluger and Zeevi (2004) is regression equivariant, but not scale and affine equivariant. However, in their model the HT treats the variables asymmetrically, designating a dependent variable and one (or more) independent variables. Our case is like orthogonal regression, which treats the variables symmetrically. It turns out that the HT in the polar parameterization is rotation equivariant. Let $R_\alpha$ be a rotation matrix with angle $\alpha$, then we can define the transformed observations by $Z^* = R_\alpha Z$. The HT estimator based on $Z^*$ will be $\hat{t}(Z^*) = \hat{t}(Z)$, and $\hat{\phi}(Z^*) + \alpha = \hat{\phi}(Z)$. It is easy to verify this argument by noting that $1\{|(R_\alpha Z)^T U_{\phi+\alpha} - t| \leq \frac{r}{2}\} = 1\{|Z^T U_\phi - t| \leq \frac{r}{2}\}$.

3.1.3 Robustness

The following theorem establishes the breakdown point (BP) of the HT estimator.

**Theorem 3.4** Let $Z_n = \{Z_1, ..., Z_n\}$ be a sample belongs to a ball of finite radius. Then

$$\epsilon_{\text{rep}}(\hat{\theta}_{r,n}; Z_n) = \frac{1}{n} \left[ \frac{n M_{r,n}(\hat{\theta}_{r,n})}{2} \right],$$

$$\epsilon_{\text{add}}(\hat{\theta}_{r,n}; Z_n) = \frac{n M_{r,n}(\hat{\theta}_{r,n}) + 1}{n + n M_{r,n}(\hat{\theta}_{r,n}) + 1}.$$
Moreover, under Assumption 1, we have
\[ \epsilon\text{rep}(\hat{\theta}; Z_n) \xrightarrow{a.s.} p/2, \quad \epsilon\text{add}(\hat{\theta}; Z_n) \xrightarrow{a.s.} p(1 + p)^{-1}, \quad n \to \infty, \]
where \( p = \int_{-r/2}^{r/2} f(x)dx. \)

We now turn to several remarks concerning the theorem. First, the BP of the HT estimator in the Cartesian parameterization depends on the assumption that there are no repeated values of the \( X_i \) [Goldenshluger and Zeevi (2004)] which can be relaxed here due to the geometrical properties of the Hough domain in the polar case. Second, the value of \( r \) controls breakdown properties of the HT estimator: the larger \( r \), the closer the breakdown point is to 1/2. An illustration for the robustness of the HT is presented in the next chapter.

3.2 Proofs

In this section we present proofs of Theorems 3.1-3.4.

3.2.1 Proof of Theorem 3.1

We prove the theorem by verifying the conditions of Theorem 2.1.

In order to verify condition (2.3) we note that when conditioning on \( \tilde{Z} \), we have for \( \theta \neq \theta_0 \),
\[ M_r(\theta) = \mathbb{E}[M_{r,n}(\theta)] \]
\[ = \mathbb{E}\left[ \frac{1}{2} \right] \left( |\tilde{Z}_i^T U_\phi + \lambda_i^T U_\phi - t| \leq \frac{r}{2} \right) \]
\[ = \mathbb{E}\left[ \mathbb{E}\left[ \frac{1}{2} \right] \left( |\tilde{Z}_i^T U_\phi + \lambda_i^T U_\phi - t| \leq \frac{r}{2} \right) | \tilde{Z} \right] \]
\[ = \mathbb{E}\left[ \frac{1}{2} \right] \left( |\tilde{Z}_i^T U_\phi + \lambda_i^T U_\phi - t| \leq \frac{r}{2} | \tilde{Z} \right) \].

Under Assumption 1 the density of \( \lambda^T U_\phi \) is \( f \), thus unimodal. Combining this result with Theorem 1 in Anderson (1955), it follows that for any \( \epsilon > 0 \) and \( \|\theta - \theta_0\| > \epsilon \)
\[ \mathbb{E}\left[ \frac{1}{2} \right] \left( -\frac{r}{2} \leq \tilde{Z}_i^T U_\phi + \lambda_i^T U_\phi - t \right) \left( \frac{r}{2} \leq \tilde{Z}_i^T U_\phi + \lambda_i^T U_\phi - t | \tilde{Z} \right) \]
\[ < \mathbb{P}\left( -\frac{r}{2} \leq \lambda^T U_\phi \leq \frac{r}{2} \right) = M_r(\theta_0). \]

Here we used the fact that \( \tilde{Z}_i^T U_{\phi_0} = t_0 \). Thus, \( \theta_0 \) is a unique point of maximum of the function \( M_r(\theta) \) for any \( r > 0 \), and condition (2.3) holds.
In order to check that (2.4) is valid, we first show that the class of sets \( D = \{ D_\theta, \theta = (\phi, t) \in \Theta \} \), where \( D_\theta \) is defined in (1.5), is a VC class. The set \( D_\theta \) consists of all points of the plane lying between two lines with parameters \((\phi, t + \frac{r}{2})\) and \((\phi, t - \frac{r}{2})\). This set is an intersection of two half planes in the \((x, y)\)-plane. There is no set of size 4 that can be shattered by a half plane, thus, a half plane is a VC-class with index 4. By Lemma 15, Chapter 2 in Pollard (1984), the intersection of two VC-classes is a VC-class. Therefore, \( D \) is a VC-class of sets. Moreover, the set of functions \( \{ m_\theta = 1_{D_\theta}, \theta = (\phi, t) \in \Theta \} \) is a VC-class of functions. Thus, by Theorem 14, Chapter 2 in Pollard (1984) it follows that

\[
\sup_{\theta \in \Theta} |M_{r,n}(\theta) - M_r(\theta)| = \sup_{D \in \mathcal{D}} |\mathbb{P}_n(D) - \mathbb{P}(D)| \xrightarrow{a.s.} 0, \ n \to \infty.
\]

Thus, condition (2.4) holds and we conclude that the HT estimator is consistent.

### 3.2.2 Proof of Theorem 3.2

In order to derive the rate of convergence of the HT estimator we will verify the conditions in Theorem 2.2. First we check condition (2.5).

Let \( V(\theta) \) denote the second derivative matrix of the function \( M_r(\theta) \) given in (3.8). Thus write

\[
M_r(\theta) = \mathbb{E}\left\{ F\left[ \frac{r}{2} - \tilde{Z}^T U_\phi + t \right] - F\left[ -\frac{r}{2} - \tilde{Z}^T U_\phi + t \right] \right\},
\]

where the expected value is taken w.r.t. the distribution of \( \tilde{Z} \), and \( F \) is the distribution function of \( \eta \), thus Gaussian. Now, the normal distribution is continuously differentiable with bounded derivative, and under Assumption 2, \( \mathbb{E}X^2 \) and \( \mathbb{E}Y^2 < \infty \). When differentiating \( M_r(\theta) \) w.r.t \( \theta \) we can apply the dominant convergence theorem (DCT) to interchange the order of expectation and differentiation for the expression on the RHS of (3.9). Here we note that differentiation of the expression \(-\tilde{Z}^T U_\phi \) yields \( \tilde{Z}^T U_{\phi - \pi/2} \). In particular, differentiating (3.9) twice w.r.t. \( \theta \) under the integral sign, and setting \( \theta = \theta_0 \) yields

\[
V_0 := \nabla^2_\theta M_r(\theta) |_{\theta = \theta_0} = [f'(\frac{r}{2}) - f'(-\frac{r}{2})] \mathbb{E}A(\tilde{Z}) A^T(\tilde{Z}),
\]

where \( A(\tilde{Z}) = (\tilde{Z}^T U_{\phi_0 - \pi/2}, 1)^T \). The matrix \( V_0 \) is negative definite. This follows because \( f'(x) - f'(-x) < 0 \), for all \( x > 0 \), and the determinant of the matrix \( \mathbb{E}A(\tilde{Z}) A^T(\tilde{Z}) \) is

\[
\mathbb{E}\left( \tilde{Z}^T U_{\phi_0 - \pi/2} U_{\phi_0 - \pi/2}^T \tilde{Z} \right) - \left( \mathbb{E}\tilde{Z} U_{\phi_0 - \pi/2} \right)^2
\]

which is the variance of the random variable \( \tilde{Z}^T U_{\phi_0 - \pi/2} \), thus \( \mathbb{E}A(\tilde{Z}) A^T(\tilde{Z}) \) is positive definite. Therefore, we conclude that (2.5) holds.
Now we verify condition (2.6). Using Lemma 28, Chapter 2 in Pollard (1984) we have that $\mathcal{M}_\delta$ defined in (2.7) is a VC-class of functions. The envelope function of this class, given in (2.8) is bounded as follows.

$$
B_\delta = \sup_{\|\theta-\theta_0\|<\delta} \left| 1 \left\{ |Z^TU_\phi - t| \leq \frac{r}{2} \right\} - 1 \left\{ |Z^TU_{\phi_0} - t_0| \leq \frac{r}{2} \right\} \right|
$$

$$
= \sup_{\|\theta-\theta_0\|<\delta} \left| 1 \left\{ |Z^TU_\phi + \lambda^T U_\phi - t + t_0 - \lambda^T U_{\phi_0} - \lambda^T U_{\phi_0} - \lambda^T U_{\phi_0}| \leq \frac{r}{2} \right\} \right|
$$

$$
-1 \left\{ |Z^TU_{\phi_0} + \lambda^T U_{\phi_0} - t_0| \leq \frac{r}{2} \right\} \right|
$$

$$
= \sup_{\|\theta-\theta_0\|<\delta} \left| 1 \left\{ |(\tilde{Z} + \lambda)^T (U_\phi - U_{\phi_0}) - (t - t_0) + \lambda^T U_{\phi_0}| \leq \frac{r}{2} \right\} \right|
$$

$$
-1 \left\{ |\lambda^T U_{\phi_0}| \leq \frac{r}{2} \right\} \right|
$$

$$
= \sup_{\|\theta-\theta_0\|<\delta} \left| 1 \left\{ \frac{-r}{2} - Q(\tilde{Z}) \leq \lambda^T U_{\phi_0} \leq \frac{-r}{2} - Q(\tilde{Z}) \right\} - \left\{ \frac{-r}{2} \leq \lambda^T U_{\phi_0} \leq \frac{-r}{2} \right\} \right|
$$

where we define

$$
Q(\tilde{Z}) = (\tilde{Z} + \lambda)^T (U_\phi - U_{\phi_0}) - (t - t_0).
$$

Because for any $A$ and $B$, $|1_A - 1_B| = 1_{A\setminus B} + 1_{B\setminus A}$, we have for $\delta$ small enough

$$
B_\delta = \sup_{\|\theta-\theta_0\|<\delta} \left[ 1 \left\{ \min \left( \frac{-r}{2}, \frac{r}{2} - Q(\tilde{Z}) \right) \leq \lambda^T U_{\phi_0} \leq \max \left( \frac{-r}{2}, \frac{r}{2} - Q(\tilde{Z}) \right) \right\} \right]
$$

$$
+1 \left\{ \min \left( \frac{r}{2}, \frac{r}{2} - Q(\tilde{Z}) \right) \leq \lambda^T U_{\phi_0} \leq \max \left( \frac{r}{2}, \frac{r}{2} - Q(\tilde{Z}) \right) \right\} \right]
$$

$$
\leq \sup_{\|\theta-\theta_0\|<\delta} \left[ 1 \left\{ \frac{-r}{2} - |Q(\tilde{Z})| \leq \lambda^T U_{\phi_0} \leq \frac{r}{2} + |Q(\tilde{Z})| \right\} \right]
$$

$$
+1 \left\{ \frac{r}{2} - |Q(\tilde{Z})| \leq \lambda^T U_{\phi_0} \leq \frac{r}{2} + |Q(\tilde{Z})| \right\} \right]
$$

(3.10)

Observe that by Schwarz’s inequality

$$
|Q(\tilde{Z})| = |(\tilde{Z} + \lambda)^T (U_\phi - U_{\phi_0}) - (t - t_0)|
$$

$$
\leq |(\tilde{Z} + \lambda)^T (U_\phi - U_{\phi_0})| + |t - t_0|
$$

$$
\leq (\|\tilde{Z}\| + \|\lambda\|)\|U_\phi - U_{\phi_0}\| + |t - t_0|.
$$

Because $\|U_\phi - U_{\phi_0}\| = 2\sin \frac{\phi - \phi_0}{2}$, it follows that for $\delta$ small enough

$$
\sup_{\|\theta-\theta_0\|<\delta} \left( (\|\tilde{Z}\| + \|\lambda\|)\|U_\phi - U_{\phi_0}\| + |t - t_0| \right) \leq 2\sin \frac{\delta}{2}(\|\tilde{Z}\| + \|\lambda\|) + \delta
$$

$$
\leq 2\frac{\delta}{2}(\|\tilde{Z}\| + \|\lambda\|) + \delta
$$

$$
= \delta(\|\tilde{Z}\| + \|\lambda\| + 1).
$$
Combining this with (3.10) we obtain

\[
B_\delta \leq 1 \left\{ \frac{-r}{2} - \delta(\|\hat{Z}\| + \|\lambda\| + 1) \leq \lambda^T U_{\phi_0} \leq \frac{-r}{2} + \delta(\|\hat{Z}\| + \|\lambda\| + 1) \right\}
\]

\[+ 1 \left\{ \frac{r}{2} - \delta(\|\hat{Z}\| + \|\lambda\| + 1) \leq \lambda^T U_{\phi_0} \leq \frac{r}{2} + \delta(\|\hat{Z}\| + \|\lambda\| + 1) \right\}
\]

\[\leq 1 \left\{ \frac{-r}{2} - \delta(\|\hat{Z}\| + |\gamma| + |\eta| + 1) \leq \lambda^T U_{\phi_0} \leq \frac{-r}{2} + \delta(\|\hat{Z}\| + |\gamma| + |\eta| + 1) \right\}
\]

\[+ 1 \left\{ \frac{r}{2} - \delta(\|\hat{Z}\| + |\gamma| + |\eta| + 1) \leq \lambda^T U_{\phi_0} \leq \frac{r}{2} + \delta(\|\hat{Z}\| + |\gamma| + |\eta| + 1) \right\}.
\]

The last inequality follows because \(\|\lambda\| = \sqrt{\gamma^2 + \eta^2} \leq |\gamma| + |\eta|\). Recall that the distribution of \(\lambda^T U_{\phi_0}\) is the same as the distribution of \(\eta\) which is normal. Thus, conditioning on \(\hat{Z}\) and \(\gamma\), and taking expectation of \(\lambda^T U_{\phi_0}\) we have

\[
\mathbb{E}B_\delta^2 \leq \mathbb{E}\left\{ \mathbb{P}\left( \left. \frac{-r}{2} - \delta(\|\hat{Z}\| + |\gamma| + |\eta| + 1) \leq \eta \leq \frac{-r}{2} + \delta(\|\hat{Z}\| + |\gamma| + |\eta| + 1) \right| \hat{Z}, \gamma \right) \right\}
\]

\[+ \mathbb{E}\left\{ \mathbb{P}\left( \left. \frac{r}{2} - \delta(\|\hat{Z}\| + |\gamma| + |\eta| + 1) \leq \eta \leq \frac{r}{2} + \delta(\|\hat{Z}\| + |\gamma| + |\eta| + 1) \right| \hat{Z}, \gamma \right) \right\}
\]

\[= \mathbb{E}P_1 + \mathbb{E}P_2.
\]

For brevity we define \(b(\hat{Z}, \gamma) = \|\hat{Z}\| + |\gamma| + 1\). Thus write

\[
P_1 = \mathbb{P}\left( \left. \frac{-r}{2} - \delta b(\hat{Z}, \gamma) - \delta |\eta| \leq \eta \leq \frac{-r}{2} \right| \hat{Z}, \gamma \right)
\]

\[+ \mathbb{P}\left( \left. \frac{-r}{2} \leq \eta \leq \frac{-r}{2} + \delta b(\hat{Z}, \gamma) + \delta |\eta| \right| \hat{Z}, \gamma \right)
\]

\[= P_{11} + P_{12},
\]

and

\[
P_2 = \mathbb{P}\left( \left. \frac{r}{2} - \delta b(\hat{Z}, \gamma) - \delta |\eta| \leq \eta \leq \frac{r}{2} \right| \hat{Z}, \gamma \right)
\]

\[+ \mathbb{P}\left( \left. \frac{r}{2} \leq \eta \leq \frac{r}{2} + \delta b(\hat{Z}, \gamma) + \delta |\eta| \right| \hat{Z}, \gamma \right)
\]

\[= P_{21} + P_{22}.
\]

We bound the two terms on the RHS of (3.11) separately. Starting with \(P_{11}\), note that

\[
\{\eta + \delta |\eta| \geq \frac{-r}{2} - \delta b(\hat{Z}, \gamma)\} \subseteq \{\eta(1 + \delta) \geq \frac{-r}{2} - \delta b(\hat{Z}, \gamma)\} \cup \{\eta(1 - \delta) \geq \frac{-r}{2} - \delta b(\hat{Z}, \gamma)\}.
\]

Now, for \(\delta\) small enough (< 1/4), each of the events on the RHS is contained in

\[
\{\eta \geq \frac{-r}{2} - \delta b(\hat{Z}, \gamma)(1 + 2\delta)\},
\]

29
Again, by the Lipschitz property of the normal distribution, there exists a constant $L$ depending on $\sigma$ only such that the last expression is less than or equal to
\[
2L\left| -\frac{r}{2} - \left(\frac{-r}{2} - \delta b(\tilde{Z}, \gamma)\right)(1 + 2\delta) \right| = 2L\left(\delta b(\tilde{Z}, \gamma) + \delta r + 2\delta^2 b(\tilde{Z}, \gamma)\right) \leq 2L\delta\left(3b(\tilde{Z}, \gamma) + r\right).
\]
We apply the same argument in order to bound $\mathbb{P}_{12}$
\[
\{\eta - \delta|\eta| \leq -\frac{r}{2} + \delta b(\tilde{Z}, \gamma)\} \subseteq \{\eta(1 + \delta) \leq -\frac{r}{2} + \delta b(\tilde{Z}, \gamma)\} \cup \{\eta(1 - \delta) \leq -\frac{r}{2} + \delta b(\tilde{Z}, \gamma)\}.
\]
Now, we have two cases to consider. The first case is $-r/2 + \delta b(\tilde{Z}, \gamma) < 0$, then we use $(1 + \delta)^{-1} \geq 1 - \delta$ in order to bound the probability of the events above. The second case is $-r/2 + \delta b(\tilde{Z}, \gamma) > 0$ where we can bound the probability with $1 + 2\delta$. However, the result will be the same in both cases so we explore only the first.
\[
\mathbb{P}_{12} = \mathbb{P}\left( -\frac{r}{2} \leq \eta \leq -\frac{r}{2} + \delta b(\tilde{Z}, \gamma) + \delta|\eta| |\tilde{Z}, \gamma\right) \leq \mathbb{P}\left( -\frac{r}{2} \leq \eta \leq -\frac{r}{2} + \delta b(\tilde{Z}, \gamma)(1 - \delta) |\tilde{Z}, \gamma\right) + \mathbb{P}\left( -\frac{r}{2} \leq \eta \leq -\frac{r}{2} + \delta b(\tilde{Z}, \gamma)(1 + \delta) |\tilde{Z}, \gamma\right) \leq \left|\mathbb{P}\left( \eta \leq -\frac{r}{2} + \delta b(\tilde{Z}, \gamma)(1 - \delta) |\tilde{Z}, \gamma\right) - \mathbb{P}\left( \eta \leq -\frac{r}{2}\right)\right| + \left|\mathbb{P}\left( \eta \leq -\frac{r}{2} + \delta b(\tilde{Z}, \gamma)(1 + \delta) |\tilde{Z}, \gamma\right) - \mathbb{P}\left( \eta \leq -\frac{r}{2}\right)\right|.
\]
Again, by the Lipschitz property of the normal distribution there exists a constant $L$ depending only on $\sigma$ such that the last expression is less than or equal to
\[
L\left|\frac{-r}{2} + \delta b(\tilde{Z}, \gamma)(1 - \delta) + \frac{r}{2}\right| + L\left|\frac{-r}{2} + \delta b(\tilde{Z}, \gamma)(1 + \delta) + \frac{r}{2}\right| = L\left|\delta b(\tilde{Z}, \gamma) + \delta^r - \delta^2 b(\tilde{Z}, \gamma)\right| + L\left|\delta b(\tilde{Z}, \gamma) - \delta^r + \delta^2 b(\tilde{Z}, \gamma)\right| \leq 2L\left(\delta b(\tilde{Z}, \gamma) + \delta^r + \delta^2 b(\tilde{Z}, \gamma)\right) \leq 2L\delta\left(2b(\tilde{Z}, \gamma) + \frac{r}{L}\right).
\]
Since the normal distribution is symmetric, the arguments we used in order to bound $P_{11}$, and $P_{12}$ are applied to $P_{22}$, and $P_{21}$ respectively, so we obtain

$$P_{21} = P\left( \frac{r}{2} - \delta b(\tilde{Z}, \gamma) - \delta |\eta| \leq \eta \leq \frac{r}{2} |\tilde{Z}, \gamma| \right) \leq 2L\delta \left( 2b(\tilde{Z}, \gamma) + \frac{r}{2} \right),$$

$$P_{22} = P\left( \frac{r}{2} \leq \eta \leq \frac{r}{2} + \delta b(\tilde{Z}, \gamma) + \delta |\eta| |\tilde{Z}, \gamma| \right) \leq 2L\delta \left( 3b(\tilde{Z}, \gamma) + r \right).$$

Thus, we conclude from the above calculations that

$$E\{P_1 + P_2\} = E\{P_{11} + P_{12} + P_{21} + P_{22}\} = E\{4L\delta \left( 3b(\tilde{Z}, \gamma) + r \right) + 4L\delta \left( 2b(\tilde{Z}, \gamma) + \frac{r}{2} \right) \leq 8L\delta \left( 3b(\tilde{Z}, \gamma) + r \right) \leq 24L\delta E(\|\tilde{Z}\| + |\gamma| + 1) + 8L\delta r \leq C_1\delta + C_2\delta = C\delta,$$

for some constant $C$. This follows because the expectations above are bounded. Now, since we bounded the second moment of the envelope function, here $\delta = \phi^2(\delta)$, so $\phi(\delta) = \delta^{1/2}$. If we take $\delta = 1/R_n$, then the solution of $R_n^2 (\frac{1}{R_n})^{1/2} \leq \sqrt{n}$, is $R_n \leq n^{1/3}$, and we conclude that $n^{1/3}||\hat{\theta}_{r,n} - \theta_0|| = O_p(1)$.

### 3.2.3 Proof of Theorem 3.3

The proof is based on verifying conditions of Theorem 2.3. We have already obtained the consistence of the HT estimator and the rate of convergence $R_n = n^{1/3}$. We have also showed that $\mathcal{M}_\delta$ is a VC-class of functions, and that $E B_\delta^2 \leq \phi^2(\delta) = C\delta$. In the next three steps we verify conditions (2.9)-(2.11).

#### Step 1: Verification of condition (2.9)

Note that in our case $\delta^{-2}\phi^2(\delta) = C\delta^{-1}$. Therefore for any fixed $\zeta > 0$ we have that $\zeta \delta^{-2}\phi^2(\delta) = C\zeta/\delta \to \infty$, as $\delta \to 0$. Since $B_\delta \leq 2$ for all $\delta$, it follows that $1\{B_\delta >$
\[ \zeta \delta^{-2} \phi^2(\delta) \leq 1 \{ 2 > C\zeta/\delta \} = 0, \text{ as } \delta \to 0. \] So we conclude that
\[
\lim_{\delta \downarrow 0} \frac{\mathbb{E} B_{\delta}^2 \mathbf{1} \{ B_{\delta} > \zeta \delta^{-2} \phi^2(\delta) \}}{\phi^2(\delta)} = 0,
\]
and condition (2.9) holds.

**Step 2: Verification of condition (2.10)**

Condition (2.10) in our case takes the form
\[
\lim_{\xi \downarrow 0} \limsup_{\delta \downarrow 0} \frac{1}{\delta} \sup_{\|h-g\|<\xi, \|h\|\|g\|\leq K} \mathbb{E}(m_{\theta_0+\delta g} - m_{\theta_0+\delta h}) = 0.
\]

Unlike bounding the envelope function, here we first take expectation and then calculate the supremum, thus the treatment is much easier. Let \( g = (g_\phi, g_t), h = (h_\phi, h_t) \in \Theta \). Using the definition of \( D_\theta \) given in (1.5) we have
\[
m_{\theta_0+\delta g} - m_{\theta_0+\delta h} = \mathbf{1}_{D_{\theta_0+\delta g}} - \mathbf{1}_{D_{\theta_0+\delta h}}
= \mathbf{1} \{ |Z^T U_{\phi_0+\delta g_\phi} - (t_0 + \delta g_t)| \leq \frac{r}{2} \}
- \mathbf{1} \{ |Z^T U_{\phi_0+\delta g_\phi} - (t_0 + \delta h_t)| \leq \frac{r}{2} \}
= \mathbf{1} \{ -\frac{r}{2} - W(\delta) \leq \lambda^T U_{\phi_0+\delta h_\phi} \leq \frac{r}{2} - W(\delta) \}
- \mathbf{1} \{ -\frac{r}{2} - W_h(\delta) \leq \lambda^T U_{\phi_0+\delta h_\phi} \leq \frac{r}{2} - W_h(\delta) \},
\]
where we define
\[
W(\delta) = W_{\phi}(\delta) + W_{\phi,h}(\delta),
W_\phi(\delta) = \tilde{Z}^T (U_{\phi_0+\delta g_\phi} - U_{\phi_0}),
W_{\phi,h}(\delta) = \lambda^T (U_{\phi_0+\delta g_\phi} - U_{\phi_0+\delta h_\phi}),
W_h(\delta) = \tilde{Z}^T (U_{\phi_0+\delta h_\phi} - U_{\phi_0}) - \delta h_t.
\]
Therefore
\[
(m_{\theta_0+\delta g} - m_{\theta_0+\delta h})^2 = \mathbf{1} \{ -\frac{r}{2} - \max[W(\delta), W_h(\delta)] \leq \lambda^T U_{\phi_0+\delta h_\phi} \leq -\frac{r}{2} - \min[W(\delta), W_h(\delta)] \}
+ \mathbf{1} \{ \frac{r}{2} - \max[W(\delta), W_h(\delta)] \leq \lambda^T U_{\phi_0+\delta h_\phi} \leq \frac{r}{2} - \min[W(\delta), W_h(\delta)] \}.
\]

32
Let us consider the case where \( W(\delta) > W_h(\delta) \). Then (3.12) becomes

\[
1 \left\{ \frac{r}{2} - W(\delta) \leq \lambda^T U_{\phi_0 + \delta h_\phi} \leq \frac{r}{2} - W_h(\delta) \right\} \\
+ 1 \left\{ \frac{r}{2} - W(\delta) \leq \lambda^T U_{\phi_0 + \delta h_\phi} \leq \frac{r}{2} - W_h(\delta) \right\}.
\]

Conditioning on \( \tilde{Z} \) and taking expectation of \( \lambda^T U_\phi \) we get the expression

\[
E \left[ \mathbb{P} \left( \lambda^T U_{\phi_0 + \delta h_\phi} \leq -\frac{r}{2} - W_h(\delta) \mid \tilde{Z} \right) \right] - \mathbb{P} \left( \lambda^T U_{\phi_0 + \delta g_\phi} \leq -\frac{r}{2} - W_g(\delta) \mid \tilde{Z} \right) \\
+ E \left[ \mathbb{P} \left( \lambda^T U_{\phi_0 + \delta h_\phi} \leq -\frac{r}{2} - W_h(\delta) \mid \tilde{Z} \right) \right] - \mathbb{P} \left( \lambda^T U_{\phi_0 + \delta h_\phi} \leq -\frac{r}{2} - W_h(\delta) \mid \tilde{Z} \right)
\]

\[
\leq E \mathbb{P} \left( \eta \leq -\frac{r}{2} - W_h(\delta) \mid \tilde{Z} \right) - \mathbb{P} \left( \eta \leq -\frac{r}{2} - W_g(\delta) \mid \tilde{Z} \right) \\
+ E \mathbb{P} \left( \eta \leq \frac{r}{2} - W_h(\delta) \mid \tilde{Z} \right) - \mathbb{P} \left( \eta \leq \frac{r}{2} - W_g(\delta) \mid \tilde{Z} \right)
\]

\[
\leq LE \left[ -\frac{r}{2} - \tilde{Z}^T (U_{\phi_0 + \delta g_\phi} - U_{\phi_0}) + \delta h_t + \frac{r}{2} + \tilde{Z}^T (U_{\phi_0 + \delta g_\phi} - U_{\phi_0}) - \delta g_t \right] \\
+ LE \left[ \frac{r}{2} - \tilde{Z}^T (U_{\phi_0 + \delta h_\phi} - U_{\phi_0}) + \delta h_t - \frac{r}{2} + \tilde{Z}^T (U_{\phi_0 + \delta h_\phi} - U_{\phi_0}) - \delta g_t \right],
\]

for some constant \( L \) depending on \( \sigma \) only. Note that \( \|U_{\phi_0 + \delta g_\phi} - U_{\phi_0 + \delta h_\phi}\| \leq \delta |g_\phi - h_\phi| \), thus we obtain

\[
\lim_{\xi \to 0} \lim_{\delta \to 0} \frac{1}{\delta} \sup_{\|h - g\| < \xi} 2LE \left[ \|\tilde{Z}^T (U_{\phi_0 + \delta g_\phi} - U_{\phi_0 + \delta h_\phi}) - \delta (g_t - h_t) \right] \\
\leq \lim_{\xi \to 0} \lim_{\delta \to 0} \frac{1}{\delta} \sup_{\|h - g\| < \xi} 2LE \left[ \|\tilde{Z}\| \|U_{\phi_0 + \delta g_\phi} - U_{\phi_0 + \delta h_\phi}\| + \delta |g_t - h_t| \right] \\
\leq \lim_{\xi \to 0} \lim_{\delta \to 0} \frac{1}{\delta} 2LE \left( \|\tilde{Z}\| \delta \xi + \delta \xi \right) \\
= \lim_{\xi \to 0} 2LE \left( \|\tilde{Z}\| \xi + \xi \right) = 0.
\]

The argument above holds for the case where \( W(\delta) < W_h(\delta) \), and for any \( K \) such that \( \|h\| \vee \|g\| \leq K \). Therefore we conclude that condition (2.10) is verified.
Step 3: Verification of condition (2.11)

Here we need to calculate
\[ \lim_{\delta \downarrow 0} \frac{\mathbb{E}(m_{\theta_0+\delta g} - m_{\theta_0+\delta h})^2}{\delta^2(\delta)} = \lim_{\delta \downarrow 0} \frac{1}{\delta} \mathbb{E}(m_{\theta_0+\delta g} - m_{\theta_0+\delta h})^2. \]

Consider the case in (3.12) where \( W(\delta) > W_h(\delta) \). Thus, conditioning on \( \tilde{Z} \) we get
\[ \frac{1}{\delta} \mathbb{E}(m_{\theta_0+\delta g} - m_{\theta_0+\delta h})^2 = \frac{1}{\delta} \mathbb{E} \left\{ f \left( -\frac{r}{2} - \tilde{Z}^T(U_{\phi_0+\delta \phi} - U_{\phi_0}) + \delta h \right) - f \left( -\frac{r}{2} - \tilde{Z}^T(U_{\phi_0+\delta \phi} - U_{\phi_0}) + \delta g \right) \right\} \]
\[ + \frac{1}{\delta} \mathbb{E} \left\{ f \left( \frac{r}{2} - \tilde{Z}^T(U_{\phi_0+\delta \phi} - U_{\phi_0}) + \delta h \right) - f \left( \frac{r}{2} - \tilde{Z}^T(U_{\phi_0+\delta \phi} - U_{\phi_0}) + \delta g \right) \right\}. \]

Our goal is to take the limit w.r.t \( \delta \) for this expression which in our case is the same as derivative w.r.t. \( \delta \). We recall that \( A(\tilde{Z}) = (\tilde{Z}^T(U_{\phi_0-\pi/2},1)^T, and derivate under the expectation (DCT) as follows.
\[
\lim_{\delta \downarrow 0} \frac{1}{\delta} \mathbb{E}(m_{\theta_0+\delta g} - m_{\theta_0+\delta h})^2
= \mathbb{E} \left\{ f \left( -\frac{r}{2} \right) h^T A(\tilde{Z}) - f \left( -\frac{r}{2} \right) g^T A(\tilde{Z}) \right\}
+ \mathbb{E} \left\{ f \left( \frac{r}{2} \right) h^T A(\tilde{Z}) - f \left( \frac{r}{2} \right) g^T A(\tilde{Z}) \right\}
= 2f \left( \frac{r}{2} \right) \mathbb{E} |(h - g)^T A(\tilde{Z})|.
\]

since \( f(x) = f(-x) \). Now, consider the case where \( W(\delta) < W_h(\delta) \) we get the variance given in (3.7)
\[ 2f \left( \frac{r}{2} \right) \mathbb{E} |(h - g)^T A(\tilde{Z})|. \]

3.2.4 Proof of Theorem 3.4

Under the premise of the theorem, there exist a constant \( c \) such that all sinusoids in the Hough domain are in \([0, 2\pi) \times [0, c]\). By the definition of the HT, for fixed \( n \), \( \hat{\theta}_{r,n} \) is the center of the finite line-segment of length \( r \) that crosses over the maximal number of these sinusoids. Thus, \( nM_{r,n}(\hat{\theta}_{r,n}) \) is the corresponding number of such sinusoids. Now, in order to shift this estimate to infinity, one should add at least \( nM_{r,n}(\hat{\theta}_{r,n}) + 1 \) sinusoids at infinity. In particular, in order to shift one sinusoid (in the parameter space) to infinity one should replace the corresponding point \( Z_i \) (in the observations space) by a point at infinity. This point should lie on the same ray crossing the origin and \( Z_i \). Thus, the smallest
contamination fraction under which \( \hat{\theta}_{r,n} \) breaks down is 
\[
\frac{(nM_{r,n}(\hat{\theta}_{r,n}) + 1)}{(n + nM_{r,n}(\hat{\theta}_{r,n}) + 1)}.
\]
Because \( M_{r,n}(\hat{\theta}_{r,n}) \rightarrow M_r(\theta_0) = \int_{-r/2}^{r/2} f(x)dx \), therefore the result for \( \epsilon_{add}(\hat{\theta}_{r,n}; Z_n) \) follows. For the replacement BP, it is sufficient to note that under the premise of the theorem at least \( \lceil nM_{r,n}(\hat{\theta}_{r,n})/2 \rceil \) sinusoids should be replaced.
Chapter 4

Simulations results

In this chapter we present numerical examples illustrating several properties of the HT estimator discussed in previous chapters.

4.1 Choice of the parameter $r$

The properties of the HT estimator depend on a choice of the parameter $r$. In the previous chapter we showed that the HT is consistent for any choice of $r$, and that its asymptotic distribution depends on $r$. For example, if $r$ is very small, then the matrix $V_0$ given in (3.6) is nearly singular.

An important consequence of the choice of $r$ is that large values of $r$ lead to a large connected solution set, and in this case the estimation accuracy depends on the way the estimator is chosen from the solution set. On the other hand, small values of $r$ lead to ”under-smoothed” dual plot, and the solution set is a union of many disconnected sets. In this case estimation accuracy of the average estimator may be very poor. Figure 4.1(a) displays a large connected solution set while in Figure 4.1(c) an ”under-smoothed” dual plot is presented. Since we seek for a distinguished point of maximum, Figure 4.1(b) would be the optimal dual plot.

An optimal choice of $r$ can be defined as follows. Let $\mathcal{R}$ be a set of possible values of $r$. A reasonable choice of $r$ would be

$$
\hat{r} = \arg\min_{r \in \mathcal{R}} tr\{cov(W)\}, \tag{4.1}
$$

where $W$ is the limiting random variable given in Theorem 3.3. Since we do not have an
analytic expression for the distribution of $W$, it is impossible to calculate (4.1) directly. However, here we tried to define a ”rule of thumb” based on simulations.

We performed extensive experiments as follows. Let $n$ be the sample size, $i = 1, ..., N$ be the simulation index, and $j = 1, ..., M$ be the experiment index. Set $\mathcal{R} = [0.276 : (0.025) : 0.976]$ and define $\hat{\theta}_{i,j,r,n}$ to be the estimator computed in the $i$-th simulation in experiment $j$. In each experiment $j$ we calculate for each value of $r$

$$S(r,j) = \sum_{i=1}^{N} \| \hat{\theta}_{i,j,r,n} - \theta_0 \|^2.$$  \hspace{1cm} (4.2)

Then we compute $\hat{r}_j = \arg \min_{r \in \mathcal{R}} S(r,j)$ which is an (sort of) empirical version of (4.1).

For the implementation we used the square $[0,2\pi) \times [0,1]$ as the search region. The HT estimator is computed by direct maximization of the objective function on the above square using a grid comprised of 10,000 points. Because the solution is not unique, the average of the grid points where the maximum is achieved is taken as the estimate. In particular we define $n = 200$, $N = 200$, $M = 192$. In all experiments we set $\phi_0 = \pi/4$, $t_0 = 0.5$, $\lambda_i \sim N_2(0,0.2^2I)$ for all $i$. Figure 4.2 presents the histogram of $\hat{r}_j$, $j = 1, ..., 192$, while a boxplot of the data is displayed in Figure 4.3. In Figure 4.4 we illustrate (4.2) as a function of $r$ for $j = 1, .., 192$.

Based on the numerical results, we can say that the optimal $r$ is about $3\sigma$, where $\sigma$ is the standard deviation of the noise variable $\eta$. In spite of the fact that our experiment is based on just one set of parameters and one kind of noise, it is extensive enough for
Figure 4.2: Histogram of $\hat{r}_j$, $j = 1, ..., M$ for $M = 192$ experiments. In each experiment $r \in \mathcal{R} = [0.276 : (0.025) : 0.976]$, i.e., 29 potential values. For each value of $r$ (4.2) is calculated based on a sample size $n = 200$, and $N = 200$ simulations.

Figure 4.3: Boxplot for the experiments data.
motivating further investigation in that direction.

Figures 4.5-4.10 may give some intuition regarding the limiting distribution of $W_{r,n} := n^{1/3}(\hat{\theta}_{r,n} - \theta_0)$. Here we run the same algorithm used for finding the optimal $r$, but only for 3 values, $r = 0.276$, 0.6, 0.976 which are the minimum, optimal, and maximum values respectively. Now we set $n = 300$, $N = 250$, and used the same $\theta_0$ and $\sigma$ given above. A look at these figures suggests that there is a negative correlation between the estimators $\hat{\phi}$, $\hat{t}$. It turns out that if $t_0$ is over estimated then $\phi_0$ is underestimated and vice versa. Figure 4.7 displays the empirical distribution of $W_{r,n}$ for the optimal $r$. It is clear that the cloud of data is more concentrated around zero than in the other cases.

### 4.2 Vertical line

As already mentioned in Chapter 1, one of the drawbacks of the HT estimator in the Cartesian parameterization is its unbounded parameter space. As a result, vertical lines cannot be estimated by this version of the HT. The following example illustrates that this is not the case for the HT in the polar parameterization. We set $\phi_0 = 0$, $t_0 = 0.6$, $\lambda \sim N_2(0, \sigma^2 I)$, $\sigma = 0.01$, and generate 200 observations from this model. Note that when
Figure 4.5: Distribution of $W_{0.276,300}$ based on 250 simulations.

Figure 4.6: Histogram of $W_{0.276,300}$. 
Figure 4.7: Distribution of $W_{0.6,300}$ based on 250 simulations.

Figure 4.8: Histogram of $W_{0.6,300}$. 

41
Figure 4.9: Distribution of $W_{0.976,300}$ based on 250 simulations.

Figure 4.10: Histogram of $W_{0.976,300}$. 

42
Figure 4.11: The accumulator array for vertical line detection. The estimated parameters are $\hat{\phi} = 0$, $\hat{t} = 0.59596$.

the parameter $\phi_0 = 0$, the line under detection is vertical. Here we used $r = 0.03$ for the searching line. In Figure 4.11 we present the accumulator array generated by the HT algorithm. We see that there is a clear point of maximum in the neighborhood of $\theta_0$. In particular, we obtained $\hat{\phi} = 0$, $\hat{t} = 0.59596$ which is a very good result. The HT in the Cartesian parameterization could not estimate the above line.

4.3 Robustness

4.3.1 Breakdown point

In order to demonstrate the robustness properties of the HT estimator we consider a numerical example similar to the example given in Rousseeuw and Leroy (1987). The authors there consider a Cartesian parameterization model while we are interested in the model discussed in Chapter 3, $\tilde{Z}_i^T U_{\phi_0} = t_0$, where we observe $Z_i = \tilde{Z}_i + \lambda_i$, and $\lambda_i \sim N_2(0, \sigma^2 I)$ for all $i$. 

43
Figure 4.12: (a) An illustration of the breakdown properties of the HT estimator. The cluster on the right is the contamination of 20 "bad" observations. (b) The accumulator array.

In the first illustration we present 30 "good" observations uniformly distributed on the line with parameters $\theta_0 = (3\pi/4, \sqrt{2})$, where $\sigma = 0.2$. Then a cluster of 20 "bad" observations is added. These observations follow a bivariate Gaussian distribution with expectation $(7, 2)$ and covariance matrix $0.25I$. The parameter $r$ of the HT is set to 0.6. The HT is calculated by direct maximization of (3.4) on $[0, 2\pi] \times [0, 10]$ using a uniform rectangular grid comprised of 10,000 points. Because the solution is not unique, the average of the grid points where the maximum is achieved is taken as the estimate.

Here the density of the noise $f$ is $N(0, 0.2^2)$, thus under conditions of the experiment $\int_{-0.3}^{0.3} f(x)dx = 0.8664$, which in view of Theorem 3.4 approximately corresponds to a 43.32% replacement BP. In Figure 4.12(a) we can see that the HT fits the "good" observations very well under the contamination in the data while the least squares (LS) yields very poor result. In the accumulator array displayed in Figure 4.12(b) we see that there is a clear point of maximum of intersections, corresponds to the "good" data set. In particular $\hat{\theta}_{0.6,50} = (2.3483, 1.4176)$ which is very close to the original values of $\theta_0$. Figure 4.13 displays the second illustration. Here we used the same data, but now the contamination data set contains 24 data points. That means 48% of contamination which is more then the theoretical value given above. Indeed, the estimation of the HT now seems very poor.
Figure 4.13: (a) An illustration of the breakdown properties of the HT estimator. The cluster on the right is the contamination of 24 ”bad” observations. (b) The accumulator array.

4.3.2 Equivariance

Here we illustrate the equivariance property of the HT estimator discussed in Chapter 3. Figure 4.14(a) displays a blue cloud of 100 observations generated from the model $Z_i^T U_{\pi/4} - \lambda_i^T U_{\pi/4} = \sqrt{2}$, where the pairs $\tilde{Z}_i$ are uniformly distributed on the line with the parameters given above, and $\lambda_i \sim N_2(0, 0.01^2 I)$ for all $i$. The parameter $r$ of the HT is set to 0.03. The red cloud is a result of rotating the original observations with an angle $\alpha = \pi/8 = 0.3927$. We denote the original observations by $Z$, and the transformed observations by $Z^*$. Figure 4.14(b) is the corresponding HT. In particular, $\hat{\phi}(Z^*) + 0.3808 = \hat{\phi}(Z)$, thus the equivariance property holds.
Figure 4.14: Equivariance property of the HT. The transformation is rotation of $\pi/8$ of the data set. (a) The blue cloud is the original data set while the red is the transformed data. (b) The blue and red sinusoids are defined respectively to the data sets.
Appendix

Here we present the Matlab code used in simulations reported in this work.

```matlab
%======================================================================
% USAGE: observation_space_f(phi_0, t_0, n, sigma)
% GENERATION OF RANDOM SAMPLE FROM THE MODEL
% Input: model parameters and sample definition
% "t_0" - model parameter
% "phi_0" - model parameter
% "n" - sample size
% "sigma" - noise standard deviation
% This function creates: random sample in the original domain
%======================================================================

function observation_space=observation_space_f(phi_0, t_0, n, sigma)

x_dist=unifrnd(0,1,n,1);
gamma=normrnd(0, sigma,n,1); % horizontal noise
eta=normrnd(0, sigma,n,1); % vertical noise

U_0=[cos(phi_0);sin(phi_0)]; % direction vector
A=[sin(phi_0),-cos(phi_0);cos(phi_0),sin(phi_0)]; % rotation matrix

V=(x_dist,x_dist*0); % points on the x axis
W=V*A+t_0*(ones(n,1)*U_0'); % unobserved data

observation_space=[W(:,1)+gamma,W(:,2)+eta]; % observed (noisy) data
```
function parameter_space_f(m, k, Z)

phi = 0:2*pi/(m-1):2*pi;
t = 0:1/(k-1):1;

U = [(cos(phi));(sin(phi))]

parameter_space = Z*U'; % each line is a sine graph

function estimation_f(t_sample, r, phi_0, t_0, m, k, n)

phi = 0:2*pi/(m-1):2*pi;
t = 0:1/(k-1):1;

t_deterministic = reshape(repmat(t,m,1),1,k*m);
t_deterministic = repmat(t_deterministic,n,1);
t_sample = repmat(t_sample,1,k);

M_rn = (1/n)*sum((abs(t_sample-t_deterministic)<=0.5*r));
M_rn = reshape(M_rn,k,m);

[ii,jj] = find(max(max(M_rn))==M_rn);
t_est = mean(t(ii));
phi_est = mean(phi(jj));

estimation = (t_est-t_0)^2+(phi_est-phi_0)^2;

estimation

48
% USAGE: main program
% GENERATION OF SIMULATIONS AND EXPERIMENTS

\[
\begin{align*}
t_0 &= 0.5; \\
\phi_0 &= \pi/4; \\
n &= 200; \\
sigma &= 0.2; \\
m &= 100; \\
k &= 100; \\
M &= 192; \quad \text{% experiments number} \\
N &= 200; \quad \text{% simulations number} \\
r &= 0.276:0.025:0.976; \\
\end{align*}
\]

\[
\text{for experiment=1:M} \\
\quad \text{for r_number=1:size(r,2)} \\
\quad \quad \text{for simulation_number=1:N} \\
\quad \quad \quad Z = \text{observation_space_f}(t_0, \phi_0, n, \sigma); \\
\quad \quad \quad t_sample = \text{parameter_space_f}(m, k, Z); \\
\quad \quad \quad \text{variance(simulation_number)} = \text{estimation_f}(t_sample, r(r_number), \phi_0, t_0, m, k, n); \\
\quad \quad \quad \text{end} \\
\quad \quad \text{var_W_n(experiment, r_number) = sum(variance);} \\
\quad \quad \text{end} \\
\quad \text{end} \\
\text{min_var_W_n_index = find(min(var_W_n(experiment, :)) == var_W_n(experiment, :));} \\
\text{optimal_r(experiment, 1:size(min_var_W_n_index, 2)) = r(min_var_W_n_index);} \\
\text{end}
\]
Bibliography


51